

Ré-identification sans coordination dans les types de données répliquées sans conflits (CRDTs)

THÈSE

présentée et soutenue publiquement le TODO : Définir une date

pour l'obtention du

Doctorat de l'Université de Lorraine
(mention informatique)

par

Matthieu Nicolas

Composition du jury

<i>Président :</i>	Stephan Merz
<i>Rapporteurs :</i>	Le rapporteur 1 de Paris
	Le rapporteur 2
	suite taratata
	Le rapporteur 3
<i>Examineurs :</i>	L'examineur 1 d'ici
	L'examineur 2
<i>Membres de la famille :</i>	Mon frère
	Ma sœur

Mis en page avec la classe thesul.

Remerciements

Les remerciements.

*Je dédie cette thèse
à ma machine.
Oui, à Pandore,
qui fut la première de toutes.*

Sommaire

Introduction	1
1 Contexte	1
2 Questions de recherche et contributions	2
2.1 Ré-identification sans coordination pour Conflict-free Replicated Data Types (CRDTs) pour Séquence	2
2.2 Éditeur de texte collaboratif pair-à-pair	3
3 Plan du manuscrit	3
4 Publications	3
Chapitre 1	
État de l’art	5
1.1 Systèmes distribués	5
1.2 Transformées opérationnelles	6
1.3 Types de données répliquées sans conflits	7
1.3.1 Sémantique en cas de concurrence	9
1.3.2 Modèle de synchronisation	13
1.3.3 Adoption dans la littérature et l’industrie	20
1.4 Séquences répliquées sans conflits	21
1.4.1 Approche à pierres tombales	21
1.4.2 Approche à identifiants densément ordonnés	26
1.4.3 Synthèse	29
1.5 LogootSplit	30
1.5.1 Identifiants	30
1.5.2 Aggrégation dynamique d’éléments en blocs	31
1.5.3 Modèle de données	32
1.5.4 Modèle de livraison	33
1.5.5 Limites	36

1.6	Mitigation du surcoût des séquences répliquées sans conflits	37
1.7	Synthèse	38
1.8	Proposition	38

Chapitre 2

Conclusions et perspectives 39

2.1	Résumé des contributions	40
2.2	Perspectives	40
2.2.1	Définition de relations de priorité pour minimiser les traitements . .	40
2.2.2	Redéfinition de la sémantique du renommage en déplacement d'élé- ments	40
2.2.3	Définition de types de données répliquées sans conflits plus complexes	40
2.2.4	Étude comparative des différentes familles de CRDTs	40
2.2.5	Définition d'opérations supplémentaires pour fonctionnalités liées à l'édition collaborative	41
2.2.6	Conduction d'expériences utilisateurs d'édition collaborative	41
2.2.7	Comparaison des mécanismes de synchronisation	42
2.2.8	Distance entre versions d'un document	42
2.2.9	Contrôle d'accès	42
2.2.10	Détection et éviction de pairs malhonnêtes	42
2.2.11	Vecteur <i>epoch-based</i>	43
2.2.12	Fusion de versions distantes d'un document collaboratif	44
2.2.13	Rôles et places des bots dans systèmes collaboratifs	44

Annexe A

Algorithmes RENAMEID

Annexe B

Algorithmes REVERTRENAMEID

Index	49
-------	----

Bibliographie

Table des figures

1.1	Modifications concurrentes d'un Ensemble répliqué provoquant un conflit suite à l'ajout et la suppression d'un même élément	8
1.2	Résolution du conflit en utilisant la sémantique <i>Last-Writer-Wins</i> (LWW)	9
1.3	Résolution du conflit en utilisant la sémantique <i>Multi-Value</i> (MV)	10
1.4	Résolution du conflit en utilisant soit la sémantique <i>Add-Wins</i> (AW), soit la sémantique <i>Remove-Wins</i> (RW)	12
1.5	Résolution du conflit en utilisant la sémantique <i>Causal-Length</i> (CL)	12
1.6	Modifications en concurrence d'un Ensemble répliqué par les noeuds A et B	13
1.7	Synchronisation des noeuds A et B en adoptant le modèle de synchronisation par états	15
1.8	Synchronisation des noeuds A et B en adoptant le modèle de synchronisation par opérations	17
1.9	Synchronisation des noeuds A et B en adoptant le modèle de synchronisation par différences d'états	19
1.10	Modifications concurrentes d'une séquence répliquée WOOT	22
1.11	Modifications concurrentes d'une séquence répliquée Replicated Growable Array (RGA)	25
1.12	Entrelacement d'éléments insérés de manière concurrente	26
1.13	Identifiants de positions	27
1.14	Modifications concurrentes d'une séquence répliquée Treedoc	28
1.15	Représentation d'une séquence LogootSplit contenant les éléments "HLO"	31
1.16	Modifications concurrentes d'une séquence répliquée LogootSplit	33
1.17	Résurgence d'un élément supprimé suite à la relivraison de son opération <i>insert</i>	34
1.18	Non-effet de l'opération <i>remove</i> car reçue avant l'opération <i>insert</i> correspondante	35
1.19	Insertion menant à une augmentation de la taille des identifiants	36

Introduction

1 Contexte

- Systèmes collaboratifs (wikis, plateformes de contenu, réseaux sociaux) et leurs bienfaits (qualité de l'info, vitesse de l'info (exemple de crise ?), diffusion de la parole). Démocratisation (sic) de ces systèmes au cours de la dernière décennie.
- En raison du volume de données et de requêtes, adoptent architecture décentralisée. Permet ainsi de garantir disponibilité, tolérance aux pannes et capacité de passage à l'échelle.
- Mais échoue à adresser problèmes non-techniques : confidentialité, souveraineté, protection contre censure, dépendance et nécessité de confiance envers autorité centrale.
- À l'heure où les entreprises derrière ces systèmes font preuve d'ingérence et d'intérêts contraires à ceux de leurs utilisateur·rices (Cambridge Analytica, Prism, non-modération/mise en avant de contenus racistes, invisibilisation de contenus féministes), paraît fondamental de proposer les moyens technologiques accessibles pour concevoir et déployer des alternatives.
- *Matthieu: TODO : Voir si angle écologique/réduction consommation d'énergie peut être pertinent.*
- Systèmes pair-à-pair sont une direction intéressante pour répondre à ces problématiques, de part leur absence d'autorité centrale, la distribution des tâches et leur conception mettant le pair au centre. Mais posent de nouvelles problématiques de recherche.
- Ces systèmes ne disposent d'aucun contrôle sur les noeuds qui les composent. Le nombre de noeuds peut donc croître de manière non-bornée et atteindre des centaines de milliers de noeuds. La complexité des algorithmes de ces systèmes ne doit donc pas dépendre de ce paramètre, ou alors de manière logarithmique.
- De plus, ces noeuds n'offrent aucune garantie sur leur stabilité. Ils peuvent donc rejoindre et participer au système de manière éphémère. S'agit du phénomène connu sous le nom de churn. Les algorithmes de ces systèmes ne peuvent donc pas reposer sur des mécanismes nécessitant une coordination synchrone d'une proportion des noeuds.
- Finalement, ces noeuds n'offrent aucune garanties sur leur fiabilité et intentions. Les noeuds peuvent se comporter de manière byzantine. Pour assurer la confidentialité, l'absence de confiance requise et le bon fonctionnement du système, ce dernier doit être conçu pour résister aux comportements byzantins de ses acteurs.

- Ainsi, il est nécessaire de faire progresser les technologies existantes pour les rendre compatible avec ce nouveau modèle de système. Dans le cadre de cette thèse, nous nous intéressons aux mécanismes de réplication de données dans les systèmes collaboratifs pair-à-pair temps réel.

2 Questions de recherche et contributions

2.1 Ré-identification sans coordination pour CRDTs pour Séquence

- Systèmes collaboratifs permettent aux utilisateur-rices de manipuler et éditer un contenu partagé. Pour des raisons de performance, ces systèmes autorisent généralement les utilisateur-rices à effectuer des modifications sans coordination. Leur copies divergent alors momentanément. Un mécanisme de synchronisation leur permet ensuite de récupérer l'ensemble des modifications et de les intégrer, de façon à converger. Cependant, des modifications peuvent être incompatibles entre elles, car de sémantiques contraires. Un mécanisme de résolution de conflits est alors nécessaire.
- Les CRDTs sont des types de données répliquées. Ils sont conçus pour être répliqués par les noeuds d'un système et pour permettre à ces derniers de modifier les données partagées sans aucune coordination. Dans ce but, ils incluent des mécanismes de résolution de conflits directement au sein leur spécification. Ces mécanismes leur permettent de résoudre le problème évoqué précédemment. Cependant, ces mécanismes induisent un surcoût, aussi bien d'un point de vue consommation mémoire et réseau que computationnel. Notamment, certains CRDTs comme ceux pour la Séquence souffrent d'une croissance monotone de leur surcoût. Ce surcoût s'avère handicapant dans le contexte des collaborations à large échelle.
- Pouvons-nous proposer un mécanisme sans coordination de réduction du surcoût des CRDTs pour Séquence, c.-à-d. compatible avec les systèmes pair-à-pair ?
- Dans le cadre des CRDTs pour Séquence, le surcoût du type de données répliquées provient de la croissance de leurs métadonnées. Métadonnées proviennent des identifiants associés aux éléments de la Séquence par les CRDTs. Ces identifiants sont nécessaires pour le bon fonctionnement de leur mécanisme de résolution de conflits.
- Plusieurs approches ont été proposées pour réduire le coût de ces identifiants. Notamment, [1, 2] proposent un mécanisme de ré-assignation d'identifiants de façon à réduire leur taille. Mécanisme non commutatif avec les modifications concurrentes de la Séquence, c.-à-d. l'insertion ou la suppression. Propose ainsi un mécanisme de transformation des modifications concurrentes pour gérer ces conflits. Mais mécanisme de ré-assignation n'est pas non plus commutatif avec lui-même. De fait, utilisent un algorithme de consensus pour empêcher l'exécution du mécanisme en concurrence.
- Proposons RenamableLogootSplit, un nouveau CRDT pour Séquence. Intègre un mécanisme de renommage directement au sein de sa spécification. Intègre un mé-

canisme de résolution de conflits pour les renommages concurrents. Permet ainsi l'utilisation du mécanisme de renommage par les noeuds sans coordination.

2.2 Éditeur de texte collaboratif pair-à-pair

- Systèmes collaboratifs adoptent généralement architecture décentralisée. Disposent d'autorités centrales qui facilitent la collaboration, l'authentification des utilisateur-rices, la communication et le stockage des données.
- Mais ces systèmes introduisent une dépendance des utilisateur-rices envers ces mêmes autorités centrales, une perte de confidentialité et de souveraineté.
- Pouvons-nous concevoir un éditeur de texte collaboratif sans autorités centrales, c.-à-d. un éditeur de texte collaboratif à large échelle pair-à-pair ?
- Ce changement de modèle, d'une architecture décentralisée à une architecture pair-à-pair, introduit un ensemble de problématiques de domaines variés, e.g.
 - (i) Comment permettre aux utilisateur-rices de collaborer en l'absence d'autorités centrales pour résoudre les conflits de modifications ?
 - (ii) Comment authentifier les utilisateur-rices en l'absence d'autorités centrales ?
 - (iii) Comment structurer le réseau de manière efficace, c.-à-d. en limitant le nombre de connexion par pair ?
- Présentons Multi User Text Editor (MUTE) [3]. S'agit, à notre connaissance, du seul prototype complet d'éditeur de texte collaboratif temps réel pair-à-pair chiffré de bout en bout. Allie ainsi les résultats issus des travaux de l'équipe sur les CRDTs pour Séquence [4, 5] et l'authentification des pairs dans systèmes distribués [6, 7] aux résultats de la littérature sur mécanismes de conscience de groupe *Matthieu: TODO : Trouver et ajouter références*, les protocoles d'appartenance aux groupe [8, 9], les réseaux pair-à-pair [10] et les protocoles d'établissement de clés de groupe [11].

3 Plan du manuscrit

4 Publications

Chapitre 1

État de l’art

Sommaire

1.1	Systèmes distribués	5
1.2	Transformées opérationnelles	6
1.3	Types de données répliquées sans conflits	7
1.3.1	Sémantique en cas de concurrence	9
1.3.2	Modèle de synchronisation	13
1.3.3	Adoption dans la littérature et l’industrie	20
1.4	Séquences répliquées sans conflits	21
1.4.1	Approche à pierres tombales	21
1.4.2	Approche à identifiants densément ordonnés	26
1.4.3	Synthèse	29
1.5	LogootSplit	30
1.5.1	Identifiants	30
1.5.2	Aggrégation dynamique d’éléments en blocs	31
1.5.3	Modèle de données	32
1.5.4	Modèle de livraison	33
1.5.5	Limites	36
1.6	Mitigation du surcoût des séquences répliquées sans conflits	37
1.7	Synthèse	38
1.8	Proposition	38

1.1 Systèmes distribués

- Contexte des systèmes distribués à large échelle
- Réplique les données afin de pouvoir supporter les pannes
- Adopte le paradigme de la réplication optimiste [12]
- Autorise les noeuds à consulter et à modifier la donnée sans aucune coordination entre eux

- Autorise alors les noeuds à diverger temporairement
- Permet d'être toujours disponible, de toujours répondre aux requêtes même en cas de partition réseau
- Permet aussi, en temps normal, de réduire le temps de réponse (privilégie la latence) [13]
- Comme ce modèle autorise les noeuds à modifier la donnée sans se coordonner, possible d'effectuer des modifications concurrentes
- Généralement, un mécanisme de résolution de conflits est nécessaire afin d'assurer la convergence des noeuds dans une telle situation
- Plusieurs approches ont été proposées pour implémenter un tel mécanisme

1.2 Transformées opérationnelles

- Approche permettant de gérer des modifications concurrentes sur un type de données
- Consiste à transformer les opérations par rapport aux effets des opérations concurrentes pour rendre les rendre commutatives. Permet de rendre l'ordre d'intégration des opérations sans importance par rapport à l'état final obtenu
- Se décompose en 2 parties : algorithmes (génériques) et fonctions de transformations (spécifiques au type de données)
- Plusieurs algorithmes OT adoptent une architecture centralisée (trouver citations)
- Cette architecture pose des problèmes de performances (bottleneck), sécurité (SPOF), coût, d'utilisabilité (mode offline), pérennité (disparition du service), vie privée et de résistance à la censure.
- Pour ces raisons, des algorithmes reposant sur une architecture décentralisée ont été proposés
- Mais ne règlent qu'en partie ces limites
- Notamment, ne sont pas adaptés à des systèmes P2P dynamiques
- Besoin de vector clocks sur chaque opération pour détecter la concurrence. Vector clocks adaptés dans systèmes à nombre de pairs fixe, mais pas aux systèmes dynamiques (revoir causal barrier pour p-e nuancer ce propos).
- Néanmoins, cette approche a permis de démocratiser les systèmes collaboratifs via son adoption par des services tels que Google Docs, Overleaf, Framapad
- De plus, dans le cadre de ces travaux, ont été définies les propriétés CCI [14].
- Remettre en question la propriété Causalité des CCI. Généralement, confond causalité et happen-before et exprime en finalité une contrainte trop forte. Cette contrainte peut réduire la réactivité du système (exemple avec 2 insertions sans liens mais qui force d'attendre la 1ère pour intégrer la 2nde). Causalité pose aussi des problèmes de passage à l'échelle car repose sur vector clocks. IMO, doit relaxer cette propriété pour pouvoir construire systèmes à large échelle.

Matthieu: TODO : Mentionner TP1 et TP2

Matthieu: TODO : Spécification faible et forte des séquences répliquées

1.3 Types de données répliquées sans conflits

Pour limiter la coordination entre les noeuds, les systèmes distribués adoptent le paradigme de la réplication optimiste. Ce paradigme consiste à ce que chaque noeud possède une copie de la donnée. Chaque noeud possède le droit de la modifier sans se coordonner avec les autres noeuds. Noeuds peuvent alors temporairement diverger, c.-à-d. posséder des états différents. Un mécanisme de synchronisation leur permet ensuite de partager leurs modifications respectives et de nouveau converger. Ce paradigme offre ainsi aux noeuds une haute disponibilité *Matthieu: TODO : insérer ref sticky model* ainsi qu'une faible latence.

Afin d'ordonner les modifications effectués dans un système, la littérature repose généralement sur la relation *happens-before* [15]. Nous l'adaptions ci-dessous à notre contexte :

Définition 1 (Relation *happens-before*) *La relation happens-before indique qu'une modification m_1 a eu lieu avant une modification m_2 , notée $m_1 \rightarrow m_2$, si et seulement si une des conditions suivantes est respectée :*

- (i) m_1 a eu lieu avant m_2 sur le même noeud.
- (ii) m_1 a été délivrée au noeud auteur de m_2 avant la génération de m_2 .
- (iii) Il existe une modification m telle que $m_1 \rightarrow m \wedge m \rightarrow m_2$.

Dans le cadre d'un système distribué, on note que la relation *happens-before* ne permet pas d'établir un ordre total entre les modifications¹. En effet, deux modifications m_1 et m_2 peuvent être effectuées en parallèle par deux noeuds différents, sans avoir connaissance de la modification de leur pair respectif. De telles modifications sont alors dites *concurrentes* :

Définition 2 (Concurrence) *Deux modifications m_1 et m_2 sont concurrentes, noté $m_1 \parallel m_2$, si et seulement si $m_1 \nrightarrow m_2 \wedge m_1 \nrightarrow m_2$.*

Lorsque les modifications possibles sur un type de données sont commutatives, l'intégration des modifications effectuées par les autres noeuds, même concurrentes, ne nécessite aucun mécanisme particulier. Cependant, les modifications permises par un type de données ne sont généralement pas commutatives car de sémantiques contraires. Ainsi, une exécution dans un système distribué suivant le paradigme de réplication optimiste peut mener à la génération de modifications concurrentes non commutatives. Nous parlons alors de conflits. La figure Figure 1.1 présente un scénario où des modifications de sémantiques opposées sont générées en concurrence.

Dans cet exemple, deux noeuds A et B répliquent et partagent un même Ensemble. Les deux noeuds possèdent le même état initial : $\{a\}$. Le noeud A retire l'élément a de l'ensemble, *rmv(a)*. Puis, le noeud A ré-ajoute l'élément a dans l'ensemble via l'opération de modification *add(a)*. En concurrence, le noeud B retire lui aussi l'élément a de l'ensemble. Les deux noeuds se synchronisent ensuite.

À l'issue de ce scénario, l'état à produire n'est pas trivial : le noeud A a exprimé son intention d'ajouter l'élément a à l'ensemble, tandis que le noeud B a exprimé son intention

1. Nous utilisons le terme *modifications* pour désigner les *opérations de modifications* des types abstraits de données afin d'éviter une confusion avec le terme *opération* introduit ultérieurement.

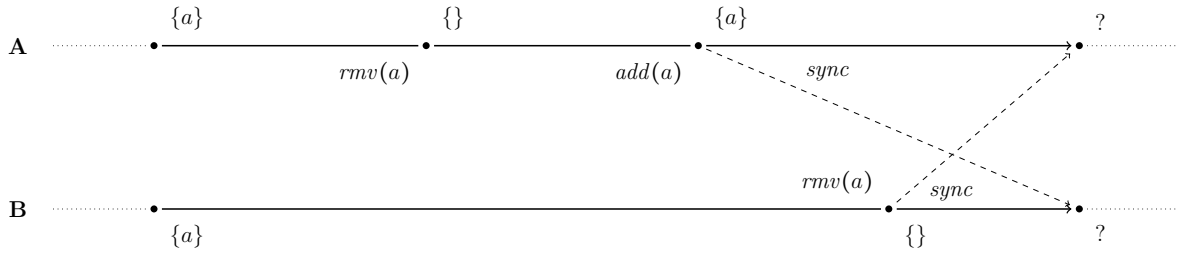


FIGURE 1.1 – Modifications concurrentes d'un Ensemble répliqué provoquant un conflit suite à l'ajout et la suppression d'un même élément

contraire de retirer l'élément a de ce même ensemble. Ainsi, les états $\{a\}$ et $\{\}$ semblent tous les deux corrects et légitimes dans cette situation. Il est néanmoins primordial que les noeuds choisissent et convergent vers un même état pour leur permettre de poursuivre leur collaboration. Pour ce faire, il est nécessaire de mettre en place un mécanisme de résolution de conflits, potentiellement automatique.

Les CRDTs [16, 17, 18] répondent à ce besoin. Les CRDTs sont de nouvelles spécifications des types de données existants, e.g. l'Ensemble ou la Séquence. Ces nouvelles spécifications sont conçues pour être utilisées dans des systèmes distribués adoptant la réplication optimiste. Ainsi, elles offrent les deux propriétés suivantes :

- (i) Les CRDTs peuvent être modifiés sans coordination avec les autres noeuds.
- (ii) Tous les noeuds ayant observés le même ensemble de modifications obtiendront des états équivalents. Cette propriété est connue sous le nom de *convergence forte* Matthieu: TODO : Ajouter ref.

Pour offrir la propriété de *convergence forte*, la spécification des CRDTs reposent sur la théorie des treillis :

- (i) Les différents états possibles d'un CRDT forment un sup-demi-treillis, possédant une relation d'ordre partiel \leq .
- (ii) Les modifications génèrent par inflation un nouvel état supérieur ou égal à l'état original d'après \leq .
- (iii) Il existe une fonction de fusion qui, pour toute paire d'états, génère l'état minimal supérieur d'après \leq aux deux états fusionnés. On parle aussi de borne supérieure ou de Least Upper Bound (LUB) pour catégoriser l'état résultant de cette fusion.

Malgré leur spécification différente, les CRDTs partagent la même sémantique et interface que les types non-répliqués correspondant du point de vue des utilisateurs. Ainsi, les CRDTs partagent le comportement des types non-répliqués dans le cadre d'exécutions séquentielles. Cependant, ils définissent aussi une nouvelle sémantique pour chaque nouveau scénario ne pouvant se produire que dans le cadre d'une exécution distribuée.

Plusieurs sémantiques valides peuvent être proposées pour ces nouveaux scénarios. Un CRDT se doit donc de préciser quelle sémantique il choisit.

L'autre aspect définissant un CRDT donné est le modèle qu'il adopte pour propager les modifications. Au fil des années, la littérature a établi et défini plusieurs modèles dit de synchronisation, chacun ayant ses propres besoins et avantages. De fait, plusieurs CRDTs

peuvent être proposés pour un même type donné en fonction du modèle de synchronisation choisi.

Ainsi, ce qui définit un CRDT est sa sémantique en cas de concurrence et son modèle de synchronisation. Dans les prochaines sections, nous présenterons les différentes sémantiques possibles pour un type donné, l'Ensemble, en guise d'exemple. Puis nous présenterons les différents modèles de synchronisation proposés dans la littérature, et détaillerons leurs contraintes et impact sur les CRDT les adoptant, toujours en utilisant le même exemple.

Matthieu: TODO : Faire le lien avec les travaux de Burckhardt [19] et les MRDTs [20]

1.3.1 Sémantique en cas de concurrence

Plusieurs sémantiques peuvent être proposées pour résoudre les conflits. Certaines de ces sémantiques ont comme avantage d'être générique, c.-à-d. applicable à l'ensemble des types de données. En contrepartie, elles souffrent de cette même généralité, en ne permettant que des comportements simples en cas de conflits.

À l'inverse, la majorité des sémantiques proposées dans la littérature sont spécifiques à un type de données. Elles visent ainsi à prendre plus finement en compte l'intention des modifications pour proposer des comportements plus précis.

Dans la suite de cette section, nous présentons ces sémantiques génériques ainsi que celles spécifiques à l'Ensemble et, à titre d'exemple, les illustrons à l'aide du scénario présenté dans la Figure 1.1.

Sémantique *Last-Writer-Wins*

Une manière simple pour résoudre un conflit consiste à trancher de manière arbitraire et de sélectionner une modification parmi l'ensemble des modifications en conflit. Pour faire cela de manière déterministe, une approche est de reproduire et d'utiliser l'ordre total sur les modifications qui serait instauré par une horloge globale pour choisir la modification à prioriser.

Cette approche, présentée dans [21], correspond à la sémantique nommée *Last-Writer-Wins* (LWW). De par son fonctionnement, cette sémantique est générique et est donc utilisée par une variété de CRDTs pour des types différents. La Figure 1.2 illustre son application à l'Ensemble pour résoudre le conflit de la Figure 1.1.

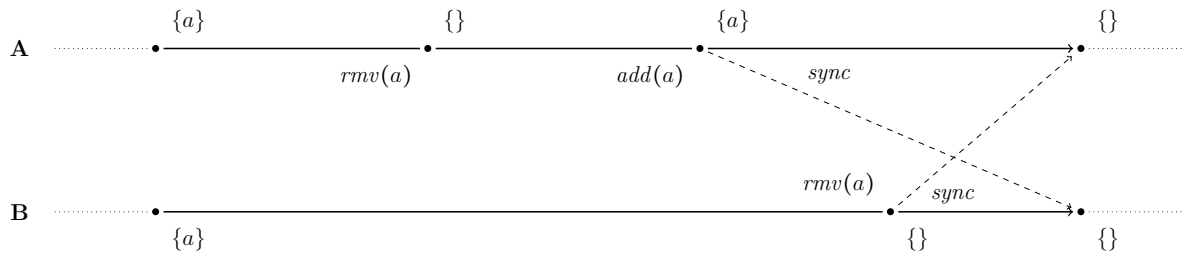


FIGURE 1.2 – Résolution du conflit en utilisant la sémantique LWW

Comme indiqué précédemment, le scénario illustré dans la Figure 1.2 présente un conflit entre les modifications concurrentes $add(a)$ et $rmv(a)$ générées de manière concurrente respectivement par les noeuds A et B. Pour le résoudre, la sémantique LWW associe à chaque modification une estampille. L'ordre créé entre les modifications par ces dernières permet de déterminer quelle modification désigner comme prioritaire. Ici, nous considérons que $add(a)$ a eu lieu plus tôt que $rmv(a)$. La sémantique LWW désigne donc $rmv(a)$ comme prioritaire et ignore $add(a)$. L'état obtenu à l'issue de cet exemple par chaque noeud est donc $\{\}$.

Il est à noter que si la modification $rmv(a)$ du noeud B avait eu lieu plus tôt dans notre exemple, l'état final obtenu aurait été $\{a\}$. Ainsi, des exécutions reproduisant le même ensemble de modifications produiront des résultats différents en fonction de l'ordre créé par les estampilles associées à chaque modification. Ces estampilles étant des métadonnées du mécanisme de résolution de conflits, elles sont dissimulées aux utilisateur-rices. Le comportement de cette sémantique peut donc être perçu comme aléatoire et s'avérer perturbant pour les utilisateur-rices.

La sémantique LWW repose sur l'horloge de chaque noeud pour attribuer une estampille à chacune de leurs modifications. Les horloges physiques étant sujettes à des imprécisions et notamment des décalages, utiliser les estampilles qu'elles fournissent peut provoquer des anomalies vis-à-vis de la relation *happens-before*. Les systèmes distribués préfèrent donc généralement utiliser des horloges logiques [15]. *Matthieu: TODO : Ajouter refs des horloges logiques plus intelligentes (Interval Tree Clock, Hybrid Clock...)*

Sémantique *Multi-Value*

Une seconde sémantique générique² est la sémantique *Multi-Value* (MV). Cette approche propose de gérer les conflits de la manière suivante : plutôt que de prioriser une modification par rapport aux autres modifications concurrentes, la sémantique MV maintient l'ensemble des états résultant possibles. Nous présentons son application à l'Ensemble dans la Figure 1.3.

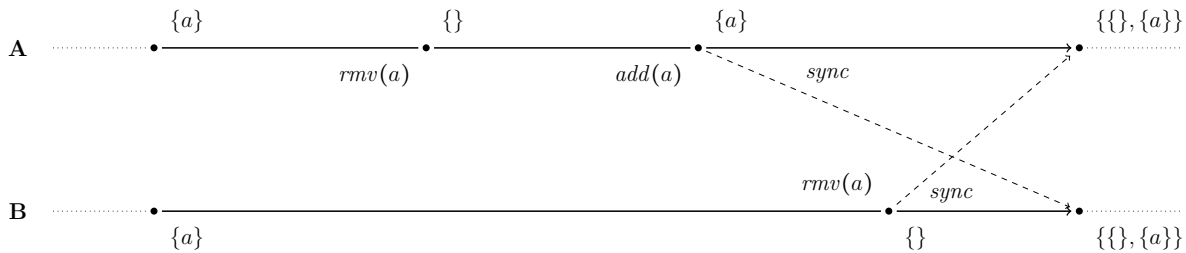


FIGURE 1.3 – Résolution du conflit en utilisant la sémantique MV

La Figure 1.3 présente la gestion du conflit entre les modifications concurrentes $add(a)$ et $rmv(a)$ par la sémantique MV. Devant ces modifications contraires, chaque noeud calcule chaque état possible, c.-à-d. un état sans l'élément a , $\{\}$, et un état avec ce dernier, $\{a\}$. Le CRDT maintient alors l'ensemble de ces états en parallèle. L'état obtenu est donc $\{\{\}, \{a\}\}$.

2. Bien que généralement associée au type *Registre* uniquement.

Ainsi, la sémantique MV expose les conflits aux utilisateur-rices lors de leur prochaine consultation de l'état du CRDT. Les utilisateur-rices peuvent alors prendre connaissance des intentions de chacun-e et résoudre le conflit manuellement. Dans la Figure 1.3, résoudre le conflit revient à re-effectuer une modification $add(a)$ ou $rmv(a)$ selon l'état choisi. Ainsi, si plusieurs personnes résolvent en concurrence le conflit de manière contraire, la sémantique MV exposera de nouveau les différents états proposés sous la forme d'un conflit.

Il est intéressant de noter que cette sémantique mène à un changement du domaine du CRDT considéré : en cas de conflit, la valeur retournée par le CRDT correspond à un Ensemble de valeurs du type initialement considéré. E.g. si nous considérons que le type correspondant au CRDT dans la Figure 1.3 est le type $Set\langle V \rangle$, nous observons que la valeur finale obtenue a pour type $Set\langle Set\langle V \rangle \rangle$. Il s'agit à notre connaissance de la seule sémantique opérant ce changement.

Sémantiques *Add-Wins* et *Remove-Wins*

Comme évoqué précédemment, d'autres sémantiques sont spécifiques au type de données concerné. Ainsi, nous abordons à présent des sémantiques spécifiques au type de l'Ensemble.

Dans le cadre de l'Ensemble, un conflit est provoqué lorsque des modifications add et rmv d'un même élément sont effectuées en concurrence. Ainsi, deux approches peuvent être proposées pour résoudre le conflit :

- (i) Une sémantique où la modification add d'un élément prend la précedence sur les modifications concurrentes rmv du même élément, nommée *Add-Wins* (AW). L'élément est alors présent dans l'état obtenu à l'issue de la résolution du conflit.
- (ii) Une sémantique où la modification rmv d'un élément prend la précedence sur les opérations concurrentes add du même élément, nommée *Remove-Wins* (RW). L'élément est alors absent de l'état obtenu à l'issue de la résolution du conflit.

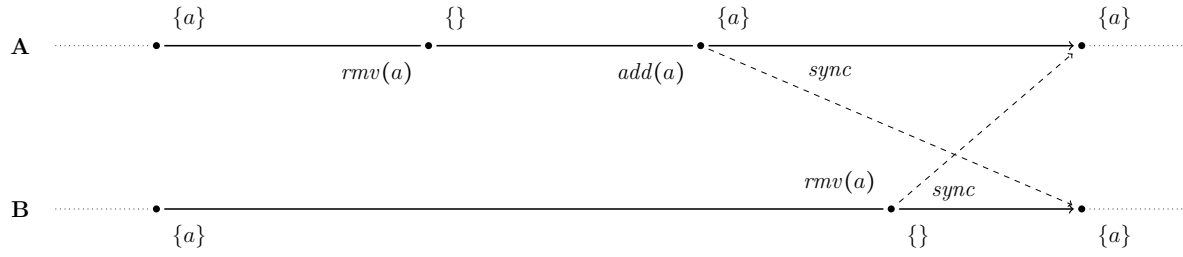
La Figure 1.4 illustre l'application de chacune de ces sémantiques sur notre exemple.

Sémantique *Causal-Length*

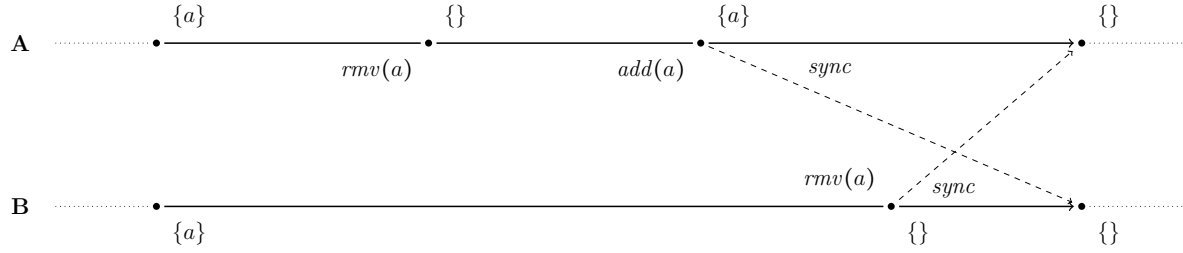
Une nouvelle sémantique pour l'Ensemble fut proposée [22] récemment. Cette sémantique se base sur les observations suivantes :

- (i) add et rmv d'un élément prennent place à tour de rôle, chaque modification invalidant la précédente.
- (ii) add (resp. rmv) concurrents d'un même élément représentent la même intention. Prendre en compte une de ces modifications concurrentes revient à prendre en compte leur ensemble.

À partir de ces observations, YU et ROSTAD proposent de déterminer pour chaque élément la chaîne d'ajouts et retraits la plus longue. C'est cette chaîne, et précisément son dernier maillon, qui indique si l'élément est présent ou non dans l'ensemble final. La Figure 1.5 illustre son fonctionnement.



(a) Application de la sémantique AW



(b) Application de la sémantique RW

FIGURE 1.4 – Résolution du conflit en utilisant soit la sémantique AW, soit la sémantique RW

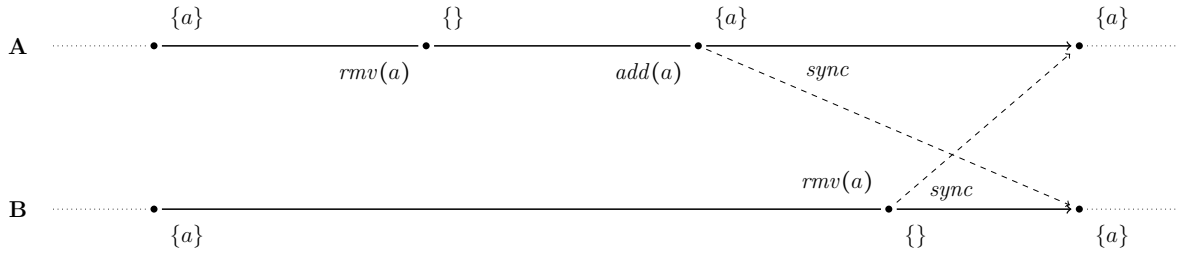


FIGURE 1.5 – Résolution du conflit en utilisant la sémantique CL

Dans notre exemple, la modification $rmv(a)$ effectuée par B est en concurrence avec une modification identique effectuée par A. La sémantique CL définit que ces deux modifications partagent la même intention. Ainsi, A ayant déjà appliqué sa propre modification préalablement, il ne prend pas en compte *de nouveau* cette modification lorsqu'il la reçoit de B. Son état reste donc inchangé.

À l'inverse, la modification $add(a)$ effectuée par A fait suite à sa modification $rmv(a)$. La sémantique CL définit alors qu'elle fait suite à toute autre modification $rmv(a)$ concurrente. Ainsi, B intègre cette modification lorsqu'il la reçoit de A. Son état évolue donc pour devenir $\{a\}$.

Synthèse

Dans cette section, nous avons mis en lumière l'existence de solutions différentes pour résoudre un même conflit. Chacune de ces solutions correspond à une sémantique spécifique de résolution de conflits. Ainsi, pour un même type de données, différents CRDTs

peuvent être spécifiés. Chacun de ces CRDTs est spécifié par la combinaison de sémantiques qu'il adopte, chaque sémantique servant à résoudre un des types de conflits du type de données.

Il est à noter qu'aucune sémantique n'est intrinsèquement meilleure et préférable aux autres. Il revient aux concepteur-rices d'applications de choisir les CRDTs adaptés en fonction des besoins et des comportements attendus en cas de conflits.

Par exemple, pour une application collaborative de listes de courses, l'utilisation d'un MV-Registre pour représenter le contenu de la liste se justifie : cette sémantique permet d'exposer les modifications concurrentes aux utilisateur-rices. Ainsi, les personnes peuvent détecter et résoudre les conflits provoqués par ces éditions concurrentes, e.g. l'ajout de l'élément *lait* à la liste, pour cuisiner des crêpes, tandis que les *oeufs* nécessaires à ces mêmes crêpes sont retirés. En parallèle, cette même application peut utiliser un LWW-Registre pour représenter et indiquer aux utilisateur-rices la date de la dernière modification effectuée.

1.3.2 Modèle de synchronisation

Dans le modèle de réplication optimiste, les noeuds divergent momentanément lorsqu'ils effectuent des modifications locales. Pour ensuite converger vers des états équivalents, les noeuds doivent propager et intégrer l'ensemble des modifications. La Figure 1.6 illustre ce point.

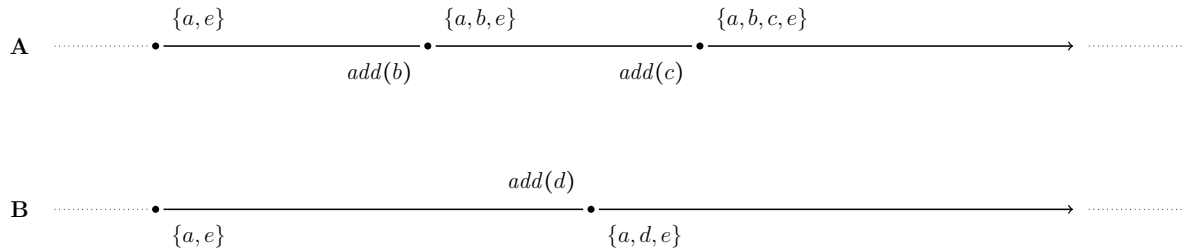


FIGURE 1.6 – Modifications en concurrence d'un Ensemble répliqué par les noeuds A et B

Dans cet exemple, deux noeuds A et B partagent et éditent un même Ensemble à l'aide d'un CRDT. Les deux noeuds possèdent le même état initial : $\{a, e\}$.

Le noeud A effectue les modifications $add(b)$ puis $add(c)$. Il obtient ainsi l'état $\{a, b, c, e\}$. De son côté, le noeud B effectue la modification suivante : $add(d)$. Son état devient donc $\{a, d, e\}$. Ainsi, les noeuds doivent encore s'échanger leur modifications pour converger vers l'état souhaité³, c.-à-d. $\{a, b, c, d, e\}$.

Dans le cadre des CRDTs, le choix de la méthode pour synchroniser les noeuds n'est pas anodin. En effet, ce choix impacte la spécification même du CRDT et ses prérequis.

Initialement, deux approches ont été proposées : une méthode de synchronisation par états [16, 23] et une méthode de synchronisation par opérations [16, 23, 24, 25]. Une

3. Le scénario ne comportant uniquement des modifications add , aucun conflit n'est produit malgré la concurrence des modifications.

troisième approche, nommée synchronisation par différence d'états [26, 27], fut spécifiée par la suite. Le but de cette dernière est d'allier le meilleur des deux approches précédentes.

Dans la suite de cette section, nous présentons ces approches ainsi que leurs caractéristiques respectives. Pour les illustrer, nous complétons l'exemple décrit ici. Cependant, nous nous focalisons uniquement sur les messages envoyés par les noeuds et n'évoquons seulement les métadonnées introduites par chaque modèle de synchronisation, par soucis de clarté et de simplicité.

Synchronisation par états

L'approche de la synchronisation par états propose que les noeuds diffusent leurs modifications en transmettant leur état. Les CRDTs adoptant cette approche doivent définir une fonction `merge`. Cette fonction correspond à la fonction de fusion mentionnée précédemment en (iii) : elle prend en paramètres une paire d'états et génère en retour l'état correspondant à leur LUB. Cette fonction doit être associative, commutative et idempotente.

Ainsi, lorsqu'un noeud reçoit l'état d'un autre noeud, il fusionne ce dernier avec son état courant à l'aide de la fonction `merge`. Il obtient alors un nouvel état intégrant l'ensemble des modifications ayant été effectuées sur les deux états.

La nature croissante des états des CRDTs couplée aux propriétés d'associativité, de commutativité et d'idempotence de la fonction `merge` permettent de reposer sur la couche réseau sans lui imposer de contraintes fortes : les messages peuvent être perdus, réordonnés ou même dupliqués. Les noeuds convergeront tant que la couche réseau garantit que les noeuds seront capables de transmettre leur état aux autres à terme. Il s'agit là de la principale force des CRDTs synchronisés par états.

Néanmoins, la définition de la fonction `merge` offrant ces propriétés peut s'avérer complexe et a des répercussions sur la spécification même du CRDT. Notamment, les états doivent conserver une trace de l'existence des éléments et de leur suppression afin d'éviter qu'une fusion d'états ne les fassent ressurgir. Ainsi, les CRDTs synchronisés par états utilisent régulièrement des pierres tombales.

En plus de l'utilisation de pierres tombales, la taille de l'état peut croître de manière non-bornée dans le cas de certains types de données, e.g. l'Ensemble ou la Séquence. Ainsi, ces structures peuvent atteindre à terme des tailles conséquentes. Dans de tels cas, diffuser l'état complet à chaque modification induirait alors un coût rédhibitoire. L'approche de la synchronisation par états s'avère donc inadaptée aux systèmes temps réel et repose généralement sur une synchronisation périodique.

Nous illustrons le fonctionnement de cette approche avec la Figure 1.7. Dans cet exemple, après que les noeuds aient effectués leurs modifications respectives, le mécanisme de synchronisation périodique de chaque noeud se déclenche. Le noeud A (resp. B) diffuse alors son état $\{a, b, c, e\}$ (resp. $\{a, d, e\}$) à B (resp. A).

À la réception de l'état, chaque noeud utilise la fonction `merge` pour intégrer les modifications de l'état reçu dans son propre état. Dans le cadre de l'Ensemble répliqué, cette fonction consiste généralement à faire l'union des états, en prenant en compte l'estampille et le statut (présent ou non) associé à chaque élément. Ainsi la fusion de leur état respectif, $\{a, b, c, e\} \cup \{a, d, e\}$, permet aux noeuds de converger à l'état souhaité : $\{a, b, c, d, e\}$.

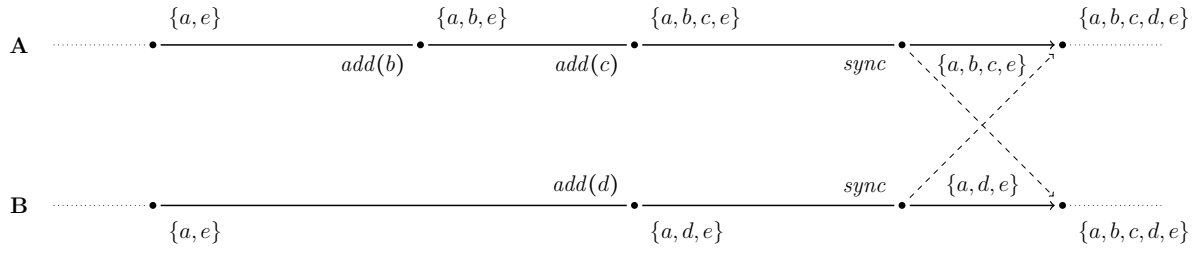


FIGURE 1.7 – Synchronisation des noeuds A et B en adoptant le modèle de synchronisation par états

Avant de conclure, il est intéressant de noter que ce modèle de synchronisation propose par nature une diffusion des modifications suivant le modèle de cohérence causale *Mathieu: TODO : Ajouter ref.* En effet, ce modèle de synchronisation assure une livraison soit de toutes les modifications connues d'un noeud, soit d'aucune. Par exemple, dans la Figure 1.7, le noeud B ne peut pas recevoir et intégrer l'élément c sans l'élément b . Ainsi, ce modèle permet naturellement d'éviter ce qui pourrait être interprétées comme des anomalies par les utilisateur-rices.

Synchronisation par opérations

L'approche de la synchronisation par opérations propose quant à elle que les noeuds diffusent leurs modifications sous la forme d'opérations. Pour chaque modification possible, les CRDTs synchronisés par opérations doivent définir deux fonctions : **prepare** et **effect**.

La fonction **prepare** a pour but de générer une opération correspondant à la modification effectuée, et commutative avec les potentielles opérations concurrentes. Elle prend en paramètres la modification ainsi que ses paramètres, et l'état courant du noeud. Cette fonction n'a pas d'effet de bord, c.-à-d. ne modifie pas l'état courant, et génère en retour l'opération à diffuser à l'ensemble des noeuds.

Une opération est un message. Son rôle est d'encoder la modification sous la forme d'un ou plusieurs éléments irréductibles du sup-demi-treillis.

Définition 3 (Élément irréductible) *Un élément irréductible d'un sup-demi-treillis est un élément atomique de ce dernier. Il ne peut être obtenu par la fusion d'autres états.*

Il est à noter que dans le cas des CRDTs purs synchronisés par opérations [25], les modifications labellisées avec leur information de causalité correspondent à des éléments irréductibles, c.-à-d. à des opérations. La fonction **prepare** peut donc être omise pour cette sous-catégorie de CRDTs synchronisés par opérations.

La fonction **effect** permet quant à elle d'intégrer les effets d'une opération générée ou reçue. Elle prend en paramètre l'état courant et l'opération, et retourne un nouvel état. Ce nouvel état correspond à la LUB entre l'état courant et le ou les éléments irréductibles encodés par l'opération.

La diffusion des modifications par le biais d'opérations présentent plusieurs avantages. Tout d'abord, la taille des opérations est généralement fixe et inférieure à la taille de

l'état complet du CRDT, puisque les opérations servent à encoder un de ses éléments irréductibles. Ensuite, l'expressivité des opérations permet de proposer plus simplement des algorithmes efficaces pour leur intégration par rapport aux modifications équivalentes dans les CRDTs synchronisés par états. Par exemple, la suppression d'un élément dans un Ensemble se traduit en une opération de manière presque littérale, tandis que pour les CRDTs synchronisés par états, c'est l'absence de l'élément dans l'état qui va rendre compte de la suppression effectuée. Ces avantages rendent possible la diffusion et l'intégration une à une des modifications et rendent ainsi utilisables les CRDTs synchronisés par opérations pour construire des systèmes temps réels.

Il est à noter que la seule contrainte imposée aux CRDTs synchronisés par opérations est que leurs opérations concurrentes soient commutatives. Ainsi, il n'existe aucune contrainte sur la commutativité des opérations liées causalement. De la même manière, aucune contrainte n'est définie sur l'idempotence des opérations. Ces libertés impliquent qu'il peut être nécessaire que les opérations soient délivrées au CRDT en respectant un ordre donné et en garantissant leur livraison en exactement une fois pour garantir la convergence. Ainsi, un intergiciel chargé de la diffusion et de la livraison des opérations est usuellement associé aux CRDTs synchronisés par opérations pour respecter ces contraintes.

Généralement, les CRDTs synchronisés par opérations sont présentés dans la littérature comme nécessitant une livraison causale des opérations. Ce modèle de livraison permet de respecter le modèle de cohérence causale et ainsi de simplifier raisonnement sur exécutions.

Ce modèle introduit néanmoins plusieurs effets négatifs. Tout d'abord, ce modèle peut provoquer un délai dans la diffusion des modifications. En effet, la perte d'une opération par le réseau provoque la mise en attente de la livraison des opérations suivantes. Les opérations mises en attente ne sont délivrées qu'une fois l'opération perdue re-diffusée et délivrée.

De plus, il nécessite que des informations de causalité précises soient attachées à chaque opération. Pour cela, les systèmes reposent généralement sur l'utilisation de vecteurs de version *Matthieu: TODO : Insérer ref.* Or, la taille de cette structure de données croît de manière linéaire avec le nombre de noeuds du système. Les métadonnées de causalité peuvent ainsi représenter la majorité des données diffusées sur le réseau⁴. *Matthieu: TODO : Ajouter mention que OT a été abandonné à cause de cette contrainte même.* Cependant, nous observons que la livraison dans l'ordre causal de toutes les opérations n'est pas toujours nécessaire. Par exemple, l'ordre d'intégration de deux opérations d'ajout d'éléments différents dans un Ensemble n'a pas d'importance. Nous pouvons alors nous affranchir de la livraison dans l'ordre causal pour accélérer la vitesse d'intégration des modifications et pour réduire les métadonnées envoyées.

Une autre contrainte généralement associée aux CRDTs synchronisés par opérations est la nécessité d'une livraison en exactement un exemplaire de chaque opération. Cette contrainte dérive de :

4. La relation de causalité étant transitive, les opérations et leurs relations de causalité forment un DAG. [28] propose d'ajouter en dépendances causales d'une opération seulement les opérations correspondant aux extrémités du DAG au moment de sa génération. Ce mécanisme plus complexe permet de réduire la consommation réseau, mais induit un surcoût en calculs et en mémoire utilisée.

- (i) La potentielle non-idempotence des opérations.
- (ii) La nécessité de la livraison de chaque opération pour la livraison causale.

Toutefois, nous observons que des opérations peuvent être sémantiquement rendues obsolètes par d'autres opérations, e.g. une opération d'ajout d'un élément dans un Ensemble est rendue obsolète par une opération de suppression ultérieure du même élément. Ainsi, l'intergiciel de livraison peut se contenter d'assurer une livraison en un exemplaire au plus des opérations non-obsolètes. Ce choix permet de réduire la consommation réseau en évitant la diffusion d'opérations désormais non-pertinentes.

Pour compenser la perte d'opérations par le réseau et ainsi garantir la livraison à terme des opérations pertinentes, l'intergiciel de livraison des opérations doit mettre en place un mécanisme d'anti-entropie. Plusieurs mécanismes de ce type ont été proposés dans la littérature [29, 30, 31, 32] *Matthieu: TODO : Ajouter refs Scuttlebutt si applicable à Op-based* et proposent des compromis variés entre complexité en temps, complexité spatiale et consommation réseau.

Nous illustrons le modèle de synchronisation par opérations à l'aide de la Figure 1.8. Dans ce nouvel exemple, les noeuds diffusent les modifications qu'ils effectuent sous la forme d'opérations. Nous considérons que le CRDT utilisé est un CRDT pur synchronisé par opérations, c.-à-d. que les modifications et opérations sont confondues.

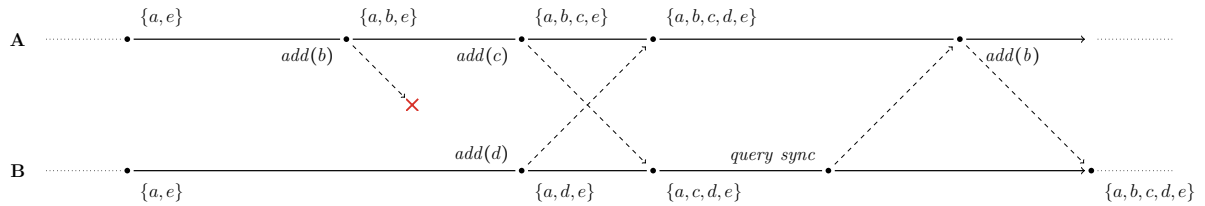


FIGURE 1.8 – Synchronisation des noeuds A et B en adoptant le modèle de synchronisation par opérations

Le noeud A diffuse donc les opérations $add(b)$ et $add(c)$. Il reçoit ensuite l'opération $add(d)$ de B, qu'il intègre à sa copie. Il obtient alors l'état $\{a, b, c, d, e\}$.

De son côté, le noeud B ne reçoit initialement pas l'opération $add(b)$ suite à une défaillance réseau. Il génère et diffuse $add(d)$ puis reçoit l'opération $add(c)$. Nous considérons que la livraisons des opérations add n'est pas obligatoire dans cet exemple, cette opération est alors intégrée. Le noeud B obtient alors l'état $\{a, c, d, e\}$.

Ensuite, le mécanisme d'anti-entropie du noeud B se déclenche. Le noeud B envoie alors à A une demande de synchronisation contenant un résumé de son état, e.g. son vecteur de version. À partir de cette donnée, le noeud A détermine que B n'a pas reçu l'opération $add(a)$. Il génère alors une réponse contenant cette opération et lui envoie. À la réception de l'opération, le noeud B l'intègre. Il obtient l'état $\{a, b, c, d, e\}$ et converge ainsi avec A.

Avant de conclure, nous noterons qu'il est nécessaire pour les noeuds de maintenir leur journal d'opérations. En effet, les noeuds l'utilisent pour renvoyer les opérations manquées lors de l'exécution du mécanisme d'anti-entropie évoqué ci-dessus. Ceci se traduit par une augmentation perpétuelle des métadonnées des CRDTs synchronisés par opérations. Pour

y pallier, des travaux [25, 33] proposent de tronquer le journal des opérations pour en supprimer les opérations connues de tous. Les noeuds reposent alors sur la notion de stabilité causale pour déterminer les opérations supprimables de manière sûre.

Définition 4 (Stabilité causale) *Une opération est stable causalement lorsqu'elle a été observée par l'ensemble des noeuds du système. Ainsi, toute opération future dépend causalement des opérations causalement stables, c.-à-d. les noeuds ne peuvent plus générer d'opérations concurrentes aux opérations causalement stables de manière honnête.*

Un mécanisme d'instantané *Matthieu: TODO : Ajouter refs* doit néanmoins être associé au mécanisme de troncage du journal pour générer un état équivalent à celui résultant des opérations supprimées. Ce mécanisme est en effet nécessaire pour permettre un nouveau noeud de rejoindre le système et d'obtenir l'état courant à partir de l'instantané et du journal tronqué.

Pour résumer, cette approche permet de mettre en place simplement un système fonctionnel à l'aide d'un CRDT synchronisé par opérations et d'un intergiciel de diffusion et de livraison RCB. Mais comme illustré ci-dessus, chaque CRDT synchronisé par opérations établit les propriétés de ses différentes opérations et délègue potentiellement des responsabilités à l'intergiciel de diffusion et livraison. La complexité de cette approche réside ainsi dans l'ajustement du couple $\langle \text{CRDT}, \text{intergiciel} \rangle$ pour régler finement et optimiser leur fonctionnement en tandem. Des approches [25, 33] ont été proposées ces dernières années pour concevoir et structurer plus proprement ces composants et leurs relations, mais reposent encore sur une livraison causale des opérations. *Matthieu: TODO : Vérifier que c'est bien le cas dans [33]*

Synchronisation par différences d'états

Dans [26], ALMEIDA, SHOKER et BAQUERO introduisent un nouveau modèle de synchronisation pour CRDTs. La proposition de ce modèle est nourrie par les observations suivantes :

- (i) Les CRDTs synchronisés par opérations sont vulnérables aux défaillances du réseau et nécessitent généralement pour pallier à cette vulnérabilité une livraison des opérations en exactement un exemplaire et respectant l'ordre causal.
- (ii) Les CRDTs synchronisés par états pâtissent du surcoût induit par la diffusion de leurs états complets, généralement croissant continuellement.

Pour pallier aux faiblesses de chaque approche et allier le meilleur des deux mondes, les auteurs proposent les CRDTs synchronisés par différences d'états [26, 27, 34]. Il s'agit en fait d'une sous-famille des CRDTs synchronisés par états. Ainsi, comme ces derniers, ils disposent d'une fonction `merge` permettant de produire la LUB entre deux états, et qui est associative, commutative et idempotente.

La spécificité des CRDTs synchronisés par différences d'états est qu'une modification locale produit en retour un delta. Un delta encode la modification effectuée sous la forme d'un état du lattice. Les deltas étant des états, ils peuvent être diffusés puis intégrés par les autres noeuds à l'aide de la fonction `merge`. Ceci permet de bénéficier des propriétés

d'associativité, de commutativité et d'idempotence offertes par cette fonction. Les CRDTs synchronisés par différences d'états offrent ainsi :

- (i) Une diffusion des modifications avec un surcoût pour le réseau proche de celui des CRDTs synchronisés par opérations.
- (ii) Une résistance aux défaillances réseaux similaire celle des CRDTs synchronisés par états.

Cette définition des CRDTs synchronisés par différences d'états, introduite dans [26, 27], fut ensuite précisée dans [34]. Dans cet article, ENES et al. précisent qu'utiliser des éléments irréductibles (cf. Définition 3) comme deltas est optimal du point de vue de la taille des deltas produits.

Concernant la diffusion des modifications, les CRDTs synchronisés par différences d'états autorisent un large éventail de possibilités. Par exemple, les deltas peuvent être diffusés et intégrés de manière indépendante. Une autre approche possible consiste à tirer avantage du fait que les deltas sont des états : il est possible d'agréger plusieurs deltas à l'aide la fonction `merge`, éliminant leurs éventuelles redondances. Ainsi, la fusion de deltas permet ensuite de diffuser un ensemble de modifications par le biais d'un seul et unique delta, minimal. Et en dernier recours, les CRDTs synchronisés par différences d'états peuvent adopter le même schéma de diffusion que les CRDTs synchronisés par états, c.-à-d. diffuser leur état complet de manière périodique. Plusieurs de ces approches sont évaluées et comparées de manière expérimentale dans [34].

Nous illustrons cette approche avec la Figure 1.9. Dans cet exemple, nous considérons que les noeuds adoptent la seconde approche évoquée, c.-à-d. que périodiquement les noeuds agrègent les deltas issus de leurs modifications et diffusent le delta résultant.

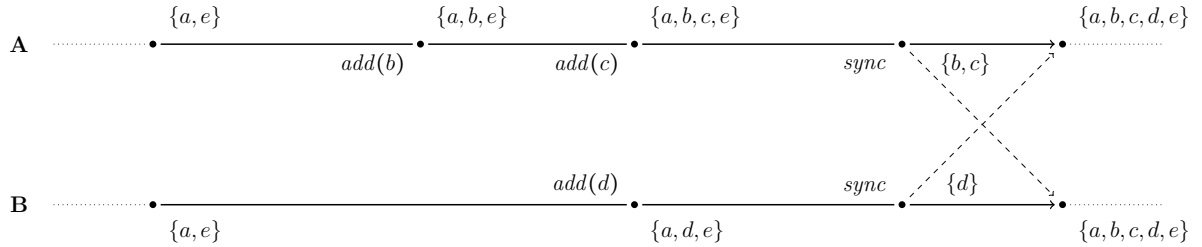


FIGURE 1.9 – Synchronisation des noeuds A et B en adoptant le modèle de synchronisation par différences d'états

Le noeud A effectue les modifications $add(b)$ et $add(c)$, qui retournent respectivement les deltas $\{b\}$ et $\{c\}$. Le noeud A agrège ces deltas et diffuse donc le delta suivant b, c . Quant au noeud B, il effectue la modification $add(d)$ qui produit le delta $\{d\}$. S'agissant de son unique modification, il diffuse ce delta inchangé.

Quand A (resp. B) reçoit le delta $\{d\}$ (resp. $\{b, c\}$), il l'intègre à sa copie en utilisant la fonction `merge`. Les deux noeuds convergent alors à l'état $\{a, b, c, d, e\}$.

La synchronisation par différences d'états permet donc de réduire la taille des messages diffusés sur le réseau par rapport à la synchronisation par états. Cependant, il est important de noter que la décomposition en deltas entraîne la perte d'une des propriétés

intrinsèques des CRDTs synchronisés par états : le respect du modèle de cohérence causale. En effet, sans mécanisme supplémentaire, la perte ou le ré-ordonnement de deltas par le réseau peut mener à une livraison dans le désordre des modifications à l'un des noeuds. S'ils souhaitent assurer une diffusion causale des modifications, les CRDTs synchronisés par différences d'états doivent donc définir et intégrer dans leur spécification un mécanisme similaire à l'intergiciel de livraison des CRDTs synchronisés par opérations. En revanche et à l'instar des CRDTs synchronisés par opérations, les CRDTs synchronisés par différences d'états peuvent aussi faire le choix inverse : s'affranchir du modèle de cohérence causale pour accélérer la diffusion des modifications et minimiser le surcoût du type répliqué.

Ainsi, les CRDTs synchronisés par différences d'états sont une évolution prometteuse des CRDTs synchronisés par états. Ce modèle de synchronisation rend ces CRDTs utilisables dans les systèmes temps réels sans introduire de contraintes sur la fiabilité du réseau. Mais pour cela, il ajoute une couche supplémentaire de complexité à la spécification des CRDTs synchronisés par états, c.-à-d. le mécanisme dédié à la livraison des deltas, qui peut s'avérer source d'erreurs et de coûts supplémentaires.

Synthèse

Nous récapitulons les principales propriétés et différences des modèles de synchronisations pour CRDTs dans Tableau 1.1.

TABLE 1.1 – Récapitulatif comparatif des différents modèles de synchronisation pour CRDTs

	State-based	Op-based	Delta-based
États forment sup-demi-treillis	✓	✓	✓
Intègre modifications par fusion d'états	✓	✗	✓
Intègre modifications de manière atomique	✗	✓	✓
Résistant aux défaillances réseau	✓	✗	✓
Peut s'affranchir de la cohérence causale	✗	✓	✓
Adapté pour systèmes temps réel	✗	✓	✓

1.3.3 Adoption dans la littérature et l'industrie

- Proposition et conception de CRDTs pour une variété de types de données : Registre, Compteur *Matthieu: TODO : Ajouter IPA*, Ensemble, Liste/Sequence, Graphe, JSON, Filesystem, Access Control. Propose généralement plusieurs sémantiques de résolution de conflits par type de données.
- Conception et développement de bibliothèques mettant à disposition des développeurs d'applications des types de données composés [35, 36, 37, 38, 39] *Matthieu: TODO : Revoir et ajouter Melda (PaPoC'22) si fitting*
- Conception de langages de programmation intégrant des CRDTs comme types primitifs, destinés au développement d'applications distribuées [40, 41]

- Conception et implémentation de bases de données distribuées, relationnelles ou non, privilégiant la disponibilité et la minimisation de la latence à l’aide des CRDTs [42, 43, 44, 45, 46] *Matthieu: TODO : Ajouter Redis et Akka*
- Conception d’un nouveau paradigme d’applications, Local-First Software, dont une des fondations est les CRDTs [47, 48] *Matthieu: TODO : Vérifier et ajouter l’article avec Digital Garden (PaPoC’22 ?) si fitting*
- Éditeurs collaboratifs temps réel à large échelle et offrant de nouveaux scénarios de collaboration grâce aux CRDTs [49, 3]

1.4 Séquences répliquées sans conflits

- Présentation du type Sequence et de son utilisation
- Spécification du type : données, modifications
- Présentation de conflits
 - dOpt puzzle
- Introduction des principales approches pour Sequence CRDTs. Préciser que, sauf mention contraire, ces CRDTs adoptent le modèle de synchronisation par opérations en raison de la taille non-bornée de l’état.

1.4.1 Approche à pierres tombales

WOOT

- WOOT [50] fait suite aux travaux présentés dans [51]. Il est considéré a posteriori comme le premier CRDT synchronisé par opérations pour Séquence⁵.
- Conçu pour l’édition collaborative Pair-à-Pair (P2P), son but est de surmonter les limites de l’approche Operational Transformation (OT). En effet, l’approche OT souffre de la nécessité d’associer à chaque opération des informations de causalité, c.-à-d. un vecteur d’horloges. La taille de cette structure de données augmentant avec le nombre de noeuds du système, cette approche se révèle inadaptée aux systèmes P2P dynamique, c.-à-d. systèmes dans lesquels un grand nombre de noeuds peuvent rejoindre puis quitter la collaboration.
- L’intuition de WOOT est la suivante : WOOT modifie la sémantique de la modification *ins* pour qu’elle corresponde à l’insertion d’un nouvel élément entre deux autres, et non plus à l’insertion d’un nouvel élément à une position donnée. Ce changement, qui est compatible avec l’intention des utilisateur-rices, n’est cependant pas anodin. En effet, il permet à WOOT de rendre *ins* commutative avec les modifications concurrentes. en exprimant la position du nouvel élément de manière relative à d’autres éléments et non plus via un index qui est spécifique à un état donné.

5. [52] n’ayant formalisé les CRDTs qu’en 2007.

- Afin de préciser quels éléments correspondent aux prédécesseur et successeur de l'élément inséré, WOOT repose sur un système d'identifiants. WOOT associe ainsi un identifiant unique à chaque élément de la Séquence. La modification *rmv* utilise aussi les identifiants pour indiquer l'élément à supprimer.
- WOOT utilise des pierres tombales pour rendre *ins*, qui nécessite la présence des deux éléments entre lesquels nous insérons un nouvel élément, commutative avec *rmv*. Ainsi, une pierre tombale est conservée est conservée dans la Séquence pour indiquer sa présence passée, mais les données de l'élément sont supprimées. Dans le cadre d'une Séquence WOOT de caractères, *rmv* a donc pour effet de masquer l'élément.
- Finalement, WOOT définit $<_{id}$, un ordre strict total sur les identifiants associés aux éléments. En effet, il convient de noter que la relation $<$ ne spécifie qu'un ordre partiel entre les éléments. Ainsi, $<$ ne permet pas d'ordonner les éléments insérés en concurrence et possédant les mêmes prédécesseur et successeur, e.g. $ins(a < 1 < b)$ et $ins(a < 2 < b)$. Pour que tous les noeuds convergent, ils doivent choisir comment ordonner ces éléments de manière déterministe et indépendante de l'ordre de réception des modifications. Ils utilisent pour cela $<_{id}$.
- Ainsi WOOT offre une spécification de la Séquence dont les opérations sont commutatives, c.-à-d. ne génèrent pas de conflits. Nous illustrons son fonctionnement à l'aide de la Figure 1.10.

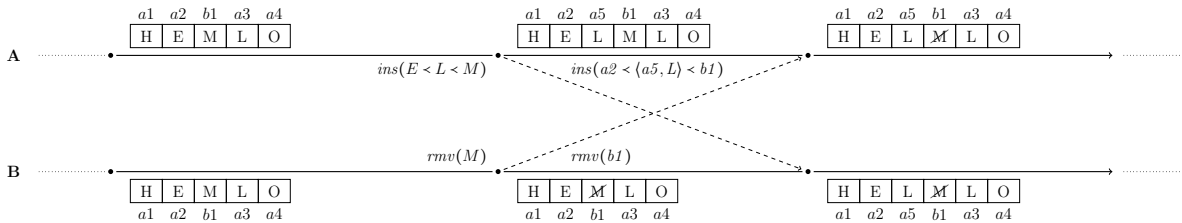


FIGURE 1.10 – Modifications concurrentes d'une séquence répliquée WOOT

Dans cet exemple, deux noeuds A et B partagent et éditent collaborativement une séquence répliquée WOOT. Initialement, ils possèdent le même état : la séquence contient les éléments "HEMLO", et à chaque élément est associé un identifiant, e.g. $a1$, $b1$, $a2$...

- Le noeud A insère l'élément "L" entre les éléments "E" et "M", c.-à-d. $ins(E < L < M)$. WOOT convertit cette modification en opération $ins(a2 < \langle a5, L \rangle < b1)$. L'opération est intégrée à la copie locale, ce qui produit l'état "HELMLO", puis diffusée sur le réseau.
- En concurrence, le noeud B supprime l'élément "M" de la séquence, c.-à-d. $rmv(M)$. De la même manière, WOOT génère l'opération correspondante $rmv(b1)$. Comme expliqué précédemment, l'intégration de cette opération ne supprime pas l'élément "M" de l'état mais se contente de le masquer. L'état produit est donc "HEMLO". L'opération est ensuite diffusée.

- A (resp. B) reçoit ensuite l'opération de B, $rmv(b1)$ (resp. A, $ins(a2 < \langle a5, L \rangle < b1)$), et l'intègre à sa copie. Les opérations de WOOT étant commutatives, les noeuds obtiennent le même état final : "HELMLO".
- Grâce à la commutativité de ses opérations, WOOT s'affranchit du modèle de livraison causale nécessitant l'utilisation coûteuse de vecteurs d'horloges. WOOT met en place un modèle de livraison sur-mesure basé sur les pré-conditions des opérations :
 - (i) L'opération $ins(predId < \langle id, elt \rangle < succId)$ ne peut être délivrée qu'après la livraison des opérations d'insertion des éléments associés à $predId$ et $succId$.
 - (ii) L'opération $rmv(id)$ ne peut être délivrée qu'après la livraison de l'opération d'insertion de l'élément associé à id .

Ce modèle de livraison ne requiert qu'une quantité fixe de métadonnées associées à chaque opération pour être respecté. WOOT est donc adapté aux systèmes P2P dynamiques.

- *Matthieu: TODO : Mentionner que WOOT ne propose pas de mécanisme pour purger les tombstones, par choix : souhaite mettre en place des fonctionnalités d'undo.*
- Des améliorations de WOOT furent par la suite proposées : WOOTO [53] et WOOTH [54]. *Matthieu: TODO : Ajouter limite de WOOT : complexité en temps de l'algorithme d'intégration des insertions distantes trop importante* Dans [53], WEISS, URSO et MOLLI remanient la structure des identifiants associés aux éléments. Cette modification permet un algorithme d'intégration des opérations ins plus efficace. Dans [54], AHMED-NACER et al. se basent sur WOOTO et proposent l'utilisation de structures de données améliorant la complexité des algorithmes d'intégration des opérations, au détriment des métadonnées stockées localement par chaque noeud.
- Néanmoins, l'évaluation expérimentale des différentes approches pour l'édition collaborative P2P en temps réel menée dans [54] a montré que les CRDTs de la famille WOOT n'étaient pas assez efficaces. Dans le cadre de cette expérience, des utilisateur-rices effectuaient des tâches d'édition collaborative données. Les traces de ces sessions d'édition collaboratives furent ensuite rejouées en utilisant divers mécanismes de résolution de conflits, dont WOOT, WOOTO et WOOTH. Le but était de mesurer les performances de ces mécanismes, notamment leurs temps d'intégration des modifications et opérations. Dans le cas de la famille WOOT, AHMED-NACER et al. ont constaté que ces temps dépassaient parfois 50ms. Il s'agit là de la limite des délais acceptables par les utilisateur-rices d'après [55, 56]. Ces performances disqualifient donc les CRDTs de la famille WOOT comme approches viables pour l'édition collaborative P2P temps réel.

Replicated Growable Array

- Replicated Growable Array (RGA) [57] est le second CRDT appartenant à l'approche à pierres tombales. Spécifié dans le cadre d'un effort pour établir les principes nécessaires à la conception de Replicated Abstract Data Types (RADTs).
- Dans cet article, définissent et se basent sur 2 principes pour concevoir RADTs. Le premier d'entre eux est Operation Commutativity (OC). Celui-ci indique que toutes

paires possibles d'opérations concurrentes doit être commutative. Permet de garantir que l'intégration par des noeuds différents d'une même séquence d'opérations concurrentes, mais dans des ordres différents, produira un état équivalent.

- Le second principe sur lequel repose les RADTs est la Precedence Transitivity (PT). Définit une relation de précédence, \rightarrow , qui permet d'établir un ordre de priorité sur les intentions des opérations, pour déterminer laquelle conserver en cas de conflits. Precedence Transitivity définit que pour tout trio d'opérations a , b et c , si on a $a \rightarrow b$ et $b \rightarrow c$, on a $a \rightarrow c$.
- À partir de ces principes, proposent plusieurs RADTs : Replicated Fixed-Size Array (RFA), Replicated Hash Table (RFT) et Replicated Growable Array (RGA), qui nous intéresse ici.
- Dans RGA, l'intention de l'insertion est définie comme l'insertion d'un nouvel élément directement après un élément existant. Ainsi, RGA se base sur le prédecesseur d'un élément pour déterminer où l'insérer. De fait, tout comme WOOT, RGA repose sur un système d'identifiants qu'il associe aux éléments pour pouvoir s'y référer par la suite. Nommé S4Vector, ces identifiants sont de la forme suivante $\langle ssid, sum, ssn, seq \rangle$ avec :
 - $ssid$, identifiant de la session de collaboration
 - sum , somme du vecteur d'horloges courant du noeud auteur de l'élément à son insertion
 - ssn , identifiant du noeud auteur de l'élément
 - seq , numéro de séquence de l'auteur de l'élément à son insertion

L'insertion est donc définie de la manière suivante : $ins(predId < \langle id, elt \rangle)$. Dans présentations suivantes de RGA, utilisent horloge de Lamport [15] plutôt. Nous abstrayons donc ici la structure des identifiants avec le symbole t .

- Puisque plusieurs éléments peuvent être insérés en concurrence à la même position, c.-à-d. avec le même prédecesseur, il est nécessaire de définir une relation d'ordre strict total pour ordonner les éléments de manière déterministe et indépendante de l'ordre de réception des modifications. Pour cela, utilise l'ordre lexicographique sur les composants des identifiants des éléments. Par exemple, dans le cadre des S4Vector, on a $\langle 1, 10, A, 2 \rangle < \langle 2, 5, B, 1 \rangle$ ou $\langle 1, 10, A, 2 \rangle < \langle 1, 15, B, 1 \rangle$. L'utilisation de cet ordre comme stratégie de résolution de conflits permet de rendre commutative les modifications ins concurrentes.
- Concernant les suppressions, RGA se comporte de manière similaire à WOOT : la séquence conserve une tombstone pour chaque élément supprimé, de façon à pouvoir insérer à la bonne position un élément dont le prédecesseur a été supprimé en concurrence. Cette stratégie rend commutative les modifications ins et rmv .
- Nous récapitulons le fonctionnement de RGA à l'aide de la Figure 1.11. *Matthieu: TODO : Trouver exemple et le présenter.*
- À la différence des auteurs de WOOT, ROH et al. jugent le coût des tombstones trop élevé. Ils proposent alors un mécanisme de Garbage Collection (GC) des tombstones. Ce mécanisme repose sur deux conditions :

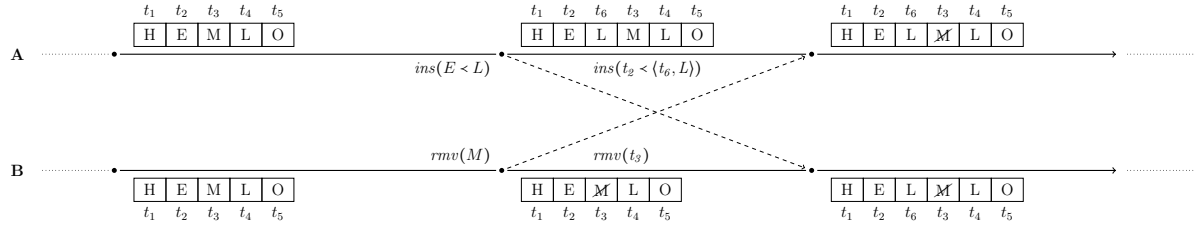


FIGURE 1.11 – Modifications concurrentes d'une séquence répliquée RGA

- (i) La stabilité causale de l'opération rmv , c.-à-d. l'ensemble des noeuds a observé la suppression de l'élément et ne peut émettre d'opérations utilisant l'élément supprimé comme prédecesseur.
 - (ii) L'impossibilité pour l'ensemble des noeuds de générer un $S4Vector$ plus petit que l'élément suivant la tombstone⁶.
- Concernant le modèle de livraison adopté, RGA repose sur une livraison causale des opérations. [57] indique que ce modèle de livraison pourrait être relaxé, de façon à ne plus dépendre de vecteurs d'horloges. Ce point est néanmoins laissé comme piste de recherche future. À notre connaissance, celle-ci n'a pas été explorée dans la littérature. Dans [5], ELVINGER indique que RGA pourrait adopter le même modèle de livraison que WOOT. Pour rappel, celui-ci consiste en :
- (i) L'opération $ins(predId < \langle id, elt \rangle)$ ne peut être délivrée qu'après la livraison de l'opération d'insertion de l'élément associé à $predId$.
 - (ii) L'opération $rmv(id)$ ne peut être délivrée qu'après la livraison de l'opération d'insertion de l'élément associé à id .

Nous secondons cette observation.

- *Matthieu: TODO : Ajouter quelques lignes sur efficacité de RGA. Meilleure complexité en temps pour l'intégration des opérations. [58, 59] montrent que RGA est optimal du point de vue complexité spatiale comme CRDT pour Séquence par élément basée sur l'approche tombstone. Implémenté dans Yjs et Automerger.*
- Plusieurs extensions de RGA furent proposées par la suite. Dans [60], BRIOT, URSO et SHAPIRO indiquent que les pauvres performances des modifications locales⁷ des CRDTs pour Séquence constituent une de leurs limites. Il s'agit en effet des performances impactant le plus l'expérience utilisateur, ceux-ci s'attendant à un feedback immédiat de la part de l'application. Les auteurs souhaitent donc réduire la complexité en temps des modifications locales à une complexité logarithmique. Pour cela, ils proposent l'*identifier structure*, une structure de données auxiliaire utilisable par les CRDTs pour Séquence. Cette structure permet de retrouver plus efficacement l'identifiant d'un élément à partir de son index, au pris d'un surcoût en métadonnées. BRIOT, URSO et SHAPIRO combinent cette structure de données à un mécanisme

6. On constate cependant que la condition (i) implique la condition (ii), puisque tout noeud ayant observé l'insertion de l'élément suivant possédera un vecteur d'horloges supérieur.

7. Relativement par rapport aux algorithmes de l'approche OT.

d'aggrégation des éléments en blocs⁸ tels que proposés par [61, 4], qui permet de réduire la quantité de métadonnées stockées par la séquence répliquée. Cette combinaison aboutit à la définition d'un nouveau CRDT pour Séquence, *RGATreeSplit*, qui offre une meilleure complexité en temps et en espace.

- Dans [62], KLEPPMANN et al. mettent en lumière un problème des CRDTs pour Séquence. Lorsque des séquences de modifications sont effectuées en concurrence par des noeuds, les CRDTs assurent la convergence des répliques mais pas la correction du résultat. Notamment, il est possible que les éléments insérés en concurrence se retrouvent entrelacés. *Matthieu: TODO : Insérer et présenter exemple* Pour remédier

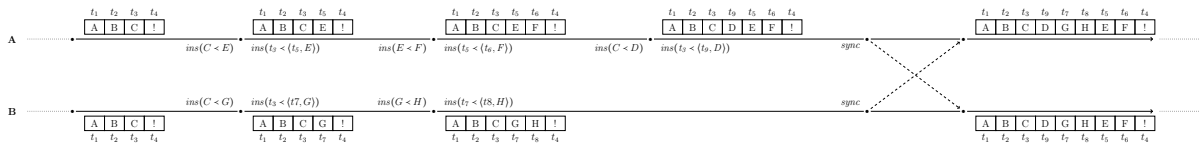


FIGURE 1.12 – Entrelacement d'éléments insérés de manière concurrente

à ce problème, les auteurs définissent une nouvelle spécification que doivent respecter les approches pour la mise en place de séquences répliquées : *la spécification forte sans entrelacement des séquences répliquées*. Basée sur la spécification forte des séquences répliquées spécifiée dans [58, 59], cette nouvelle spécification ajoute comme propriété que les éléments insérés en concurrence ne doivent pas s'entrelacer dans l'état final. KLEPPMANN et al. proposent ensuite une évolution de RGA respectant cette spécification. Pour cela, les auteurs ajoutent à l'opération *ins* un paramètre, *samePredId*, un ensemble correspondant à l'ensemble des identifiants connus utilisant le même *predId* que l'élément inséré. En maintenant en plus un exemplaire de cet ensemble pour chaque élément de la séquence, il est possible de déterminer si deux opérations *ins* sont concurrentes ou causalement liées et ainsi déterminer comment ordonner leurs éléments. Cependant, les auteurs ne prouvent pas dans [62] que cette extension empêche tout entrelacement⁹.

1.4.2 Approche à identifiants densément ordonnés

Treedoc

- [52, 64] proposent une nouvelle approche pour CRDTs pour Séquence. Se base sur des identifiants de position, respectant les propriétés suivantes :
 - Chaque élément se voit attribuer un identifiant.
 - Aucune paire d'éléments ne partage le même identifiant.
 - L'identifiant d'un élément est immuable.
 - Il existe un ordre total strict sur les identifiants, $<_{id}$, cohérent avec l'ordre des éléments dans la séquence.

8. Nous détaillerons ce mécanisme par la suite.

9. Un travail en cours [63] indique en effet qu'une séquence répliquée empêchant tout entrelacement est impossible.

- (v) Les identifiants sont tirés d'un ensemble dense, noté \mathbb{I} .
- La propriété (v) signifie que :

$$\forall predId, succId \in \mathbb{I}, \exists id \in \mathbb{I} \cdot predId <_{id} id <_{id} succId$$

Par exemple, les nombres réels forment un ensemble dense. Ceux-ci sont néanmoins inutilisables en informatique puisqu'ils nécessiteraient une précision infinie. [64] utilise donc un type dédié pour les émuler.

- L'utilisation d'identifiants de position permet de redéfinir les modifications :
 - (i) $ins(pred < elt < succ)$ devient alors $ins(id, elt)$, avec $predId <_{id} id <_{id} succId$.
 - (ii) $rmv(elt)$ devient $rmv(id)$.

Ces redéfinitions permettent d'obtenir une spécification de la séquence avec des modifications commutatives.

- À partir de cette spécification, PREGUICA et al. propose un CRDT pour Séquence : *Treedoc*. Ce dernier se base sur un arbre binaire pour générer les identifiants de position.
- La racine de l'arbre binaire, notée ϵ , correspond à l'identifiant du premier élément inséré dans la séquence répliquée. Puis, pour générer l'identifiant d'un nouvel élément inséré à gauche (resp. à droite) d'un noeud de l'arbre binaire, *Treedoc* concatène un 0 (resp. un 1) à l'identifiant de ce dernier. *Matthieu: TODO : Figure arbre binaire d'identifiants.*

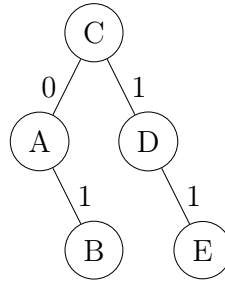


FIGURE 1.13 – Identifiants de positions

- Ce mécanisme souffre néanmoins d'un écueil : en l'état, plusieurs noeuds du système peuvent générer un même identifiant pour des éléments insérés en concurrence, contravenant alors à la propriété (ii). Pour corriger cela, *Treedoc* ajoute aux identifiants des désambiguateurs : un *Dot*. Un désambiguateur est ajouté à chaque partie d'un identifiant lorsque nécessaire, c.-à-d.
 - (i) La partie courante de l'identifiant est la fin de l'identifiant.
 - (ii) La partie courante de l'identifiant nécessite désambiguation, c.-à-d. plusieurs éléments utilisent ce même identifiant.

Il convient de noter que les noeuds de l'arbre binaire des identifiants peuvent ainsi contenir une liste d'identifiants en cas d'insertions concurrentes. *Matthieu: TODO : Figure arbre binaire d'identifiants avec désambiguateurs et mini-nodes.*

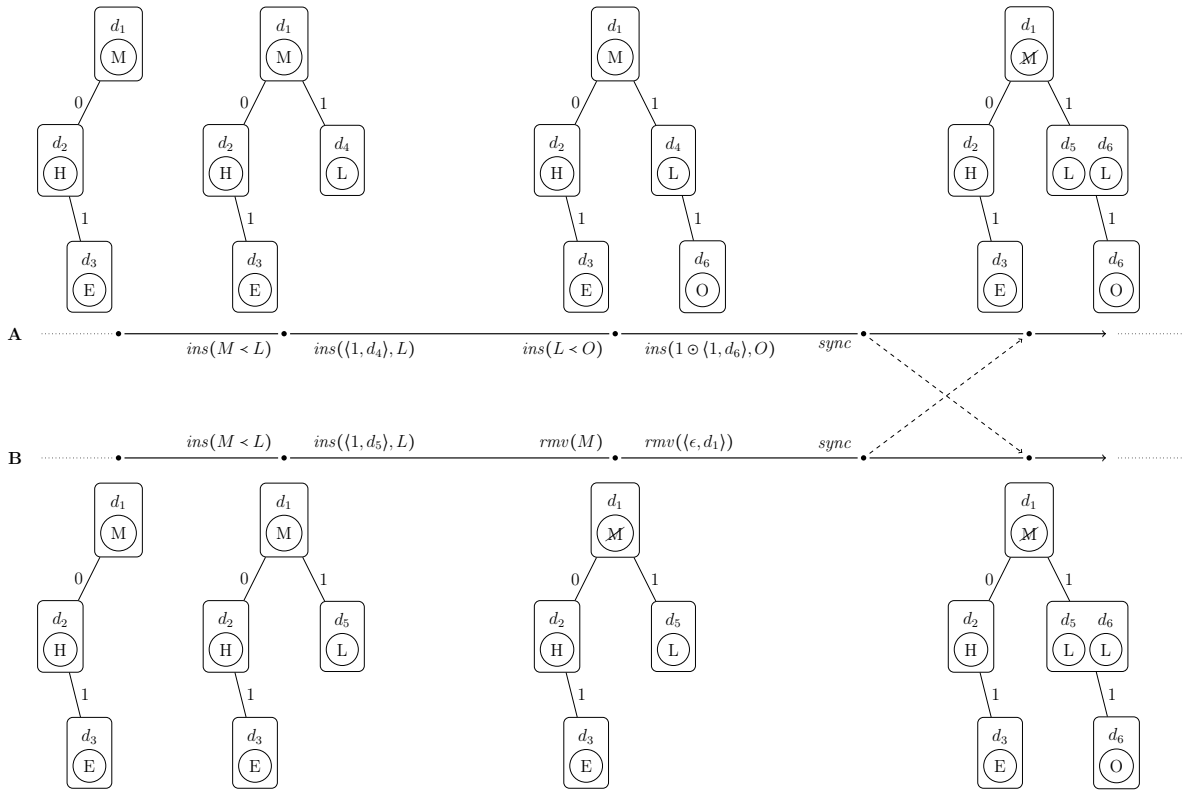


FIGURE 1.14 – Modifications concurrentes d'une séquence répliquée Treedoc

— *Matthieu: TODO : Figure exemple d'édition collaborative*

— Concernant le modèle de livraison utilisé, [64] indique reposer sur le modèle de livraison causal. En pratique, nous pouvons néanmoins relaxer le modèle de livraison comme expliqué dans [5] :

- (i) Les opérations *ins* peuvent être livrées dans n'importe quel ordre.
- (ii) L'opération *rmv(id)* ne peut être livrée qu'après la livraison de l'opération d'insertion de l'élément associé à *id*.

— Treedoc souffre néanmoins de plusieurs limites. Tout d'abord, le mécanisme d'identifiants de positions proposé est couplé à la structure d'arbre binaire. Cependant, les utilisateurs ont tendance à écrire de manière séquentielle, c.-à-d. dans le sens d'écriture de la langue utilisée. Les nouveaux identifiants forment donc généralement une liste chaînée, qui déséquilibre l'arbre.

— Ensuite, Treedoc doit conserver un noeud de l'arbre des identifiants malgré sa suppression lorsque ce dernier possède des enfants. Ce noeud de l'arbre devient alors une pierre tombale. Le mécanisme de GC des pierres tombales de Treedoc ne repose cependant pas sur la stabilité causale d'opérations, mais sur l'état du noeud de l'arbre, c.-à-d. si la pierre tombale devient une feuille. Néanmoins, l'évaluation de [64] a montré que les pierres tombales pouvaient représenter jusqu'à 95% des noeuds de l'arbre.

- Finalement, Treedoc souffre du problème de l'entrelacement d'éléments insérés de manière concurrente, contrairement à ce qui est conjecturé dans [62].

Logoot

- Logoot [65, 66]

Matthieu: NOTE : Ajouter LogootSplit de manière sommaire aussi à cet endroit ?

Matthieu: TODO : Autres Sequence CRDTs à considérer : String-wise CRDT [61], Chronofold [67]

1.4.3 Synthèse

- Deux approches différentes pour la résolution de conflits ont été proposées pour CRDTs pour Séquence. Chacune de ces approches visent à minimiser surcoût du type répliqué, que ce soit d'un point de vue mémoire, computations et réseau.
- Au fil des années, ces approches ont été raffinées avec de nouveaux CRDTs de plus en plus efficaces.
- Néanmoins, malgré les évaluations et comparaisons, la littérature n'a pas établi une supériorité d'une approche sur l'autre. Les approches proposent seulement des compromis différents sur la nature du surcoût, que nous récapitulons dans Tableau 1.2. L'approche basée sur pierres tombales offre une consommation réseau constante grâce à ses identifiants de taille fixe, mais souffre d'une consommation mémoire ne pouvant qu'augmenter. L'approche basée sur identifiants densément ordonnés bénéficie d'un meilleur délai de diffusion des modifications, les modifications pouvant être livrées dans le désordre, mais souffre d'une empreinte réseau augmentant avec la taille de ses identifiants. L'approche basée sur pierres tombales et l'approche basée sur identifiants densément ordonnés souffrent toutes les deux d'une augmentation théorique de leur surcoût en mémoire et en computations, respectivement dûe au nombre forcément croissant d'éléments stockés dans la Séquence et à la taille croissante¹⁰ des identifiants associés aux éléments de la Séquence.

TABLE 1.2 – Récapitulatif comparatif des différents approches pour CRDTs pour Séquence

	Dense ids-based	Tombstoned-based
Performances stables	✗	✗
Identifiants de taille fixe	✗	✓
Éléments réellement supprimés	✓	✗
Empreinte réseau fixe	✗	✓
Peut s'affranchir de la cohérence causale	✓	✓

- Pour la suite de ce manuscrit, nous prenons LogootSplit comme base de travail. Nous détaillons donc son fonctionnement dans la section suivante.

10. [54] montre expérimentalement que les performances de l'approche basée sur identifiants densément ordonnés restent stables tout au long des tâches d'édition collaborative proposées.

1.5 LogootSplit

LogootSplit [4] est l'état de l'art des séquences répliquées à identifiants densément ordonnés. Comme expliqué précédemment, LogootSplit utilise des identifiants provenant d'un ordre total dense pour positionner les éléments dans la séquence répliquée.

1.5.1 Identifiants

Pour ce faire, LogootSplit assigne des identifiants composés d'une liste de tuples aux éléments. Les tuples sont définis de la manière suivante :

Définition 5 (Tuple) *Un Tuple est un quadruplet $\langle position, nodeId, nodeSeq, offset \rangle$ où*

- *position incarne la position souhaitée de l'élément.*
- *nodeId est l'identifiant unique du noeud qui a généré le tuple.*
- *nodeSeq est le numéro de séquence courant du noeud à la génération du tuple.*
- *offset indique la position de l'élément au sein d'un bloc. Nous reviendrons plus en détails sur ce composant dans la sous-section 1.5.2.*

Matthieu: TODO : Ajouter une relation d'ordre sur les tuples

Dans ce manuscrit, nous représentons les tuples par le biais de la notation suivante : $position^{nodeId}_{nodeSeq}_{offset}$ où *position* est une lettre minuscule, *nodeId* une lettre majuscule et *nodeSeq* et *offset* des entiers, e.g. i_0^{B0} .

À partir de là, les identifiants LogootSplit sont définis de la manière suivante :

Définition 6 (Identifiant) *Un Identifiant est une liste de Tuples.*

Matthieu: TODO : Définir la notion de base (et autres fonctions utiles sur les identifiants ? genre isPrefix, concat, getTail...)

Nous représentons les identifiants en listant les tuples qui les composent. Par exemple, l'identifiant composé des tuples $\langle \langle i, B, 0, 0 \rangle \langle f, A, 0, 0 \rangle \rangle$ est présenté de la manière suivante : $i_0^{B0} f_0^{A0}$.

Les identifiants ont pour rôle d'ordonner les éléments relativement les uns par rapport aux autres. Pour ce faire, une relation d'ordre total aux identifiants est associée à l'ensemble des identifiants :

Définition 7 (Relation $<_{id}$) *La relation $<_{id}$ est un ordre strict total sur l'ensemble des identifiants. Elle permet aux noeuds de comparer n'importe quelle paire d'identifiants. Elle est définie en utilisant l'ordre lexicographique sur les composants des différents tuples des identifiants comparés.*

- *En utilisant cette relation d'ordre, les noeuds peuvent ordonner les éléments grâce à leur identifiant.*
- *Par exemple, déterminent que $i_0^{A1} <_{id} i_0^{B0}$ car les positions sont identiques et que le *nodeId* (A) du premier est plus petit que le *nodeId* (B) du second*
- *et que $i_0^{B0} <_{id} i_0^{B0} f_0^{A0}$ car le premier est un préfixe du second*

Matthieu: TODO : Montrer que cet ensemble d'identifiants est un ensemble dense

1.5.2 Aggrégation dynamique d'éléments en blocs

Au lieu de stocker les identifiants de chaque élément de la séquence, LogootSplit propose d'aggréger de façon dynamique les éléments dans des blocs. Pour cela, LogootSplit introduit la notion d'intervalle d'identifiants :

Définition 8 (IdInterval) *Un IdInterval est un couple $\langle idBegin, offsetEnd \rangle$ où*

- *$idBegin$ est l'identifiant du premier élément de l'intervalle.*
- *$offsetEnd$ est l'offset du dernier identifiant de l'intervalle.*

Les intervalles d'identifiants permettent à LogootSplit d'assigner logiquement un identifiant à un ensemble d'éléments, tout en ne stockant réellement que l'identifiant de son premier élément et le dernier offset de son dernier élément.

LogootSplit regroupe les éléments avec des identifiants *contigus* dans un interval. Nous appelons *contigus* deux identifiants qui partagent une même base (c.-à-d. qui sont identiques à l'exception de leur dernier offset) et dont les *offsets* sont consécutifs. Nous représentons un intervalle d'identifiants à l'aide du formalisme suivant : $position_{begin..end}^{nodeId\ nodeSeq}$ où *begin* est l'offset du premier identifiant de l'intervalle et *end* du dernier.

Les blocs permettent d'associer un intervalle d'identifiants aux éléments correspondant. Les blocs sont définis de la manière suivante :

Définition 9 (Bloc) *Un Bloc est un quadruplet $\langle idInterval, elts, isAppendable, isPrependable \rangle$ où*

- *$idInterval$ est l'intervalle d'identifiants formant le bloc*
- *$elts$ sont les éléments contenus dans le bloc*
- *$isAppendable$ (resp. $isPrependable$) est un booléen indiquant si l'auteur du bloc peut ajouter un nouvel élément en fin (resp. début) de bloc*

La Figure 1.15 présente un exemple de séquence LogootSplit : dans la Figure 1.15a, les identifiants i_0^{B0} , i_1^{B0} , i_2^{B0} forment une chaîne d'identifiants contigus. LogootSplit est donc capable de regrouper ces éléments en un bloc représentant l'intervalle d'identifiants $i_{0..2}^{B0}$ pour minimiser les métadonnées stockées, comme montré dans la Figure 1.15b.

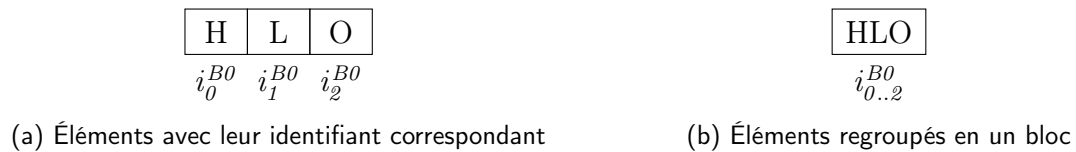


FIGURE 1.15 – Représentation d'une séquence LogootSplit contenant les éléments "HLO"

Cette fonctionnalité réduit le nombre d'identifiants stockés au sein de la structure de données, puisque les identifiants sont conservés à l'échelle des blocs plutôt qu'à l'échelle de chaque élément. Ceci permet de réduire de manière significative le surcoût en métadonnées de la structure de données. L'utilisation de blocs améliore aussi les performances de la

structure de données. En effet, l'utilisation de blocs permet de parcourir plus efficacement la structure de données. Les blocs permettent aussi d'effectuer des modifications à l'échelle de la chaîne de caractères et non plus seulement caractère par caractère.

Matthieu: TODO : indiquer que le couple $\langle nodeId, nodeSeq \rangle$ permet d'identifier de manière unique la base d'un bloc ou d'un identifiant

Notons que pour une séquence donnée, nous pouvons identifier chacun de ses identifiants par le triplet $\langle nodeId, nodeSeq, offset \rangle$ issue de leur dernier Tuple. Par exemple, le triplet $\langle B, 0, 2 \rangle$ désigne de manière unique l'identifiant i_2^{B0} dans Figure 1.15.

1.5.3 Modèle de données

ANDRÉ et al. [4] définissent une séquence LogootSplit de la manière suivante :

Définition 10 (Séquence LogootSplit) Une séquence Séquence LogootSplit est un triplet $\langle nodeId, nodeSeq, blocks \rangle$ où

- $nodeId$ est l'identifiant du noeud.
- $nodeSeq$ est le numéro de séquence courant du noeud.
- $blocks$ est une liste de Blocs correspondant à l'état actuel de la séquence répliquée.

Plusieurs fonctions sont définies sur cette structure de données et permettent de l'interroger et de la modifier :

- $ins(S, index, elts)$ permet d'insérer les éléments $elts$ à la position $index$ dans la séquence S . Cette fonction génère et associe un intervalle d'identifiants valide aux éléments insérés Elle retourne une opération *insert* permettant aux autres noeuds d'intégrer la modification à leur état.

Définition 11 (insert) Une opération *insert* est un couple $\langle id, elts \rangle$ où

- id est l'identifiant du premier élément inséré par cette opération.
- $elts$ est la liste des éléments insérés par cette opération.
- $rem(S, index, length)$ permet de supprimer $length$ éléments à partir la position $index$ dans la séquence S . Cette fonction répertorie les éléments supprimés sous la forme d'intervalles d'identifiants. Elle retourne une opération *remove* permettant aux autres noeuds d'intégrer la modification à leur état.

Définition 12 (remove) Une opération *remove* est une liste d'intervalles d'identifiants où chaque intervalle désigne un ensemble d'éléments à supprimer.

Nous présentons dans la Figure 1.16 un exemple d'utilisation de cette séquence répliquée.

Dans cet exemple, deux noeuds A et B répliquent et éditent collaborativement un document texte en utilisant LogootSplit. Ils partagent initialement le même état : une séquence composée d'un seul bloc associant les identifiants $i_{0..3}^{B0}$ aux éléments "HRLO". Les noeuds se mettent ensuite à éditer le document.

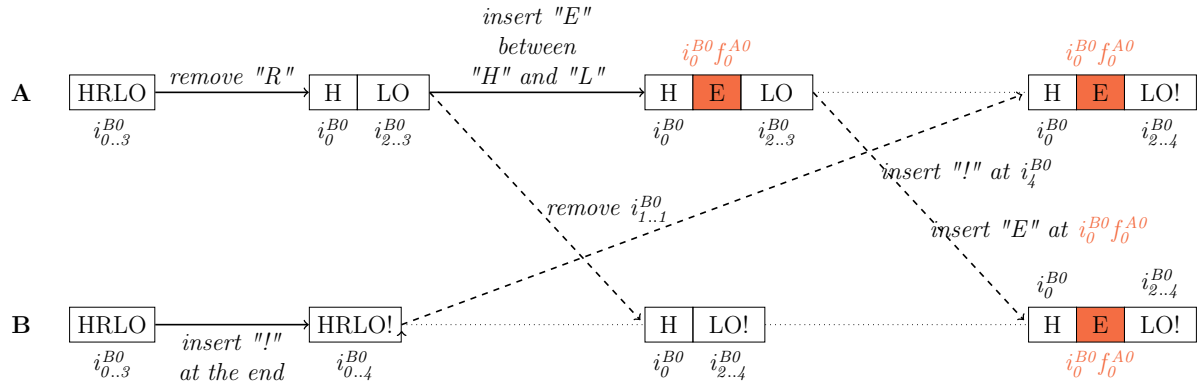


FIGURE 1.16 – Modifications concurrentes d'une séquence répliquée LogootSplit

Le noeud A commence par supprimer l'élément "R" de la séquence. LogootSplit génère l'opération *remove* correspondante en utilisant l'identifiant de l'élément supprimé (i_1^{B0}). Cette opération est envoyée au noeud B pour qu'il intègre cette modification.

Le noeud A insère ensuite un élément "E" dans la séquence, entre le "H" et le "L". LogootSplit doit alors générer un identifiant id à associer à ce nouvel élément. Ce nouvel identifiant id doit respecter la contrainte suivante : $i_0^{B0} <_{id} id <_{id} i_2^{B0}$. Cependant, LogootSplit ne peut pas générer un identifiant composé d'un seul tuple respectant cet ordre. LogootSplit génère alors id en recopiant le premier tuple (i_0^{B0}) et en y ajoutant un nouveau tuple (f_0^{A0}). LogootSplit génère l'opération *insert* correspondante, indiquant l'élément à insérer et sa position grâce à son identifiant. Cette opération est ensuite diffusée sur le réseau.

En parallèle, le noeud B insère un élément "!" à la fin de la séquence. Comme le noeud B est l'auteur du bloc i_0^{B0} , il peut y ajouter de nouveaux éléments. LogootSplit associe donc l'identifiant i_4^{B0} à l'élément "!" et l'ajoute au bloc existant.

Les noeuds se synchronisent ensuite. Le noeud A reçoit l'opération *insert* de l'élément "!" à la position i_4^{B0} . Le noeud A détermine que cet élément doit être inséré à la fin de la séquence (puisque $i_3^{B0} <_{id} i_4^{B0}$) et qu'il peut être ajouté au bloc i_2^{B0} (puisque i_3^{B0} et i_4^{B0} sont contigus).

De son côté, le noeud B reçoit tout d'abord l'opération *remove* des éléments identifiés par l'intervalle i_1^{B0} , c.-à-d. l'élément attaché à l'identifiant i_1^{B0} . Le noeud B supprime donc l'élément "R" de son état.

Il reçoit ensuite l'opération *insert* de l'élément "E" à la position $i_0^{B0} f_0^{A0}$. Le noeud B insère cet élément entre les éléments "H" et "L" (puisque $i_0^{B0} <_{id} i_0^{B0} f_0^{A0} <_{id} i_2^{B0}$), respectant ainsi l'intention du noeud A.

Matthieu: NOTE : Pourrait définir dans cette sous-section la notion de séquence bien-formée

1.5.4 Modèle de livraison

Afin de garantir son bon fonctionnement, LogootSplit doit être associé à une couche de livraison de messages garantissant plusieurs propriétés.

Livraison des opérations en exactement un exemplaire

Tout d'abord, la couche de livraison de messages doit assurer que toutes les opérations soient délivrées aux noeuds, mais qu'une seule et unique fois. La Figure 1.17 représente un exemple illustrant la nécessité de cette contrainte.

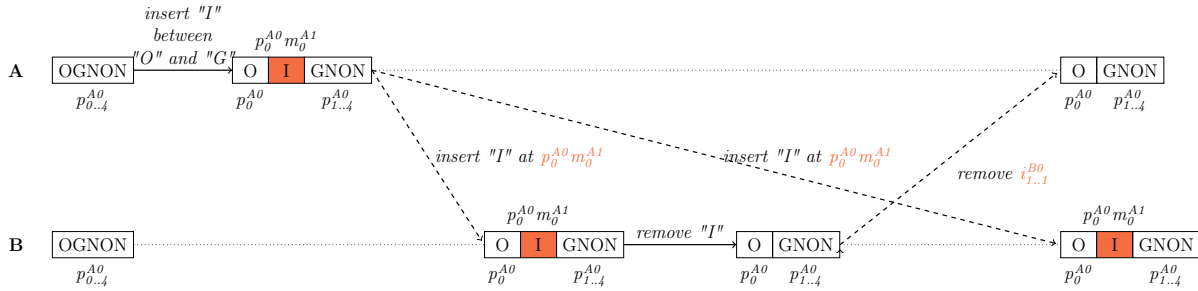


FIGURE 1.17 – Résurgence d'un élément supprimé suite à la relivraison de son opération *insert*

Dans cet exemple, deux noeuds A et B répliquent et éditent collaborativement une séquence. La séquence répliquée contient initialement les éléments "OGNON", qui sont associés à l'intervalle d'identifiants $p_{0..4}^{A0}$.

Le noeud A commence par insérer un nouvel élément, "I", dans la séquence entre les éléments "O" et "G". L'opération *insert* résultante, insérant l'élément "I" à la position $p_0^{A0} m_0^{A1}$, est diffusée au noeud B.

À la réception de l'opération *insert*, le noeud B l'intègre à son état. Puis il supprime dans la foulée ce nouvel élément. L'opération *remove* générée est envoyée au noeud A.

Le noeud A intègre l'opération *remove*, ce qui a pour effet de supprimer l'élément "I" associé à l'identifiant $p_0^{A0} m_0^{A1}$. Il obtient alors un état équivalent à celui du noeud B.

Cependant, l'opération *insert* insérant l'élément "I" à la position $p_0^{A0} m_0^{A1}$ est de nouveau délivrée au noeud B. De multiples raisons peuvent être à l'origine de cette nouvelle livraison : perte du message d'*acknowledgment*, utilisation d'un protocole de diffusion épidémique des messages, déclenchement du mécanisme d'anti-entropie en concurrence... Le noeud B ré-intègre alors l'opération *insert*, ce qui fait revenir l'élément "I" et l'identifiant associé. L'état du noeud B diverge désormais de celui-ci du noeud A.

Pour se prémunir de ce type de scénarios, LogootSplit requiert que la couche de livraison des messages assure une livraison en exactement un exemplaire des opérations. Cette contrainte permet d'éviter que d'anciens éléments et identifiants ressurgissent après leur suppression chez certains noeuds uniquement à cause d'une livraison multiple de l'opération *insert* correspondante.

Matthieu: QUESTION : Ajouter quelques lignes ici sur comment faire ça en pratique (Ajout d'un dot aux opérations, maintien d'un dot store au niveau de la couche livraison, vérification que dot pas encore présent dans dot store avant de passer opération à la structure de données) ? Ou je garde ça pour le chapitre sur MUTE ?

Livraison de l'opération *remove* après l'opération *insert*

Une autre propriété que doit assurer la couche de livraison de messages est que les opérations *remove* doivent être livrées au CRDT après les opérations *insert* correspondantes. La Figure 1.18 présente un exemple justifiant cette contrainte.

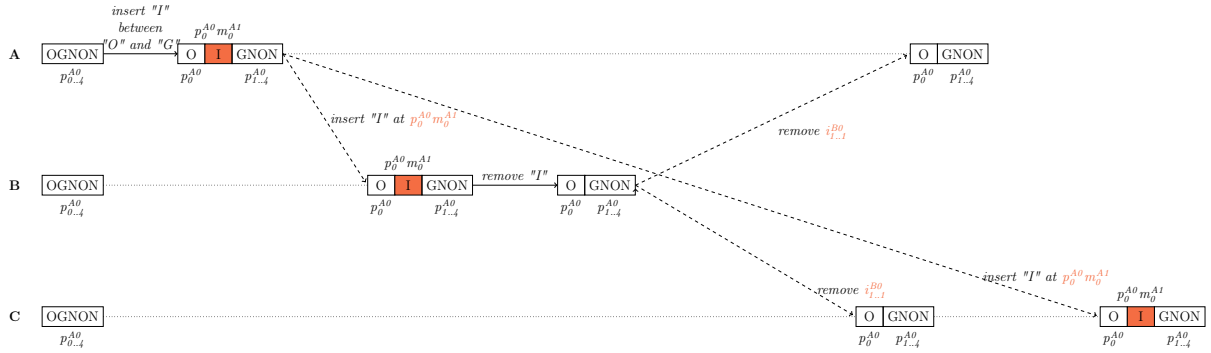


FIGURE 1.18 – Non-effet de l'opération *remove* car reçue avant l'opération *insert* correspondante

Dans cet exemple, trois noeuds A, B et C répliquent et éditent collaborativement une séquence. Le noeud A commence par insérer un nouvel élément, "I", dans la séquence entre les éléments "O" et "G". L'opération *insert* résultante, insérant l'élément "I" à la position $p_0^{A0} m_0^{A1}$, est diffusée aux autres noeuds.

À la réception de l'opération *insert*, le noeud B l'intègre à son état. Cependant, le noeud B supprime dans la foulée l'élément nouvellement ajouté. Il diffuse ensuite l'opération *remove* générée.

Toutefois, suite à un aléa du réseau, l'opération *remove* supprimant l'élément "I" est livrée au noeud C avant l'opération *insert* l'ajoutant à son état. Lorsque le noeud C reçoit l'opération *remove*, il parcourt son état à la recherche de l'élément "I" pour le supprimer. Cependant, celui-ci n'est pas présent dans son état courant. L'intégration de l'opération s'achève donc sans effectuer de modification.

Le noeud C reçoit ensuite l'opération *insert*. Le noeud C intègre ce nouvel élément dans la séquence en utilisant son identifiant ($p_0^{A0} <_{id} p_0^{A0} m_0^{A1} <_{id} p_1^{A0}$).

Ainsi, l'état du noeud C diverge de celui-ci des autres noeuds à terme, et cela malgré que les noeuds A, B et C aient intégré le même ensemble d'opérations. Ce résultat transgresse la propriété de Cohérence forte à terme (SEC) que doivent assurer les CRDTs. Afin d'empêcher ce scénario de se produire, LogootSplit impose donc la livraison causale des opérations *remove* par rapport aux opérations *insert* correspondantes.

Matthieu: QUESTION : Même que pour la exactly-once delivery, est-ce que j'explique ici comment assurer cette contrainte plus en détails (Ajout des dots des opérations insert en dépendances de l'opération remove, vérification que dots présents dans dot store avant de passer l'opération remove à la structure de données) ou je garde ça pour le chapitre sur MUTE ?

Définition du modèle de livraison

Pour résumer, la couche de livraison des opérations associée à LogootSplit doit respecter le modèle de livraison suivant :

Définition 13 (Exactly-once + Causal remove) *Le modèle de livraison Exactly-once + Causal remove définit les 3 règles suivantes sur la livraison des opérations :*

1. Une opération doit être délivrée à l'ensemble des noeuds à terme,
2. Une opération doit être délivrée qu'une seule et unique fois aux noeuds,
3. Une opération remove doit être délivrée à un noeud une fois que les opérations insert des éléments concernés par la suppression ont été délivrées à ce dernier.

Il est à noter que ELVINGER [5] a récemment proposé dans ses travaux de thèse Dotted LogootSplit, un nouveau Sequence CRDT basée sur les différences. Inspiré de Logoot et LogootSplit, ce nouveau CRDT associe une séquence à identifiants densément ordonnés à un contexte causal. Le contexte causal est une structure de données permettant à Dotted LogootSplit de représenter et de maintenir efficacement les informations des modifications déjà intégrées à l'état courant. Cette association permet à Dotted LogootSplit de fonctionner de manière autonome, sans imposer de contraintes particulières à la couche livraison autres que la livraison à terme.

1.5.5 Limites

Comme indiqué précédemment, la taille des identifiants provenant d'un ordre total dense est variable. Quand les noeuds insèrent de nouveaux éléments entre deux autres ayant la même valeur de *position*, LogootSplit n'a pas d'autre choix que d'augmenter la taille de l'identifiant résultant. La Figure 1.19 illustre de tels cas. Dans cet exemple, puisque le noeud A insère un nouvel élément entre deux identifiants contigus i_0^{B0} et i_1^{B0} , LogootSplit ne peut pas générer un identifiant adapté de la même taille. Pour respecter l'ordre souhaité, LogootSplit génère un identifiant en ajoutant un nouveau tuple à l'identifiant du prédecesseur : $i_0^{B0} f_0^{A0}$.

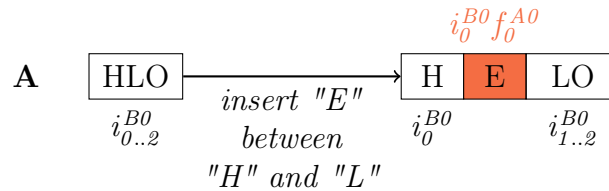


FIGURE 1.19 – Insertion menant à une augmentation de la taille des identifiants

Par conséquent, la taille des identifiants a tendance à croître alors que le système progresse. Cette croissance impacte négativement les performances de la structure de données sur plusieurs aspects. Puisque les identifiants attachés aux éléments deviennent plus long, le surcoût en métadonnées augmente. Ceci augmente aussi la consommation en bande-passante puisque les noeuds doivent diffuser les identifiants aux autres.

Matthieu: TODO : Ajouter une phrase pour expliquer que la croissance des identifiants impacte aussi le temps d'intégration des modifications

De plus, le nombre de blocs composant la séquence répliquée augmente au fil du temps. En effet, plusieurs contraintes sur la génération d'identifiants empêchent les noeuds d'ajouter des nouveaux éléments aux blocs existants. Par exemple, seul le noeud qui a généré un bloc peut ajouter un élément à ce dernier. Ces limitations provoquent la génération de nouveau blocs. La séquence se retrouve finalement fragmentée en de nombreux blocs de seulement quelques caractères chacun. Cependant, aucun mécanisme pour fusionner les blocs à posteriori n'est fourni. L'efficacité de la structure décroît donc puisque chaque bloc entraîne un surcoût.

Comme illustré plus loin, nous avons mesuré au cours de nos évaluations que le contenu représente à terme moins de 1% de taille de la structure de données. Les 99% restants correspondent aux métadonnées utilisées par la séquence répliquée. Il est donc nécessaire de proposer des mécanismes et techniques afin de mitiger les problèmes soulignés précédemment.

1.6 Mitigation du surcoût des séquences répliquées sans conflits

- Plusieurs approches ont été proposées pour réduire croissance des métadonnées dans Sequence CRDTs
- RGA (et RGASplit) propose un mécanisme de GC des pierres tombales. Nécessite cependant stabilité causale des opérations de suppression. S'agit d'une contrainte forte, peu adaptée aux systèmes dynamiques à large échelle. *Matthieu: TODO : Trouver référence sur la stabilité causale dans systèmes dynamiques*
- Core & Nebula propose un mécanisme de ré-équilibrage de l'arbre pour Treedoc. Le ré-équilibrage a pour effet de supprimer des potentielles pierres tombales et de réduire la taille des identifiants. Repose sur un algorithme de consensus. S'agit de nouveau d'une contrainte forte pour systèmes dynamique à large échelle. Pour y pallier, propose de séparer les pairs entre deux ensembles : Core et Nebula. Permet de limiter le nombre participant au consensus. Un protocole de rattrapage permet aux noeuds de la Nebula de mettre à jour leurs modifications concurrentes à un ré-équilibrage.
- LSEQ adopte une autre approche. Part du constat que les identifiants dans Logoot croissent de manière linéaire. Vise une croissance logarithmique des identifiants. Pour cela, propose de nouvelles fonctions d'allocation des identifiants visant à maximiser le nombre d'identifiants insérés avant de devoir augmenter la taille de l'identifiant. Propose aussi d'utiliser une base exponentielle pour la valeur *position* des identifiants. Atteint ainsi la croissance polylogarithmique des identifiants, sans coordination requise entre les noeuds et mécanisme supplémentaire. Solution adaptée aux systèmes distribués à large échelle. Conjecture cependant que cette approche se marie mal avec les Sequence CRDTs utilisant des blocs. En effet, ajoute une

raison supplémentaire à la croissance des identifiants : l'insertion entre identifiants contigus. Force alors la croissance des identifiants.

1.7 Synthèse

- Systèmes distribués adoptent le modèle de la réplication optimiste pour offrir de meilleures performances, c.-à-d. disponibilité et latence, et assurer la résilience du système, c.-à-d. accroître la capacité de tolérance aux pannes.
- Ce modèle autorise les noeuds à modifier leur copie sans coordination, provoquant ainsi des divergences temporaires. Pour résoudre les potentiels conflits et assurer la convergence à terme des copies, systèmes utilisent les CRDTs en place et lieu des types de données séquentiels.
- CRDTs pour Séquence ont été proposés pour conception d'éditeurs collaboratifs pair-à-pair. Deux approches sont utilisées pour concevoir leur mécanismes de résolution de conflits : l'approche basée sur les pierres tombales et l'approche basée sur les identifiants densément ordonnés.
- Chacune de ces approches introduit un surcoût croissant, pénalisant leurs performances à terme. Plusieurs travaux ont été proposés pour limiter ce surcoût, notamment [1, 2] qui présentent un mécanisme de renommage des identifiants pour les CRDTs pour Séquence basés sur identifiants densément ordonnés.
- Mais cette approche requiert un mécanisme de consensus, des renommages concurrents provoquant un nouveau conflit. Cette contrainte empêche son utilisation dans les systèmes pair-à-pair ne disposant pas de noeuds suffisamment stables et bien connectés pour exécuter le mécanisme de consensus.

1.8 Proposition

- Dans ce manuscrit, nous proposons et présentons un nouveau mécanisme de renommage pour CRDTs pour Séquence, ne nécessitant pas de coordination synchrone entre les noeuds.
- Concevons ce mécanisme pour le CRDT pour Séquence LogootSplit, mais principe de notre approche est générique. Pourrait ainsi l'adapter et proposer un équivalent pour autres CRDTs pour Séquence, e.g. RGASplit.
- Présentons et détaillons notre contribution dans le chapitre suivant.

Chapitre 2

Conclusions et perspectives

Sommaire

2.1	Résumé des contributions	40
2.2	Perspectives	40
2.2.1	Définition de relations de priorité pour minimiser les traitements	40
2.2.2	Redéfinition de la sémantique du renommage en déplacement d'éléments	40
2.2.3	Définition de types de données répliquées sans conflits plus complexes	40
2.2.4	Étude comparative des différentes familles de CRDTs	40
2.2.5	Définition d'opérations supplémentaires pour fonctionnalités liées à l'édition collaborative	41
2.2.6	Conduction d'expériences utilisateurs d'édition collaborative . .	41
2.2.7	Comparaison des mécanismes de synchronisation	42
2.2.8	Distance entre versions d'un document	42
2.2.9	Contrôle d'accès	42
2.2.10	Détection et éviction de pairs malhonnêtes	42
2.2.11	Vecteur <i>epoch-based</i>	43
2.2.12	Fusion de versions distantes d'un document collaboratif	44
2.2.13	Rôles et places des bots dans systèmes collaboratifs	44

2.1 Résumé des contributions

2.2 Perspectives

2.2.1 Définition de relations de priorité pour minimiser les traitements

2.2.2 Redéfinition de la sémantique du renommage en déplacement d'éléments

2.2.3 Définition de types de données répliquées sans conflits plus complexes

2.2.4 Étude comparative des différentes familles de CRDTs

- La spécification récente des Delta-based CRDTs . Ce nouveau type de CRDTs se base sur celui des State-based CRDTs. Partage donc les mêmes pré-requis :
 - États du type de données répliqué forment un sup-demi-treillis
 - Modifications locales entraînent une inflation de l'état
 - Possède une fonction de **merge**, permettant de fusionner deux états S et S' , et qui
 - Est associative, commutative et idempotente
 - Retourne S'' , la LUB de S et S' (c.-à-d. $\nexists S''' \cdot merge(S, S') < S''' < S''$)

Et bénéficie de son principal avantage : synchronisation possible entre deux pairs en fusionnant leur états, peu importe le nombre de modifications les séparant.

- Spécificité des Delta-based CRDTs est de proposer une synchronisation par différence d'états. Plutôt que de diffuser l'entièreté de l'état pour permettre aux autres pairs de se mettre à jour, idée est de seulement transmettre la partie de l'état ayant été mise à jour. Correspond à un élément irréductible du sup-demi-treillis. Permet ainsi de mettre en place une synchronisation en temps réel de manière efficace. Et d'utiliser la synchronisation par fusion d'états complets pour compenser les défaillances du réseau
- Ainsi, ce nouveau type de CRDTs semble allier le meilleur des deux mondes :
 - Absence de contrainte sur le réseau autre que la livraison à terme
 - Propagation possible en temps réel des modifications

Semble donc être une solution universelle :

- Utilisable peu importe la fiabilité réseau à disposition
- Empreinte réseau du même ordre de grandeur qu'un Op-based CRDT
- Utilisable peu importe la fréquence de synchronisation désirée

Pose la question de l'intérêt des autres types de CRDTs.

- Delta-based CRDT est un State-based CRDT dont on a identifié les éléments irréductibles et qui utilise ces derniers pour la propagation des modifications plutôt que l'état complet. Famille des State-based CRDTs semble donc rendue obsolète par celle des Delta-based CRDTs. À confirmer.
- Les Op-based CRDTs proposent une spécification différente du type répliqué de leur équivalent Delta-based, généralement plus simple. À première vue, famille des Op-based CRDTs semble donc avoir la simplicité comme avantage par rapport à celle des Delta-based CRDTs. S'agit d'un paramètre difficilement mesurable et auquel on peut objecter si on considère qu'un Op-based CRDT s'accompagne d'une couche livraison de messages, qui cache sa part de complexité. Intéressant d'étudier si la spécification différente des Op-based CRDTs présente d'autres avantages par rapport aux Delta-based CRDTs : performances (temps d'intégration des modifications, délai de convergence...), fonctionnalités spécifiques (composition, undo...)
- But serait de fournir des guidelines sur la famille de CRDT à adopter en fonction du cas d'utilisation.

2.2.5 Définition d'opérations supplémentaires pour fonctionnalités liées à l'édition collaborative

- Commentaires
- Suggestions

2.2.6 Conduction d'expériences utilisateurs d'édition collaborative

- Absence d'un dataset réel et réutilisable sur les sessions d'édition collaborative
- Généralement, expériences utilisent données d'articles de Wikipédia *Matthieu: TODO : Revoir références, mais me semble que c'est celui utilisé pour Logoot, LogootSplit et RGASplit entre autres.* Mais ces données correspondent à une exécution séquentielle, c.-à-d. aucune édition concurrente ne peut être réalisée avec le système de résolution de conflits de Wikipédia. *Matthieu: TODO : Me semble que Kleppmann a aussi utilisé et mis à disposition ses traces correspondant à la rédaction d'un de ses articles. Mais que cet article n'était rédigé que par lui. Peu de chances de présence d'édérations concurrentes. À retrouver et vérifier.*
- Inspiré par expériences de Claudia, pourrait mener des sessions d'édition collaborative sur des outils orchestrés pour produire ce dataset
- Devrait rendre ce dataset agnostique de l'approche choisie pour la résolution automatique de conflits
- Absence de retours sur les collaborations à grande échelle
- Comment on collabore lorsque plusieurs centaines d'utilisateur-rices ?

2.2.7 Comparaison des mécanismes de synchronisation

Serait intéressant de comparer à d'autres méthodes de synchronisation : mécanisme d'anti-entropie basé sur un Merkle Tree[30, 31, 32], synchronisation par états (state/delta-based CRDTs). Dans le cadre des Delta-based CRDTs, pourrait évaluer un protocole de diffusion épidémique des deltas comme celui proposé par SWIM[8].

2.2.8 Distance entre versions d'un document

- Est-ce que ça a vraiment du sens d'intégrer automatiquement des modifications ayant été généré sur une version du document distante de l'état actuel du document (voir distance de Hamming, Levenstein, String-to-string correction problem (Tichy et al))
- Jusqu'à quelle distance est-ce que la fusion automatique a encore du sens ? *Matthieu: NOTE : Peut connecter ça à la nécessité de conserver un chemin d'une époque à l'autre : si les opérations émises depuis cette époque ont probablement plus d'intérêt pour l'état actuel, couper l'arbre ?*

2.2.9 Contrôle d'accès

- Pour le moment, n'importe quel utilisateur ayant l'URL du document peut y accéder dans MUTE
- Pour des raisons de confidentialité, peut vouloir contrôler quels utilisateurs ont accès à un document
- Nécessite l'implémentation de liste de contrôle d'accès
- Mais s'agit d'une tâche complexe dans le cadre d'un système distribué
- Peut s'inspirer des travaux réalisés au sein de la communauté CRDTs [68, 69] pour cela

2.2.10 Détection et éviction de pairs malhonnêtes

- À l'heure actuelle, MUTE suppose qu'ensemble des collaborateurs honnêtes
- Vulnérable à plusieurs types d'attaques par des adversaires byzantins, tel que l'équivoque
- Ce type d'attaque peut provoquer des divergences durables et faire échouer des collaborations
- Dans [70, 5], ELVINGER propose un mécanisme permettant de maintenir des logs authentifiés dans un système distribué
- Les logs authentifiés permettent de mettre en lumière les comportements malveillants des adversaires et de borner le nombre d'actions malveillantes qu'ils peuvent effectuer avant d'être évincé
- Implémenter ce mécanisme permettrait de rendre compatible MUTE avec des environnements avec adversaires byzantins

- Nécessiterait tout de même de faire évoluer le CRDT pour résoudre les équivoques détectés

2.2.11 Vecteur *epoch-based*

- Comme présenté précédemment, nous utilisons plusieurs vecteurs pour représenter des données dans l'application MUTE
- Notamment pour le vecteur de version, utilisé pour respecter le modèle de livraison requis par le CRDT
- Et pour la liste des collaborateurs, utilisé pour offrir des informations nécessaires à la conscience de groupe aux utilisateurs
- Ces vecteurs sont maintenus localement par chacun des noeuds et sont échangés de manière périodique
- Cependant, la taille de ces vecteurs croît de manière linéaire au nombre de noeuds impliqués dans la collaboration
- Les systèmes P2P à large échelle sont sujets au *churn*
- Dans le cadre d'un tel système, ces structures croissent de manière non-bornée
- Ceci pose un problème de performances, notamment d'un point de vue consommation en bande-passante
- Cependant, même si on observe un grand nombre de pairs différents dans le cadre d'une collaboration à large échelle
- Intuition est qu'une collaboration repose en fait sur un petit noyau de collaborateurs principaux
- Et que majorité des collaborateurs se connectent de manière éphémère
- Serait intéressant de pouvoir réduire la taille des vecteurs en oubliant les collaborateurs éphémères
- Dynamo[30] tronque le vecteur de version lorsqu'il dépasse une taille seuil
- Conduit alors à une perte d'informations
- Pour la liste des collaborateurs, approche peut être adoptée (pas forcément gênant de limiter à 100 la taille de la liste)
- Mais pour vecteur de version, conduirait à une relivraison d'opérations déjà observées
- Approche donc pas applicable pour cette partie
- Autre approche possible est de réutiliser le système d'époque
- Idée serait de ACK un vecteur avec un changement d'époque
- Et de ne diffuser à partir de là que les différences
- Un mécanisme de transformation (une simple soustraction) permettrait d'obtenir le dot dans la nouvelle époque d'une opération concurrente au renommage
- Peut facilement mettre en place un mécanisme d'inversion du renommage (une simple addition) pour revenir à une époque précédente

- Et ainsi pouvoir circuler librement dans l'arbre des époques et gérer les opérations *rename* concurrentes
- Serait intéressant d'étudier si on peut aller plus loin dans le cadre de cette structure de données et notamment rendre commutatives les opérations de renommage concurrentes

2.2.12 Fusion de versions distantes d'un document collaboratif

2.2.13 Rôles et places des bots dans systèmes collaboratifs

- Stockage du document pour améliorer sa disponibilité
- Overleaf en P2P ?
- Comment réinsérer des bots dans la collaboration sans en faire des éléments centraux, sans créer des failles de confidentialité, et tout en rendant ces fonctionnalités accessibles ?

Annexe A

Algorithmes RENAMEID

Algorithme 1 Remaining functions to rename an identifier

```
function RENIDLESTHANFIRSTID(id, newFirstId)
  if id < newFirstId then
    return id
  else
    pos ← position(newFirstId)
    nId ← nodeId(newFirstId)
    nSeq ← nodeSeq(newFirstId)
    predNewFirstId ← new Id(pos, nId, nSeq, -1)

    return concat(predNewFirstId, id)
  end if
end function

function RENIDGREATERTHANLASTID(id, newLastId)
  if id < newLastId then
    return concat(newLastId, id)
  else
    return id
  end if
end function
```

Annexe B

Algorithmes REVERTRENAMEID

Algorithme 2 Remaining functions to revert an identifier renaming

```
function REVRENIDLESTHANNEWFIRSTID(id, firstId, newFirstId)
  predNewFirstId  $\leftarrow$  createIdFromBase(newFirstId, -1)
  if isPrefix(predNewFirstId, id) then
    tail  $\leftarrow$  getTail(id, 1)
    if tail < firstId then
      return tail
    else
       $\triangleright id$  has been inserted causally after the rename op
      offset  $\leftarrow$  getLastOffset(firstId)
      predFirstId  $\leftarrow$  createIdFromBase(firstId, offset)
      return concat(predFirstId, MAX_TUPLE, tail)
    end if
  else
    return id
  end if
end function

function REVRENIDGREATERTHANNEWLASTID(id, lastId)
  if id < lastId then
     $\triangleright id$  has been inserted causally after the rename op
    return concat(lastId, MIN_TUPLE, id)
  else if isPrefix(newLastId, id) then
    tail  $\leftarrow$  getTail(id, 1)
    if tail < lastId then
       $\triangleright id$  has been inserted causally after the rename op
      return concat(lastId, MIN_TUPLE, tail)
    else if tail < newLastId then
      return tail
    else
       $\triangleright id$  has been inserted causally after the rename op
      return id
    end if
  else
    return id
  end if
end function
```

Index

Voici un index

FiXme :

Notes :

- 10 : Matthieu : TODO : Insérer ref, 16
- 11 : Matthieu : TODO : Ajouter mention que OT a été abandonné à cause de cette contrainte même., 16
- 12 : Matthieu : TODO : Ajouter refs Scuttlebutt si applicable à Op-based, 17
- 13 : Matthieu : TODO : Ajouter refs, 18
- 14 : Matthieu : TODO : Vérifier que c'est bien le cas dans [33], 18
- 15 : Matthieu : TODO : Ajouter IPA, 20
- 16 : Matthieu : TODO : Revoir et ajouter Melda (PaPoC'22) si fitting, 20
- 17 : Matthieu : TODO : Ajouter Redis et Akka, 21
- 18 : Matthieu : TODO : Vérifier et ajouter l'article avec Digital Garden (PaPoC'22 ?) si fitting, 21
- 19 : Matthieu : TODO : Mentionner que WOOT ne propose pas de mécanisme pour purger les tombstones, par choix : souhaite mettre en place des fonctionnalités d'undo., 23
- 1 : Matthieu : TODO : Voir si angle écologique/réduction consommation d'énergie peut être pertinent., 1
- 20 : Matthieu : TODO : Ajouter limite de WOOT : complexité en temps de l'algorithme d'intégration des insertions distantes trop importante, 23
- 21 : Matthieu : TODO : Trouver exemple et le présenter., 24
- 22 : Matthieu : TODO : Ajouter quelques lignes sur efficacité de RGA. Meilleure complexité en temps pour l'intégration des opérations. [58, 59] montrent que RGA est optimal du point de vue complexité spatiale comme CRDT pour Séquence par élément basée sur l'approche tombstone. Implémenté dans Yjs et Automerge. , 25
- 23 : Matthieu : TODO : Insérer et présenter exemple, 26
- 24 : Matthieu : TODO : Figure arbre binaire d'identifiants., 27
- 25 : Matthieu : TODO : Figure arbre binaire d'identifiants avec désambiguaturs et mini-nodes., 27
- 26 : Matthieu : TODO : Figure exemple d'édition collaborative, 28
- 27 : Matthieu : NOTE : Ajouter LogootSplit de manière sommaire aussi à cet endroit ?, 29
- 28 : Matthieu : TODO : Autres Sequence CRDTs à considérer : Stringwise CRDT [61], Chronofold [67], 29
- 29 : Matthieu : TODO : Ajouter une relation d'ordre sur les tuples, 30
- 2 : Matthieu : TODO : Trouver et ajouter références, 3
- 30 : Matthieu : TODO : Définir la notion de base (et autres fonctions utiles sur les identifiants ? genre is-Prefix, concat, getTail...), 30
- 31 : Matthieu : TODO : Montrer que cet ensemble d'identifiants est un en-

- semble dense, 30
- 32 : Matthieu : TODO : indiquer que le couple $\rightarrow \text{nodeId}, \text{nodeSeq} \not\approx$ permet d'identifier de manière unique la base d'un bloc ou d'un identifiant, 32
- 33 : Matthieu : NOTE : Pourrait définir dans cette sous-section la notion de séquence bien-formée, 33
- 34 : Matthieu : QUESTION : Ajouter quelques lignes ici sur comment faire ça en pratique (Ajout d'un dot aux opérations, maintien d'un dot store au niveau de la couche livraison, vérification que dot pas encore présent dans dot store avant de passer opération à la structure de données) ? Ou je garde ça pour le chapitre sur MUTE ?, 34
- 35 : Matthieu : QUESTION : Même que pour la exactly-once delivery, est-ce que j'explique ici comment assurer cette contrainte plus en détails (Ajout des dots des opérations *insert* en dépendances de l'opération *remove*, vérification que dots présents dans dot store avant de passer l'opération *remove* à la structure de données) ou je garde ça pour le chapitre sur MUTE ?, 35
- 36 : Matthieu : TODO : Ajouter une phrase pour expliquer que la croissance des identifiants impacte aussi le temps d'intégration des modifications, 37
- 37 : Matthieu : TODO : Trouver référence sur la stabilité causale dans systèmes dynamiques, 37
- 38 : Matthieu : TODO : Revoir références, mais me semble que c'est celui utilisé pour Logoot, LogootSplit et RGASplit entre autres, 41
- 39 : Matthieu : TODO : Me semble que Kleppmann a aussi utilisé et mis à disposition ses traces correspondant à la rédaction d'un de ses articles. Mais que cet article n'était rédigé que par lui. Peu de chances de présence d'éditions concurrentes. À retrouver et vérifier. , 41
- 3 : Matthieu : TODO : Mentionner TP1 et TP2, 6
- 40 : Matthieu : NOTE : Peut connecter ça à la nécessité de conserver un chemin d'une époque à l'autre : si les opérations émises depuis cette époque ont probablement plus d'intérêt pour l'état actuel, couper l'arbre ?, 42
- 4 : Matthieu : TODO : Spécification faible et forte des séquences répliquées, 6
- 5 : Matthieu : TODO : insérer ref sticky model, 7
- 6 : Matthieu : TODO : Ajouter ref, 8
- 7 : Matthieu : TODO : Faire le lien avec les travaux de Burckhardt [19] et les MRDTs [20], 9
- 8 : Matthieu : TODO : Ajouter refs des horloges logiques plus intelligentes (Interval Tree Clock, Hybrid Clock...), 10
- 9 : Matthieu : TODO : Ajouter ref, 15
- FiXme (Matthieu) :
- Notes :
- 10 : TODO : Insérer ref, 16
- 11 : TODO : Ajouter mention que OT a été abandonné à cause de cette contrainte même., 16
- 12 : TODO : Ajouter refs Scuttlebutt si applicable à Op-based, 17
- 13 : TODO : Ajouter refs, 18
- 14 : TODO : Vérifier que c'est bien le cas dans [33], 18
- 15 : TODO : Ajouter IPA, 20
- 16 : TODO : Revoir et ajouter Melda (PaPoC'22) si fitting, 20
- 17 : TODO : Ajouter Redis et Akka, 21
- 18 : TODO : Vérifier et ajouter l'ar-

- ticle avec Digital Garden (PaPoC'22 ?) si fitting, 21
- 19 : TODO : Mentionner que WOOT ne propose pas de mécanisme pour purger les tombstones, par choix : souhaite mettre en place des fonctionnalités d'undo., 23
- 1 : TODO : Voir si angle écologique/réduction consommation d'énergie peut être pertinent., 1
- 20 : TODO : Ajouter limite de WOOT : complexité en temps de l'algorithme d'intégration des insertions distantes trop importante, 23
- 21 : TODO : Trouver exemple et le présenter., 24
- 22 : TODO : Ajouter quelques lignes sur efficacité de RGA. Meilleure complexité en temps pour l'intégration des opérations. [58, 59] montrent que RGA est optimal du point de vue complexité spatiale comme CRDT pour Séquence par élément basée sur l'approche tombstone. Implémenté dans Yjs et Automerge. , 25
- 23 : TODO : Insérer et présenter exemple, 26
- 24 : TODO : Figure arbre binaire d'identifiants., 27
- 25 : TODO : Figure arbre binaire d'identifiants avec désambiguateurs et mininodes., 27
- 26 : TODO : Figure exemple d'édition collaborative, 28
- 27 : NOTE : Ajouter LogootSplit de manière sommaire aussi à cet endroit ?, 29
- 28 : TODO : Autres Sequence CRDTs à considérer : String-wise CRDT [61], Chronofold [67], 29
- 29 : TODO : Ajouter une relation d'ordre sur les tuples, 30
- 2 : TODO : Trouver et ajouter références, 3
- 30 : TODO : Définir la notion de base (et autres fonctions utiles sur les identifiants ? genre isPrefix, concat, get-Tail...), 30
- 31 : TODO : Montrer que cet ensemble d'identifiants est un ensemble dense, 30
- 32 : TODO : indiquer que le couple $\neg \text{nodeId}, \text{nodeSeq} \not\equiv$ permet d'identifier de manière unique la base d'un bloc ou d'un identifiant, 32
- 33 : NOTE : Pourrait définir dans cette sous-section la notion de séquence bien-formée, 33
- 34 : QUESTION : Ajouter quelques lignes ici sur comment faire ça en pratique (Ajout d'un dot aux opérations, maintien d'un dot store au niveau de la couche livraison, vérification que dot pas encore présent dans dot store avant de passer opération à la structure de données) ? Ou je garde ça pour le chapitre sur MUTE ?, 34
- 35 : QUESTION : Même que pour la exactly-once delivery, est-ce que j'explique ici comment assurer cette contrainte plus en détails (Ajout des dots des opérations *insert* en dépendances de l'opération *remove*, vérification que dots présents dans dot store avant de passer l'opération *remove* à la structure de données) ou je garde ça pour le chapitre sur MUTE ?, 35
- 36 : TODO : Ajouter une phrase pour expliquer que la croissance des identifiants impacte aussi le temps d'intégration des modifications, 37
- 37 : TODO : Trouver référence sur la stabilité causale dans systèmes dynamiques, 37
- 38 : TODO : Revoir références, mais me semble que c'est celui utilisé pour Logoot, LogootSplit et RGASplit entre autres, 41
- 39 : TODO : Me semble que Klepp-

mann a aussi utilisé et mis à disposition ses traces correspondant à la rédaction d'un de ses articles. Mais que cet article n'était rédigé que par lui. Peu de chances de présence d'édicions concurrentes. À retrouver et vérifier. , 41

3 : TODO : Mentionner TP1 et TP2, 6

40 : NOTE : Peut connecter ça à la nécessité de conserver un chemin d'une époque à l'autre : si les opérations émises depuis cette époque ont probablement plus d'intérêt pour l'état actuel, couper l'arbre ?, 42

4 : TODO : Spécification faible et forte des séquences répliquées, 6

5 : TODO : insérer ref sticky model, 7

6 : TODO : Ajouter ref, 8

7 : TODO : Faire le lien avec les travaux de Burckhardt [19] et les MRDTs [20], 9

8 : TODO : Ajouter refs des horloges logiques plus intelligentes (Interval Tree Clock, Hybrid Clock...), 10

9 : TODO : Ajouter ref, 15

Bibliographie

- [1] Mihai LETIA, Nuno PREGUIÇA et Marc SHAPIRO. « Consistency without concurrency control in large, dynamic systems ». In : *LADIS 2009 - 3rd ACM SIGOPS International Workshop on Large Scale Distributed Systems and Middleware*. T. 44. Operating Systems Review 2. Big Sky, MT, United States : Assoc. for Computing Machinery, oct. 2009, p. 29–34. DOI : 10.1145/1773912.1773921. URL : <https://hal.inria.fr/hal-01248270>.
- [2] Marek ZAWIRSKI, Marc SHAPIRO et Nuno PREGUIÇA. « Asynchronous rebalancing of a replicated tree ». In : *Conférence Française en Systèmes d'Exploitation (CFSE)*. Saint-Malo, France, mai 2011, p. 12. URL : <https://hal.inria.fr/hal-01248197>.
- [3] Matthieu NICOLAS et al. « MUTE : A Peer-to-Peer Web-based Real-time Collaborative Editor ». In : *ECSCW 2017 - 15th European Conference on Computer-Supported Cooperative Work*. T. 1. Proceedings of 15th European Conference on Computer-Supported Cooperative Work - Panels, Posters and Demos 3. Sheffield, United Kingdom : EUSSET, août 2017, p. 1–4. DOI : 10.18420/ecscw2017_p5. URL : <https://hal.inria.fr/hal-01655438>.
- [4] Luc ANDRÉ et al. « Supporting Adaptable Granularity of Changes for Massive-Scale Collaborative Editing ». In : *International Conference on Collaborative Computing : Networking, Applications and Worksharing - CollaborateCom 2013*. Austin, TX, USA : IEEE Computer Society, oct. 2013, p. 50–59. DOI : 10.4108/icst.collaboratecom.2013.254123.
- [5] Victorien ELVINGER. « Réplication sécurisée dans les infrastructures pair-à-pair de collaboration ». Theses. Université de Lorraine, juin 2021. URL : <https://hal.univ-lorraine.fr/tel-03284806>.
- [6] Hoang-Long NGUYEN, Claudia-Lavinia IGNAT et Olivier PERRIN. « Trusternity : Auditing Transparent Log Server with Blockchain ». In : *Companion of the The Web Conference 2018*. Lyon, France, avr. 2018. DOI : 10.1145/3184558.3186938. URL : <https://hal.inria.fr/hal-01883589>.
- [7] Hoang-Long NGUYEN et al. « Blockchain-Based Auditing of Transparent Log Servers ». In : *32th IFIP Annual Conference on Data and Applications Security and Privacy (DBSec)*. Sous la dir. de Florian KERSCHBAUM et Stefano PARABOSCHI. T. LNCS-10980. Data and Applications Security and Privacy XXXII. Part 1 : Administration. Bergamo, Italy : Springer International Publishing, juil. 2018, p. 21–37. DOI : 10.1007/978-3-319-95729-6_2. URL : <https://hal.archives-ouvertes.fr/hal-01917636>.

- [8] A. DAS, I. GUPTA et A. MOTIVALA. « SWIM : scalable weakly-consistent infection-style process group membership protocol ». In : *Proceedings International Conference on Dependable Systems and Networks*. 2002, p. 303–312. DOI : 10.1109/DSN.2002.1028914.
- [9] Armon DADGAR, James PHILLIPS et Jon CURREY. « Lifeguard : Local health awareness for more accurate failure detection ». In : *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE. 2018, p. 22–25.
- [10] Brice NÉDELEC et al. « An adaptive peer-sampling protocol for building networks of browsers ». In : *World Wide Web* 21.3 (2018), p. 629–661.
- [11] Mike BURMESTER et Yvo DESMEDT. « A secure and efficient conference key distribution system ». In : *Advances in Cryptology — EUROCRYPT’94*. Sous la dir. d’Alfredo DE SANTIS. Berlin, Heidelberg : Springer Berlin Heidelberg, 1995, p. 275–286. ISBN : 978-3-540-44717-7.
- [12] Yasushi SAITO et Marc SHAPIRO. « Optimistic Replication ». In : *ACM Comput. Surv.* 37.1 (mar. 2005), p. 42–81. ISSN : 0360-0300. DOI : 10.1145/1057977.1057980. URL : <https://doi.org/10.1145/1057977.1057980>.
- [13] D. ABADI. « Consistency Tradeoffs in Modern Distributed Database System Design : CAP is Only Part of the Story ». In : *Computer* 45.2 (2012), p. 37–42. DOI : 10.1109/MC.2012.33.
- [14] Chengzheng SUN et al. « Achieving Convergence, Causality Preservation, and Intention Preservation in Real-Time Cooperative Editing Systems ». In : *ACM Trans. Comput.-Hum. Interact.* 5.1 (mar. 1998), p. 63–108. ISSN : 1073-0516. DOI : 10.1145/274444.274447. URL : <https://doi.org/10.1145/274444.274447>.
- [15] Leslie LAMPORT. « Time, Clocks, and the Ordering of Events in a Distributed System ». In : *Commun. ACM* 21.7 (juil. 1978), p. 558–565. ISSN : 0001-0782. DOI : 10.1145/359545.359563. URL : <https://doi.org/10.1145/359545.359563>.
- [16] Marc SHAPIRO et al. « Conflict-Free Replicated Data Types ». In : *Proceedings of the 13th International Symposium on Stabilization, Safety, and Security of Distributed Systems*. SSS 2011. 2011, p. 386–400. DOI : 10.1007/978-3-642-24550-3_29.
- [17] Nuno M. PREGUIÇA, Carlos BAQUERO et Marc SHAPIRO. « Conflict-free Replicated Data Types (CRDTs) ». In : *CoRR* abs/1805.06358 (2018). arXiv : 1805.06358. URL : <http://arxiv.org/abs/1805.06358>.
- [18] Nuno M. PREGUIÇA. « Conflict-free Replicated Data Types : An Overview ». In : *CoRR* abs/1806.10254 (2018). arXiv : 1806.10254. URL : <http://arxiv.org/abs/1806.10254>.

-
- [19] Sebastian BURCKHARDT et al. « Replicated Data Types : Specification, Verification, Optimality ». In : *Proceedings of the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. POPL '14. San Diego, California, USA : Association for Computing Machinery, 2014, p. 271–284. ISBN : 9781450325448. DOI : 10.1145/2535838.2535848. URL : <https://doi.org/10.1145/2535838.2535848>.
- [20] Gowtham KAKI et al. « Mergeable Replicated Data Types ». In : *Proc. ACM Program. Lang.* 3.OOPSLA (oct. 2019). DOI : 10.1145/3360580. URL : <https://doi.org/10.1145/3360580>.
- [21] Paul R JOHNSON et Robert THOMAS. *RFC0677 : Maintenance of duplicate databases*. RFC Editor, 1975.
- [22] Weihai YU et Sigbjørn ROSTAD. « A Low-Cost Set CRDT Based on Causal Lengths ». In : *Proceedings of the 7th Workshop on Principles and Practice of Consistency for Distributed Data*. New York, NY, USA : Association for Computing Machinery, 2020. ISBN : 9781450375245. URL : <https://doi.org/10.1145/3380787.3393678>.
- [23] Marc SHAPIRO et al. *A comprehensive study of Convergent and Commutative Replicated Data Types*. Research Report RR-7506. Inria – Centre Paris-Rocquencourt ; INRIA, jan. 2011, p. 50. URL : <https://hal.inria.fr/inria-00555588>.
- [24] Carlos BAQUERO, Paulo Sérgio ALMEIDA et Ali SHOKER. « Making Operation-Based CRDTs Operation-Based ». In : *Proceedings of the First Workshop on Principles and Practice of Eventual Consistency*. PaPEC '14. Amsterdam, The Netherlands : Association for Computing Machinery, 2014. ISBN : 9781450327169. DOI : 10.1145/2596631.2596632. URL : <https://doi.org/10.1145/2596631.2596632>.
- [25] Carlos BAQUERO, Paulo Sergio ALMEIDA et Ali SHOKER. *Pure Operation-Based Replicated Data Types*. 2017. arXiv : 1710.04469 [cs.DC].
- [26] Paulo Sérgio ALMEIDA, Ali SHOKER et Carlos BAQUERO. « Efficient State-Based CRDTs by Delta-Mutation ». In : *Networked Systems*. Sous la dir. d’Ahmed BOUAJJANI et Hugues FAUCONNIER. Cham : Springer International Publishing, 2015, p. 62–76. ISBN : 978-3-319-26850-7.
- [27] Paulo Sérgio ALMEIDA, Ali SHOKER et Carlos BAQUERO. « Delta state replicated data types ». In : *Journal of Parallel and Distributed Computing* 111 (jan. 2018), p. 162–173. ISSN : 0743-7315. DOI : 10.1016/j.jpdc.2017.08.003. URL : <http://dx.doi.org/10.1016/j.jpdc.2017.08.003>.
- [28] Ravi PRAKASH, Michel RAYNAL et Mukesh SINGHAL. « An Adaptive Causal Ordering Algorithm Suited to Mobile Computing Environments ». In : *Journal of Parallel and Distributed Computing* 41.2 (1997), p. 190–204. ISSN : 0743-7315. DOI : <https://doi.org/10.1006/jpdc.1996.1300>. URL : <https://www.sciencedirect.com/science/article/pii/S0743731596913003>.
- [29] D. S. PARKER et al. « Detection of Mutual Inconsistency in Distributed Systems ». In : *IEEE Trans. Softw. Eng.* 9.3 (mai 1983), p. 240–247. ISSN : 0098-5589. DOI : 10.1109/TSE.1983.236733. URL : <https://doi.org/10.1109/TSE.1983.236733>.

- [30] Giuseppe DECANDIA et al. « Dynamo : Amazon’s highly available key-value store ». In : *ACM SIGOPS operating systems review* 41.6 (2007), p. 205–220.
- [31] Nico KRUBER, Maik LANGE et Florian SCHINTKE. « Approximate Hash-Based Set Reconciliation for Distributed Replica Repair ». In : *2015 IEEE 34th Symposium on Reliable Distributed Systems (SRDS)*. 2015, p. 166–175. DOI : 10.1109/SRDS.2015.30.
- [32] Ricardo Jorge Tomé GONÇALVES et al. « DottedDB : Anti-Entropy without Merkle Trees, Deletes without Tombstones ». In : *2017 IEEE 36th Symposium on Reliable Distributed Systems (SRDS)*. 2017, p. 194–203. DOI : 10.1109/SRDS.2017.28.
- [33] Jim BAUWENS et Elisa Gonzalez BOIX. « Improving the Reactivity of Pure Operation-Based CRDTs ». In : *Proceedings of the 8th Workshop on Principles and Practice of Consistency for Distributed Data*. PaPoC ’21. Online, United Kingdom : Association for Computing Machinery, 2021. ISBN : 9781450383387. DOI : 10.1145/3447865.3457968. URL : <https://doi.org/10.1145/3447865.3457968>.
- [34] V. ENES et al. « Efficient Synchronization of State-Based CRDTs ». In : *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. 2019, p. 148–159. DOI : 10.1109/ICDE.2019.00022.
- [35] Petru NICOLAESCU et al. « Yjs : A Framework for Near Real-Time P2P Shared Editing on Arbitrary Data Types ». In : *15th International Conference on Web Engineering*. ICWE 2015. Springer LNCS volume 9114, juin 2015, p. 675–678. DOI : 10.1007/978-3-319-19890-3_55. URL : <http://dbis.rwth-aachen.de/~derntl/papers/preprints/icwe2015-preprint.pdf>.
- [36] Petru NICOLAESCU et al. « Near Real-Time Peer-to-Peer Shared Editing on Extensible Data Types ». In : *19th International Conference on Supporting Group Work*. GROUP 2016. ACM, nov. 2016, p. 39–49. DOI : 10.1145/2957276.2957310.
- [37] YJS. *Yjs : A CRDT framework with a powerful abstraction of shared data*. URL : <https://github.com/yjs/yjs>.
- [38] Martin KLEPPMANN et Alastair R. BERESFORD. « A Conflict-Free Replicated JSON Datatype ». In : *IEEE Transactions on Parallel and Distributed Systems* 28.10 (oct. 2017), p. 2733–2746. ISSN : 1045-9219. DOI : 10.1109/tpds.2017.2697382. URL : <http://dx.doi.org/10.1109/TPDS.2017.2697382>.
- [39] AUTOMERGE. *Automerge : data structures for building collaborative applications in Javascript*. URL : <https://github.com/automerge/automerge>.
- [40] Christopher MEIKLEJOHN et Peter VAN ROY. « Lasp : A Language for Distributed, Coordination-free Programming ». In : *17th International Symposium on Principles and Practice of Declarative Programming*. PPDP 2015. ACM, juil. 2015, p. 184–195. DOI : 10.1145/2790449.2790525.
- [41] Kevin DE PORRE et al. « CScript : A distributed programming language for building mixed-consistency applications ». In : *Journal of Parallel and Distributed Computing volume 144* (oct. 2020), p. 109–123. ISSN : 0743-7315. DOI : 10.1016/j.jpdc.2020.05.010.

-
- [42] RIAK. *Riak KV*. URL : <http://riak.com/>.
 - [43] The SyncFree CONSORTIUM. *AntidoteDB : A planet scale, highly available, transactional database*. URL : <http://antidoteDB.eu/>.
 - [44] C. WU et al. « Anna : A KVS for Any Scale ». In : *IEEE Transactions on Knowledge and Data Engineering* 33.2 (2021), p. 344–358. DOI : 10.1109/TKDE.2019.2898401.
 - [45] CONCORDANT. *Concordant*. URL : <http://www.concordant.io/>.
 - [46] Weihai YU et Claudia-Lavinia IGNAT. « Conflict-Free Replicated Relations for Multi-Synchronous Database Management at Edge ». In : *IEEE International Conference on Smart Data Services, 2020 IEEE World Congress on Services*. Beijing, China, oct. 2020. URL : <https://hal.inria.fr/hal-02983557>.
 - [47] Martin KLEPPMANN et al. « Local-First Software : You Own Your Data, in Spite of the Cloud ». In : *Proceedings of the 2019 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*. Onward! 2019. Athens, Greece : Association for Computing Machinery, 2019, p. 154–178. ISBN : 9781450369954. DOI : 10.1145/3359591.3359737. URL : <https://doi.org/10.1145/3359591.3359737>.
 - [48] Peter van HARDENBERG et Martin KLEPPMANN. « PushPin : Towards Production-Quality Peer-to-Peer Collaboration ». In : *7th Workshop on Principles and Practice of Consistency for Distributed Data*. PaPoC 2020. ACM, avr. 2020. DOI : 10.1145/3380787.3393683.
 - [49] Brice NÉDELEC, Pascal MOLLI et Achour MOSTEFAOUI. « CRATE : Writing Stories Together with our Browsers ». In : *25th International World Wide Web Conference*. WWW 2016. ACM, avr. 2016, p. 231–234. DOI : 10.1145/2872518.2890539.
 - [50] Gérald OSTER et al. « Data Consistency for P2P Collaborative Editing ». In : *ACM Conference on Computer-Supported Cooperative Work - CSCW 2006*. Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. Banff, Alberta, Canada : ACM Press, nov. 2006, p. 259–268. URL : <https://hal.inria.fr/inria-00108523>.
 - [51] Gérald OSTER et al. « Tombstone Transformation Functions for Ensuring Consistency in Collaborative Editing Systems ». In : *2006 International Conference on Collaborative Computing : Networking, Applications and Worksharing*. 2006, p. 1–10. DOI : 10.1109/COLCOM.2006.361867.
 - [52] Marc SHAPIRO et Nuno PREGUIÇA. *Designing a commutative replicated data type*. Research Report RR-6320. INRIA, 2007. URL : <https://hal.inria.fr/inria-00177693>.
 - [53] Stéphane WEISS, Pascal URSO et Pascal MOLLI. « Wooki : a P2P Wiki-based Collaborative Writing Tool ». In : t. 4831. Déc. 2007. ISBN : 978-3-540-76992-7. DOI : 10.1007/978-3-540-76993-4_42.

- [54] Mehdi AHMED-NACER et al. « Evaluating CRDTs for Real-time Document Editing ». In : *11th ACM Symposium on Document Engineering*. Sous la dir. d'ACM. Mountain View, California, United States, sept. 2011, p. 103–112. DOI : 10.1145/2034691.2034717. URL : <https://hal.inria.fr/inria-00629503>.
- [55] Ben SHNEIDERMAN. « Response Time and Display Rate in Human Performance with Computers ». In : *ACM Comput. Surv.* 16.3 (sept. 1984), p. 265–285. ISSN : 0360-0300. DOI : 10.1145/2514.2517. URL : <https://doi.org/10.1145/2514.2517>.
- [56] Caroline JAY, Mashhuda GLENCROSS et Roger HUBBOLD. « Modeling the Effects of Delayed Haptic and Visual Feedback in a Collaborative Virtual Environment ». In : *ACM Trans. Comput.-Hum. Interact.* 14.2 (août 2007), 8-es. ISSN : 1073-0516. DOI : 10.1145/1275511.1275514. URL : <https://doi.org/10.1145/1275511.1275514>.
- [57] Hyun-Gul ROH et al. « Replicated abstract data types : Building blocks for collaborative applications ». In : *Journal of Parallel and Distributed Computing* 71.3 (2011), p. 354–368. ISSN : 0743-7315. DOI : <https://doi.org/10.1016/j.jpdc.2010.12.006>. URL : <http://www.sciencedirect.com/science/article/pii/S0743731510002716>.
- [58] Hagit ATTIYA et al. « Specification and Complexity of Collaborative Text Editing ». In : *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing*. PODC '16. Chicago, Illinois, USA : Association for Computing Machinery, 2016, p. 259–268. ISBN : 9781450339643. DOI : 10.1145/2933057.2933090. URL : <https://doi.org/10.1145/2933057.2933090>.
- [59] Hagit ATTIYA et al. « Specification and space complexity of collaborative text editing ». In : *Theoretical Computer Science* 855 (2021), p. 141–160. ISSN : 0304-3975. DOI : <https://doi.org/10.1016/j.tcs.2020.11.046>. URL : <http://www.sciencedirect.com/science/article/pii/S0304397520306952>.
- [60] Loïck BRIOT, Pascal URSO et Marc SHAPIRO. « High Responsiveness for Group Editing CRDTs ». In : *ACM International Conference on Supporting Group Work*. Sanibel Island, FL, United States, nov. 2016. DOI : 10.1145/2957276.2957300. URL : <https://hal.inria.fr/hal-01343941>.
- [61] Weihai YU. « A String-Wise CRDT for Group Editing ». In : *Proceedings of the 17th ACM International Conference on Supporting Group Work*. GROUP '12. Sanibel Island, Florida, USA : Association for Computing Machinery, 2012, p. 141–144. ISBN : 9781450314862. DOI : 10.1145/2389176.2389198. URL : <https://doi.org/10.1145/2389176.2389198>.
- [62] Martin KLEPPMANN et al. « Interleaving Anomalies in Collaborative Text Editors ». In : *Proceedings of the 6th Workshop on Principles and Practice of Consistency for Distributed Data*. PaPoC '19. Dresden, Germany : Association for Computing Machinery, 2019. ISBN : 9781450362764. DOI : 10.1145/3301419.3323972. URL : <https://doi.org/10.1145/3301419.3323972>.
- [63] Matthew WEIDNER. *There Are No Doubly Non-Interleaving List CRDTs*. URL : https://mattweidner.com/assets/pdf/List_CRDT_Non_Interleaving.pdf.

-
- [64] Nuno PREGUICA et al. « A Commutative Replicated Data Type for Cooperative Editing ». In : *2009 29th IEEE International Conference on Distributed Computing Systems*. Juin 2009, p. 395–403. DOI : 10.1109/ICDCS.2009.20.
- [65] Stéphane WEISS, Pascal URSO et Pascal MOLLI. « Logoot : A Scalable Optimistic Replication Algorithm for Collaborative Editing on P2P Networks ». In : *Proceedings of the 29th International Conference on Distributed Computing Systems - ICDCS 2009*. Montreal, QC, Canada : IEEE Computer Society, juin 2009, p. 404–412. DOI : 10.1109/ICDCS.2009.75. URL : <http://doi.ieeecomputersociety.org/10.1109/ICDCS.2009.75>.
- [66] Stéphane WEISS, Pascal URSO et Pascal MOLLI. « Logoot-Undo : Distributed Collaborative Editing System on P2P Networks ». In : *IEEE Transactions on Parallel and Distributed Systems* 21.8 (août 2010), p. 1162–1174. DOI : 10.1109/TPDS.2009.173. URL : <https://hal.archives-ouvertes.fr/hal-00450416>.
- [67] Victor GRISHCHENKO et Mikhail PATRAKEEV. « Chronofold : A Data Structure for Versioned Text ». In : *Proceedings of the 7th Workshop on Principles and Practice of Consistency for Distributed Data*. PaPoC '20. Heraklion, Greece : Association for Computing Machinery, 2020. ISBN : 9781450375245. DOI : 10.1145/3380787.3393680. URL : <https://doi.org/10.1145/3380787.3393680>.
- [68] Elena YANAKIEVA et al. « Access Control Conflict Resolution in Distributed File Systems Using CRDTs ». In : *Proceedings of the 8th Workshop on Principles and Practice of Consistency for Distributed Data*. PaPoC '21. Online, United Kingdom : Association for Computing Machinery, 2021. ISBN : 9781450383387. DOI : 10.1145/3447865.3457970. URL : <https://doi.org/10.1145/3447865.3457970>.
- [69] Pierre-Antoine RAULT, Claudia-Lavinia IGNAT et Olivier PERRIN. « Distributed Access Control for Collaborative Applications Using CRDTs ». In : *Proceedings of the 9th Workshop on Principles and Practice of Consistency for Distributed Data*. PaPoC '22. Rennes, France : Association for Computing Machinery, 2022, p. 33–38. ISBN : 9781450392563. DOI : 10.1145/3517209.3524826. URL : <https://doi.org/10.1145/3517209.3524826>.
- [70] Victorien ELVINGER, G  rald OSTER et Fran  ois CHAROY. « Prunable Authenticated Log and Authenticable Snapshot in Distributed Collaborative Systems ». In : *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*. IEEE. 2018, p. 156–165.

Résumé

Afin d'assurer leur haute disponibilité, les systèmes distribués à large échelle se doivent de répliquer leurs données tout en minimisant les coordinations nécessaires entre noeuds. Pour concevoir de tels systèmes, la littérature et l'industrie adoptent de plus en plus l'utilisation de types de données répliquées sans conflits (CRDTs). Les CRDTs sont des types de données qui offrent des comportements similaires aux types existants, tel l'Ensemble ou la Séquence. Ils se distinguent cependant des types traditionnels par leur spécification, qui supporte nativement les modifications concurrentes. À cette fin, les CRDTs incorporent un mécanisme de résolution de conflits au sein de leur spécification.

Afin de résoudre les conflits de manière déterministe, les CRDTs associent généralement des identifiants aux éléments stockés au sein de la structure de données. Les identifiants doivent respecter un ensemble de contraintes en fonction du CRDT, telles que l'unicité ou l'appartenance à un ordre dense. Ces contraintes empêchent de borner la taille des identifiants. La taille des identifiants utilisés croît alors continuellement avec le nombre de modifications effectuées, aggravant le surcoût lié à l'utilisation des CRDTs par rapport aux structures de données traditionnelles. Le but de cette thèse est de proposer des solutions pour pallier ce problème.

Nous présentons dans cette thèse deux contributions visant à répondre à ce problème : (i) Un nouveau CRDT pour Séquence, *RenamableLogootSplit*, qui intègre un mécanisme de renommage à sa spécification. Ce mécanisme de renommage permet aux noeuds du système de réattribuer des identifiants de taille minimale aux éléments de la séquence. Cependant, cette première version requiert une coordination entre les noeuds pour effectuer un renommage. L'évaluation expérimentale montre que le mécanisme de renommage permet de réinitialiser à chaque renommage le surcoût lié à l'utilisation du CRDT. (ii) Une seconde version de *RenamableLogootSplit* conçue pour une utilisation dans un système distribué. Cette nouvelle version permet aux noeuds de déclencher un renommage sans coordination préalable. L'évaluation expérimentale montre que cette nouvelle version présente un surcoût temporaire en cas de renommages concurrents, mais que ce surcoût est à terme.

Mots-clés: CRDTs, édition collaborative en temps réel, cohérence à terme, optimisation mémoire, performance

Abstract

Keywords: CRDTs, real-time collaborative editing, eventual consistency, memory-wise optimisation, performance

