

Matthieu Nicolas

DOCTORANT EN INFORMATIQUE - RÉPLICATION DE DONNÉES OPTIMISTE DANS LES SYSTÈMES PAIR-À-PAIR

☎ +33 6 75 98 34 40 | ✉ matthieu.nicolas@univ-lorraine.fr | 🌐 MatthieuNICOLAS

Déroulement de carrière

DOCTORANT ATTACHÉ TEMPORAIRE D'ENSEIGNEMENT ET DE RECHERCHE

Université de Lorraine, Polytech Nancy

Septembre 2020 - Août 2022

Nancy

DOCTORANT CONTRACTUEL

Université de Lorraine, Loria, équipe Coast

Octobre 2017 - Août 2020

Nancy

- **Intitulé :** (Ré)Identification efficace dans les types de données répliquées sans conflit (CRDTs)
- **Mots clés :** systèmes distribués, pair-à-pair, réplication optimiste, CRDTs, performances
- **Directeur de thèse :** Pr. Olivier Perrin
- **Co-directeur de thèse :** Dr. G rald Oster

Dans le cadre de mes travaux de recherche, j tudie et travaille sur les Conflict-free Replicated Data Types (CRDTs). Les CRDTs sont de nouvelles sp cifications des types abstraits de donn es, tels que l'*Ensemble* ou la *S quence*. Contrairement aux sp cifications traditionnelles, les CRDTs sont con us pour supporter nativement les modifications concurrentes. Pour ce faire, ces structures de donn es int grent un m canisme de r solution de conflits directement au sein de leur sp cification. Cette sp cificit  rend les CRDTs particuli rement adapt s pour concevoir des syst mes distribu s hautement disponibles dans lesquels diff rents noeuds r pliquent et modifient une m me donn e sans aucune coordination.

Pour r soudre les conflits de mani re d terministe, les CRDTs utilisent g n ralement des identifiants qu'ils associent aux  l ments stock s au sein de la structure de donn es. Cependant, selon le type de CRDT, les identifiants doivent respecter un ensemble de contraintes telles qu' tre unique ou appartenir   un espace dense. Dans certains cas, ces contraintes emp chent de borner la taille des identifiants. La taille des identifiants cro t alors continuellement avec le nombre de modifications effectu es.

Ces identifiants repr sentent donc un surco t li    l'utilisation des CRDTs par rapport aux structures de donn es traditionnelles. Ce surco t d courage l'adoption des CRDTs dans les syst mes distribu s. Le but de cette th se est de proposer des solutions pour pallier ce probl me. L'approche que nous proposons consiste   int grer un m canisme de renommage au sein des CRDTs. Ce m canisme a pour but de permettre aux diff rents noeuds de renommer les identifiants afin de r duire leur taille, tout en respectant les contraintes impos es aux CRDTs. En particulier, le renommage doit se faire sans aucune coordination entre les noeuds. Afin de valider cette approche, nous avons con u, impl ment , et  valu  un tel m canisme pour *LogootSplit*, un CRDT souffrant particuli rement du probl me de croissance des identifiants. Ces travaux ont conduit   la conception d'un nouveau CRDT, *RenamableLogootSplit*, que j'ai pr sent  au cours du workshop PaPoC'20 puis qui a fait l'objet d'une publication dans le journal scientifique IEEE Transactions on Parallel and Distributed Systems (TPDS) en 2022.

Publications

Matthieu Nicolas, Gerald Oster, and Olivier Perrin. "Efficient Renaming in Sequence CRDTs". In: *IEEE Transactions on Parallel and Distributed Systems* 33.12 (Dec. 2022), pp. 3870–3885. DOI: [10.1109/TPDS.2022.3172570](https://doi.org/10.1109/TPDS.2022.3172570). URL: <https://hal.inria.fr/hal-03772633>

Matthieu Nicolas, G rald Oster, and Olivier Perrin. "Efficient Renaming in Sequence CRDTs". In: *7th Workshop on Principles and Practice of Consistency for Distributed Data (PaPoC'20)*. Heraklion, Greece, Apr. 2020. URL: <https://hal.inria.fr/hal-02526724>

Matthieu Nicolas. "Efficient renaming in CRDTs". In: *Middleware 2018 - 19th ACM/IFIP International Middleware Conference (Doctoral Symposium)*. Rennes, France, Dec. 2018. URL: <https://hal.inria.fr/hal-01932552>

ING NIEUR RECHERCHE & D VELOPPEMENT

INRIA,  quipe Coast

Septembre 2014 – Septembre 2017

Nancy

- **Projets :** OpenPaaS::NG et PLM

OpenPaaS::NG

Ce projet avait pour objectif la r alisation d'un r seau social d'entreprise open-source incorporant une suite d'applications collaboratives pair- -pair de bureautique. Le but  tait ainsi de proposer une alternative viable et libre   des solutions telles que Google Apps. Ce projet fut r alis  en collaboration avec l' quipe DaSciM (Data Science and Mining) du laboratoire d'informatique de l' cole Polytechnique, Linagora, XWiki SAS et Nexedi.

Dans le cadre de ce projet, l' quipe COAST travaillait sur la f d ration interorganisationnelle de syst mes pair- -pair et sur la s curisation des  changes de donn es dans ce type de collaboration. De plus, elle apportait son expertise sur les m canismes de r plication de donn es et de

cohérence à terme dans les systèmes distribués.

C'est sur ce dernier point que portaient les tâches que j'ai effectuées dans le cadre de ce projet. Afin d'être validés, ces travaux ont été intégrés dans **MUTE** (<https://www.coedit.re>), la plateforme de démonstration de l'équipe.

- Maintenance de l'implémentation de *LogootSplit*
- Étude de la littérature sur les types de données répliquées sans conflits existants et leurs cas d'utilisation
- Développement et intégration d'un système d'anti-entropie

Publications

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*. Vol. 1. Proceedings of 15th European Conference on Computer-Supported Cooperative Work - Panels, Posters and Demos 3. Sheffield, United Kingdom: EUSSET, Aug. 2017, pp. 1–4. DOI: **10.18420/ecscw2017_p5**. URL: **https://hal.inria.fr/hal-01655438**

ADT PLM

La **PLM** (<http://people.irisa.fr/Martin.Quinson/Teaching/PLM/>) est un environnement d'apprentissage de la programmation libre et ouvert. Développé par Gérard Oster et Martin Quinson, il permet d'explorer différents aspects de l'algorithmique au travers d'exercices interactifs.

Le but de ce projet était de faire évoluer cet outil en une plateforme expérimentale pour l'enseignement de la programmation informatique. Pour cela, un mécanisme de capture des traces d'utilisation des apprenant devait être intégré afin de générer un corpus de données. Ce corpus, mis à disposition de chercheurs, devait permettre la conduite de travaux de recherche, tel que la conception d'outils d'aide automatique à l'apprentissage. Un second objectif ce projet était d'effectuer le portage de l'outil, jusqu'à alors disponible sous la forme d'une application lourde Java, en une application web afin de le rendre accessible au plus grand nombre.

Mes travaux se sont principalement focalisés sur la réalisation de ce portage. Ce changement important de type d'application a entraîné l'apparition de plusieurs problématiques auxquelles il a fallu apporter des solutions.

- Implémentation et test du mécanisme de capture des traces d'utilisation
- Conception et mise en place d'une architecture distribuée assurant le passage à l'échelle de l'application
- Isolation de l'exécution du code des apprenants
- Déploiement et supervision d'une application multi-composants

STAGE ÉLÈVE-INGÉNIEUR

Avril 2014 – Août 2014

Université de Lorraine, équipe Coast

Nancy

- **Intitulé :** Réalisation d'une plateforme d'édition collaborative

Issue des travaux sur l'édition collaborative, une nouvelle famille d'algorithmes de réplcation des données et de maintien de la cohérence à terme est apparue récemment : l'approche Conflict-free Replicated Data Type (CRDT). Cette nouvelle famille d'algorithme répond à plusieurs des limites constatées chez les autres approches existantes, notamment concernant la capacité de passage à l'échelle.

L'équipe Coast, travaillant sur ce domaine de recherche, a proposé un nouvel algorithme de cette famille : *LogootSplit*.

Afin d'illustrer et de mettre en valeur les travaux de l'équipe sur cette approche, ma tâche a été de concevoir et de développer un éditeur collaboratif temps réel se basant sur cet algorithme.

- Implémentation sous forme de librairie de *LogootSplit*
- Conception et développement de **MUTE** (<https://www.coedit.re>), un éditeur collaboratif temps réel en ligne reposant sur cette librairie

STAGE ÉLÈVE-TECHNICIEN

Avril 2011 – Juin 2011

École Polytechnique de Montréal

Montréal, Canada

- **Intitulé :** Développement d'un outil d'analyse d'algorithmes d'édition collaborative

Les outils d'édition collaborative existants reposent majoritairement sur une famille spécifique d'algorithmes pour assurer le maintien de la cohérence à terme : les transformées opérationnelles.

Deux propriétés de convergence *TP1* et *TP2* existent et permettent de garantir la correction de ces algorithmes.

L'objectif de ce stage était de réaliser un outil permettant de vérifier automatiquement le respect de ces propriétés pour un algorithme donné.

- Implémentation de plusieurs algorithmes issus de la famille des transformées opérationnelles
- Développement de l'outil permettant de vérifier les propriétés de convergences *TP1* et *TP2* pour les algorithmes implémentés.

Diplômes

DIPLÔME D'INGÉNIEUR TELECOM NANCY, SPÉCIALITÉ INGÉNIERIE DU LOGICIEL

TELECOM Nancy

2011 – 2014

Nancy

DIPLÔME UNIVERSITAIRE DE TECHNOLOGIE, SPÉCIALITÉ INFORMATIQUE

IUT de Metz

2009 – 2011

Metz

Enseignement

En parallèle de mon doctorat, j'ai effectué diverses missions d'enseignements au sein des différentes composantes de l'Université de Lorraine. Ainsi, j'ai effectué un service de demi-ATER puis d'ATER complet à Polytech Nancy au cours des deux dernières années. Auparavant, j'avais assuré une charge d'enseignement en tant que DCCE à l'IUT Nancy Charlemagne et effectué plusieurs vacations à TELECOM Nancy ainsi qu'à la Faculté des Sciences et Technologies de Nancy. L'ensemble de mes activités d'enseignement est présenté ci-dessous et récapitulé dans la Table 1. Au total, je comptabilise **542h équivalent TD** d'enseignement.

Année universitaire	Licence 1	Licence 2			Licence 3				Master 1	Master 2		eq. TD
	TD	CM	TD	TP	EI	CM	TD	TP	CM	CM	TD	
2014 – 2015	-	-	-	-	-	-	-	30 h	-	-	-	20 h
2015 – 2016	-	-	-	-	-	-	-	30 h	-	-	-	20 h
2016 – 2017	-	-	-	18 h	-	-	-	30 h	-	-	-	32 h
2017 – 2018	42 h	-	-	-	-	-	-	24 h	-	-	-	58 h
2018 – 2019	32 h	-	-	-	24 h	-	-	-	-	-	-	62 h
2019 – 2020	32 h	-	-	-	24 h	-	-	-	-	-	-	62 h
2020 – 2021	44 h	-	-	-	-	12 h	34 h	-	-	-	-	96 h
2021 – 2022	-	12 h	66 h	-	-	18 h	42 h	-	6 h	12 h	12 h	192 h
Total	150 h	12 h	66 h	18 h	48 h	30 h	76 h	114 h	6 h	12 h	12 h	542 h

Table 1: Volume horaire d'enseignement par type, niveau, et année universitaire

MISE EN PRODUCTION DE PROGRAMMES

2021 - 2022

Polytech Nancy

- **Niveau :** Polytech Nancy 5A STI (Master 2)
- **Responsable :** Moi-même
- **Volume horaire :** 12h CM, 12h TD

L'objectif de ce module est d'introduire aux étudiants les concepts et pratiques de la mouvance DevOps. Ce module porte donc sur les gestionnaires de versions de code et leurs pratiques d'utilisation (Git workflows), la mise en place de tests automatiques (unitaires, intégration), la mise en place de pipelines d'intégration continue (GitHub Actions) et la conteneurisation d'applications (Docker).

Responsable du module, j'ai établi le syllabus, réalisé les supports de cours et rédigé les sujets d'exercices.

ALGORITHMIQUE DISTRIBUÉE

2021 - 2022

Polytech Nancy

- **Niveau :** Polytech Nancy 4A STI (Master 1) et 3A IA2R (Licence 3)
- **Responsable :** Dr. Vincent Despré
- **Volume horaire :** 6h CM

L'objectif de ce module est d'initier les étudiants aux systèmes distribués et leurs problématiques, ainsi qu'à l'algorithmie distribuée. Ainsi, ce module porte sur les algorithmes fondamentaux utilisés dans les systèmes distribués (élection de leader, Bellman-Ford...) et leurs complexités en temps et en messages.

Chargé de cours pour la partie théorique, j'ai réalisé les supports de cours et retravaillé les sujets d'exercices.

PROGRAMMATION ALGORITHMIQUE II

2021 - 2022

Polytech Nancy

- **Niveau :** Polytech Nancy 2A PEIP (Licence 2)
- **Responsable :** Moi-même
- **Volume horaire :** 12h CM, 66h TD

L'objectif de ce module est de poursuivre l'apprentissage entamé en 1ère année de la programmation informatique. Ce module porte sur les notions de complexité en temps et en espace des algorithmes et utilise comme cas d'études des algorithmes récursifs (tris) et structures de données (listes, arbres...)

Responsable du module, j'ai réalisé les supports de cours et retravaillé les sujets d'exercices. J'ai aussi coordonné le bon déroulement du module avec le chargé de TD.

BASE DE DONNÉES II

2020 - 2022

Polytech Nancy

- **Niveau :** Polytech Nancy 3A spécialité IA2R (Licence 3)
- **Responsable :** Pr. Claude Godart
- **Volume horaire :** 6h CM et 12h TD

Suite directe du module de Bases de Données I, ce module enseigne aux étudiants les bases du langage SQL, comment déployer et utiliser une base de données à l'aide d'un SGBD Relationnel et finalement comment interagir avec une base de données depuis un langage de programmation.

Chargé de CM et de TD, j'ai actualisé les différents supports de cours et sujets de TD fournis.

BASE DE DONNÉES I

2020 - 2022

Polytech Nancy

- **Niveau :** Polytech Nancy 3A spécialité IA2R (Licence 3)
- **Responsable :** Pr. Claude Godart
- **Volume horaire :** 6h CM et 12h TD

L'objectif de ce module est de présenter aux étudiants le concept de Base de Données et de leur enseigner comment en concevoir et en utiliser une. Ce module se focalise sur la réalisation du schéma conceptuel d'un système d'information, la méthode pour concevoir un schéma relationnel équivalent, la normalisation de ce dernier et finalement son utilisation à l'aide de l'algèbre relationnel.

Chargé de CM et de TD, j'ai actualisé les différents supports de cours et sujets de TD fournis.

BASES DE DONNÉES I

2021 - 2022

Polytech Nancy

- **Niveau :** Polytech Nancy 3A EMME (Licence 3)
- **Responsable :** Moi-même
- **Volume horaire :** 6h CM, 12h TD

Destiné à un public de non-informaticiens, l'objectif de ce module est de présenter aux étudiants le concept de Base de Données. Ce module porte brièvement sur la conception de bases de données (schéma conceptuel, schéma relationnel) puis se focalise sur leur utilisation par le biais du langage SQL.

Responsable du module, j'ai retravaillé son contenu afin de porter une plus grande importance à l'aspect pratique (création et utilisation de bases de données relationnelles). J'ai aussi coordonné le bon déroulement du module avec la chargée de TD.

PROJET INFORMATIQUE

2020 - 2021

Polytech Nancy

- **Niveau :** Polytech Nancy 3A spécialité IA2R (Licence 3)
- **Responsable :** Dr. Vincent Despré
- **Volume horaire :** 10h TD

L'objectif de ce module est de renforcer et valider les compétences acquises par les étudiants via la conception et réalisation d'une application. Ce projet étant de plus grande envergure que ceux réalisés jusqu'alors, ce module insiste sur les bonnes pratiques de la conception orientée objet (séparation des responsabilités, factorisation du code, minimisation des dépendances...) et de développement (logiciels de gestion de versions, gestionnaires de dépendances...).

Chargé de TD, j'ai encadré des groupes d'étudiants dans la conception et réalisation de l'application demandée. J'ai aussi préparé et présenté en introduction de ce module un cours portant sur la conception orientée objet.

DÉCOUVERTE DE L'INFORMATIQUE

2020 - 2021

Polytech Nancy

- **Niveau :** Polytech Nancy 1A (Licence 1)
- **Responsable :** Dr. Charles Dumenil
- **Volume horaire :** 44h TD

Le but de ce module est d'initier les étudiants à la programmation informatique. Ce module présente les bases de l'algorithmie (instructions conditionnelles, boucles, fonctions...) et met en application ces notions dans le cadre d'exercices à réaliser en Python.

Chargé de TD.

PROGRAMMATION WEB SUR CLIENT

2017 - 2020

IUT Nancy Charlemagne

- **Niveau :** Licence Pro Informatique CIASIE
- **Responsable :** Dr. Gêrôme Canals
- **Volume horaire :** 24h TP (2017-2018) puis 24h EI (2018-2020)

Destiné à des étudiants ayant déjà appris et utilisé JavaScript au cours de leur formation précédente, ce module a pour but de consolider leur connaissance et maîtrise des bases du langage (POO, manipulation du DOM...) puis d'introduire des notions plus avancées (closures, AJAX, bundling...).

Chargé de CM et TP (Enseignement Intégré), j'ai notamment retravaillé le contenu du module (cours, exercices, projet) par rapport à l'évolution du langage à partir de la seconde année.

ALGORITHMIQUE

2018 - 2020

IUT Nancy Charlemagne

- **Niveau :** DUT Informatique 1A
- **Responsable :** Dr. Yolande Belaïd
- **Volume horaire :** 32h TD

Destiné à des étudiants débutant leurs études en informatique, l'objectif de ce module est de leur présenter la notion d'algorithme et de leur enseigner comment en concevoir (instructions disponibles, décomposition de problèmes en sous-problèmes...).

Chargé de CM et TD (Enseignement Intégré).

CONCEPTION ORIENTÉE OBJET

2017 - 2018

IUT Nancy Charlemagne

- **Niveau :** DUT Informatique 1A
- **Responsable :** Dr. Vincent Thomas
- **Volume horaire :** 42h TD

L'objectif de ce module est d'enseigner aux étudiants les principes de la conception orientée objet (séparation des responsabilités, factorisation du code...) et de leur apprendre à manier les outils existants (UML, patrons de conceptions...). Une partie du module est aussi consacrée aux bonnes pratiques de développement (logiciels de gestion de versions, tests unitaires...).

Chargé de TD et TP.

BASES DE LA PROGRAMMATION OBJET

2016 - 2017

Faculté des Sciences et Technologies de Nancy

- **Niveau :** Licence 2 Informatique
- **Responsable :** Dr. Martine Gautier
- **Volume horaire :** 18h TP

L'objectif de ce module est d'enseigner aux étudiants le paradigme de la programmation orientée objet et ses spécificités (classe, héritage, polymorphisme...), ainsi que les bonnes pratiques de développement (programmation par contrat, tests...). Ces concepts sont ensuite mis en application dans le cadre de multiples exercices à réaliser en Java.

Chargé de TP, j'ai notamment participé à la conception et l'animation du TP noté.

TELECOM Nancy

- **Niveau :** TELECOM Nancy 1A (Licence 3)
- **Responsables :** Dr. Gérald Oster et Pr. Martin Quinson
- **Volume horaire :** 30h TP

Destiné aux élèves provenant de classes préparatoires, ce module a pour but de travailler les notions de bases de la programmation (instructions, conditions, boucles...) avant d'aborder des exercices plus complexes (tris, récursivité). Dans le cadre de ce module, les étudiants travaillent de façon autonome sur l'environnement d'apprentissage de la PLM.

Chargé de TP.

Encadrement

STAGE TELECOM NANCY 2A

- Ishara Chan-Tung : *Intégration d'un agent de messages basé sur des journaux au sein d'une application d'édition collaborative*, juin à août 2019, co-encadrement avec Cédric Enclos
- Pierric Grguric : *Service de compilation isolé pour la PLM*, juin à août 2015, co-encadrement avec Gérald Oster et Martin Quinson
- Alexandre Carpentier : *Remédiation de masse dans un environnement d'apprentissage*, juin à août 2015, co-encadrement avec Gérald Oster et Martin Quinson
- Tanguy Gloaguen : *Mise en place d'un environnement de qualification pour la plateforme PLM*, juin à août 2015, co-encadrement avec Gérald Oster et Martin Quinson

PROJET D'INITIATION À LA RECHERCHE TELECOM NANCY 2A

- Pierre Maeckereel, Yannick Philippe : *Simulation du comportement de collaborateurs dans une session d'édition collaborative*, janvier à mai 2017, co-encadrement avec Quentin Laporte-Chabasse

STAGE DUT

- Tom Mendez-Porcel : *Implémentation d'un protocole de gestion de groupe au sein d'une application d'édition collaborative*, co-encadrement avec Victorien Elvinger, avril à juillet 2020
- Théodore Lambolez : *Portage web d'un exerciceur de programmation*, avril à juin 2015, co-encadrement avec Gérald Oster et Martin Quinson
- Baptiste Mounier : *Langage visuel pour un exerciceur*, avril à juin 2015, co-encadrement avec Gérald Oster et Martin Quinson
- Benjamin Thirion : *Conception et réalisation d'un éditeur d'exercices de programmation*, avril à juin 2015, co-encadrement avec Gérald Oster et Martin Quinson