Thèse présentée pour l'obtention du grade de

DOCTEUR de SORBONNE UNIVERSITÉ

Spécialité

Ingénierie / Systèmes Informatiques

École doctorale

Informatique, Télécommunication et Électronique Paris (ED130)

# CALock : Topological Multi-Granularity Locking for Hierarchical data

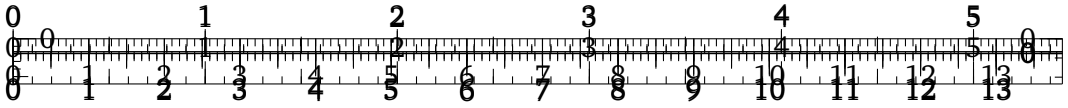
---

Ayush Pandey

Soutenue publiquement le : *21 Mars 2025*

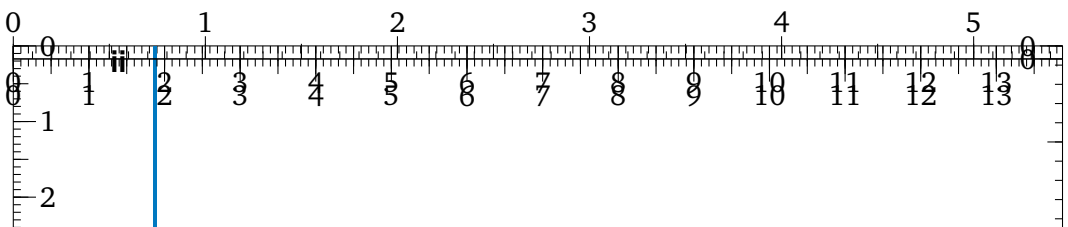
Devant un jury composé de :

| | |
|---|---|
| **Jean-Michel DUPONT**, Professeur, Sorbonne Université | *Rapporteur* |
| **Alice DUPONT**, Directrice de Recherche, CNRS | *Examinatrice* |
| **Jean-Michel DUPONT**, Professeur, Sorbonne Université | *Examinateur* |
| **Pierre AAA**, Maître de conférences, UPMC | *Examinateur* |
| **Mesaac MAKPANGOU**, Chargé de Recherche [HDR], INRIA | *Directeur de thèse* |
| **Marc SHAPIRO**, Directeur de Recherche Émérite, INRIA, Sorbonne Université, LIP6 | *Directeur de thèse* |
| **Julien SOPENA**, Maître de Conférences, Sorbonne Université, LIP6 | *Encadrant* |
| **Swan DUBOIS**, Maître de Conférences, Sorbonne Université, LIP6 | *Encadrant* |



*À Alice et Bob*





# Remerciements

A PhD is a long journey. For me, it took 3 years, 3 supervisors and education in 3 countries to finally reach here. Throughout this time, I have been supported by many people, without whom I would not have been able to reach here. I would like to take this opportunity to thank them.

First and foremost, I would like to thank the rapporteurs of this thesis, . I am grateful for their time and effort in reviewing my work and providing valuable feedback. I would also like to thank the members of the jury, for taking the time to read, evaluate and provide feedback on this body of work.

Add names of rapporteurs

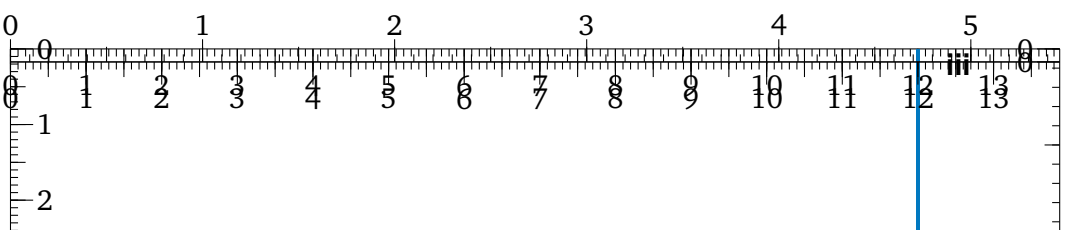
Add names of jury members

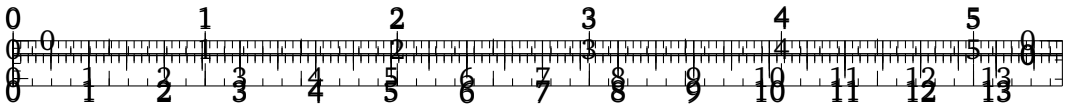
The people without whom this thesis would not have been possible are my supervisors, Marc Shapiro, Julien Sopena and Swan Dubois. From asking me if i would like to do a PhD back in 2021 while i was in Germany to the day I defended my thesis, I cannot thank Marc enough for his constant support, guidance and well placed reality checks. Julien, you have been a great inspiration, not only in the rigor of research but also in correcting my French which, even after several corrections, is still awful. Swan, you have been a great mentor and helped me navigate the finer details of theoretical computer science. I am grateful for the opportunity to have worked with you all.

To the seniors in the lab with whom i have shared numerous lunches, coffee breaks and discussions, thank you for making my time in the lab enjoyable. Thank you Pierre, Jonathan and Alessio for taking care of DELYS and your help in my research.

I have had the pleasure of working with some amazing people during my time in the lab. Saalik, Benoît and Laurent have been great friends and colleagues, helping me navigate the French bureaucracy, teaching me bits of french culture and also helping me with my research direction. I am grateful for your friendship and support. To my fellow PhDs Étienne, Aymeric, Célia, Baptiste and Daniel; Thank you for your discussions and support. I wish you well for your future.

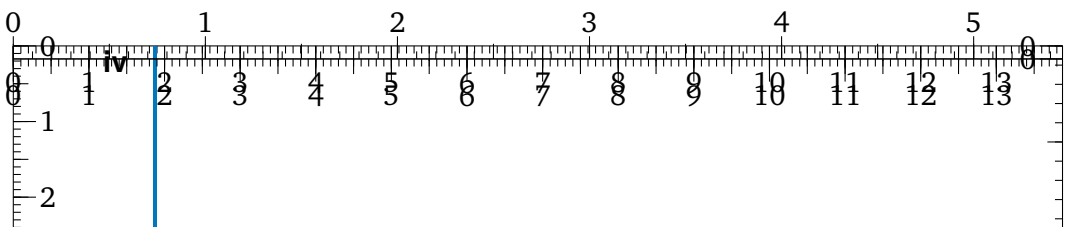
To the lehrstuhl softwaretechnik of RPTU, i extend my thanks, for introducing me to the wonderful world of parallel and distributed systems. I am grateful for the opportunity to have worked with Prof. Dr. Annette Bieniusa without whom I would not have been able to start my PhD.

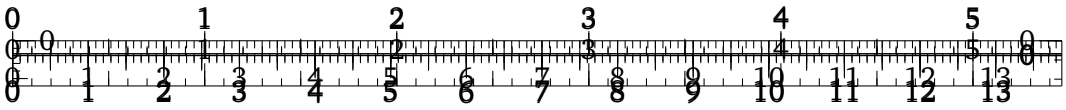




Finally, the people without whom this thesis would not have been imaginable. My family, who dealt with years of daily phone and video calls, the fear of the COVID-19 pandemic, the uncertainty of being abroad and the stresses of a PhD, i am eternally grateful for your support. I dedicate this thesis to you Papa, Mummy and my brother. To my friends from Germany, India and France, thank you for your support, your love and your friendship.

Thankyou everyone.





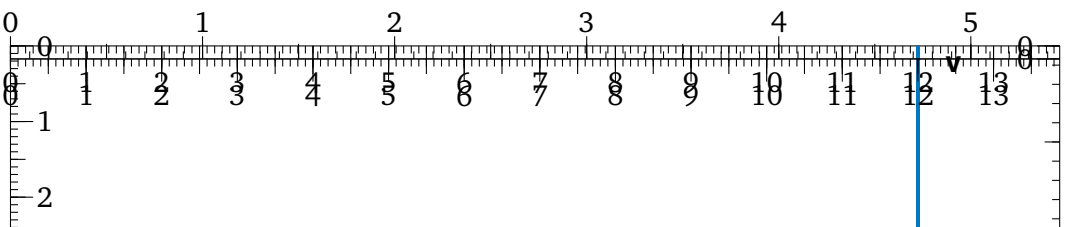
# Abstract

Hierarchies serve as fundamental structures across various disciplines, modeling hierarchical relationships in computer science, biology, social networks, and logistics. However, the dynamic and concurrent updates in real-world systems necessitate synchronization techniques for maintaining data consistency despite concurrent access. This paper explores a novel approach called CALock to synchronize operations on hierarchies by utilizing a labeling scheme that facilitates multi-granularity locking.

Our approach addresses both concurrent data reads and writes as well as structural modifications. CALock exploits the hierarchical topology via a new labeling scheme to identify the common ancestors of vertices. This enables a thread to identify an appropriate lock granule for its lock request. Leveraging variable lock granularities optimizes operations across the hierarchy while ensuring consistency and performance.

We provide a detailed discussion of the CALock labeling and the locking algorithm, prove its properties, and evaluate it experimentally. On static hierarchies CALock remains competitive with previous labeling schemes and has better concurrency and throughput when structural modifications change the hierarchy. In particular, CALock improves throughput by 4.5×, and response time by 1.5× for workloads that contain structural modifications.

**Keywords:** Multi-granularity locking, Hierarchical data, Graphs, Locking, Synchronization, Graph topology, Ancestors.



# Résumé

Les hiérarchies servent de structures fondamentales dans diverses disciplines, modélisant les relations hiérarchiques en informatique, en biologie, dans les réseaux sociaux et en logistique. Cependant, les mises à jour dynamiques et simultanées dans les systèmes du monde réel nécessitent des techniques de synchronisation pour maintenir la cohérence des données malgré l'accès simultané. Cet article explore une nouvelle approche appelée CALock pour synchroniser les opérations sur les hiérarchies en utilisant un schéma d'étiquetage qui facilite le verrouillage multi-granularité.

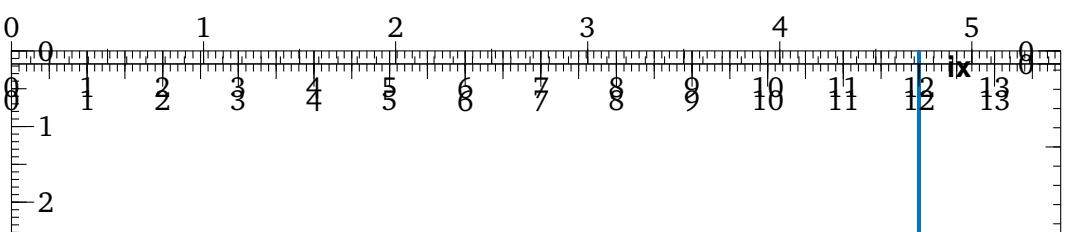
Notre approche concerne à la fois les lectures et écritures concurrentes de données et les modifications structurelles. CALock exploite la topologie hiérarchique par le biais d'un nouveau schéma d'étiquetage permettant d'identifier les ancêtres communs des sommets. Cela permet à un thread d'identifier un granule de verrouillage approprié pour sa demande de verrouillage. L'utilisation de granularités de verrouillage variables optimise les opérations à travers la hiérarchie tout en garantissant la cohérence et les performances.

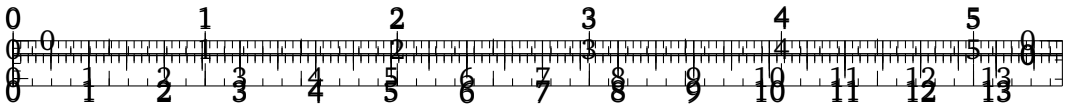
Nous présentons une discussion détaillée de l'étiquetage CALock et de l'algorithme de verrouillage, nous prouvons ses propriétés et nous l'évaluons de manière expérimentale. Sur des hiérarchies statiques, CALock reste compétitif par rapport aux schémas d'étiquetage précédents et offre une meilleure simultanéité et un meilleur débit lorsque des modifications structurelles changent la hiérarchie. En particulier, CALock améliore le débit de 4,5× et le temps de réponse de 1,5× pour les charges de travail qui contiennent des modifications structurelles.

**Mots-clés:** Verrouillage multi-granularité, Données hiérarchiques, Graphes, Verrouillage, Synchronisation, Topologie des graphes, Ancêtres.

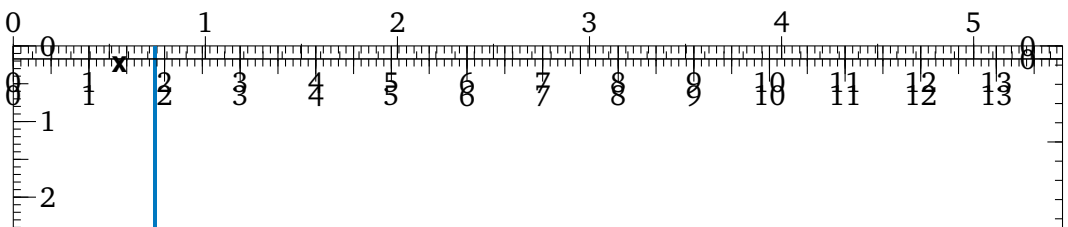
# Contents











# 1 Introduction

## 1.1 Background and Motivation

Overview of concurrency control and the importance of locking in database systems. Challenges associated with multi-granularity locking. Introduction to CALock and its significance.

## 1.2 Research Objectives

Define the main goals of the research. Specific objectives related to CALock and optimal locking grain.

## 1.3 Problem Statement

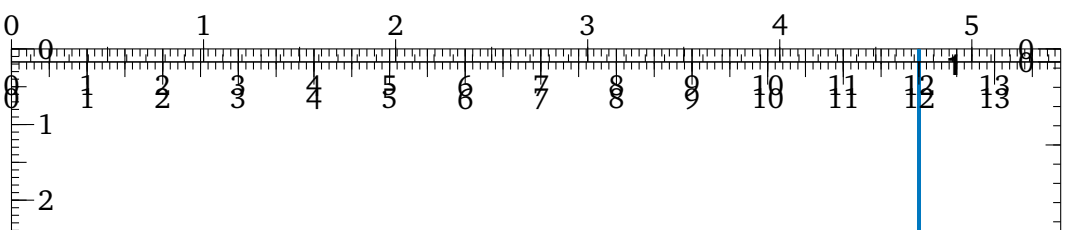
Detailed description of the problem addressed by CALock.

## 1.4 Contributions

Summary of key contributions, including the development and evaluation of CALock.

## 1.5 Thesis Structure

Brief outline of each chapter.



# 2 Background

## 2.1 Concurrency Control Mechanisms

Overview of concurrency control techniques in database systems.

## 2.2 Multi-Granularity Locking

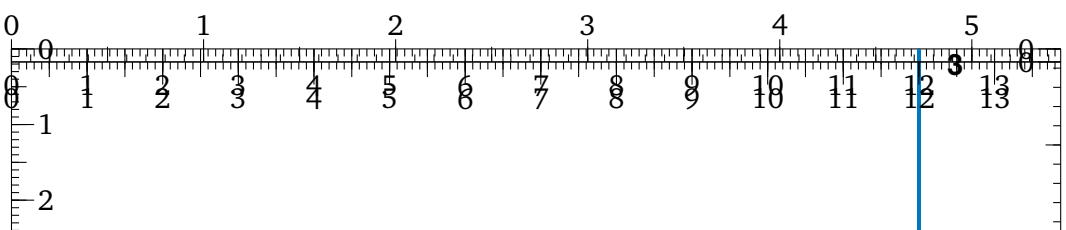
Detailed discussion of multi-granularity locking concepts. Existing algorithms and their limitations.

## 2.3 Related Work

Survey of existing research on locking mechanisms and optimization techniques. Comparison of related approaches to the CALock algorithm.

## 2.4 Summary

Gaps identified in the current literature that CALock addresses.



# 3 Theoretical Framework

## 3.1 Multi-Granularity Locking Model

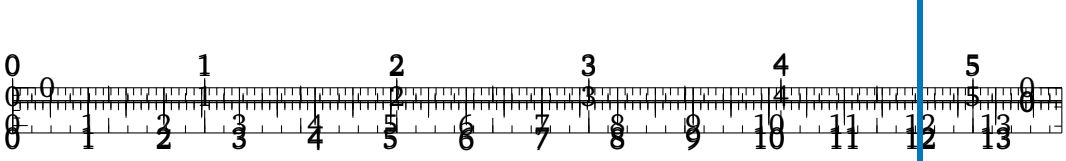
Formal definitions and theoretical underpinnings.

## 3.2 Lowest Common Ancestors (LCA) in Graphs

Overview of LCA concepts and algorithms. Application of LCA in the context of locking granularity.

## 3.3 Problem Formulation

Mathematical and conceptual formulation of the optimal locking grain problem.



# 4

# CALock: A New Approach to Multi-Granularity Locking

## 4.1 Overview of CALock

Introduction to the CALock algorithm and its design principles.

## 4.2 Algorithm Design

Step-by-step description of the CALock algorithm. Use of LCA to find optimal locking grain.

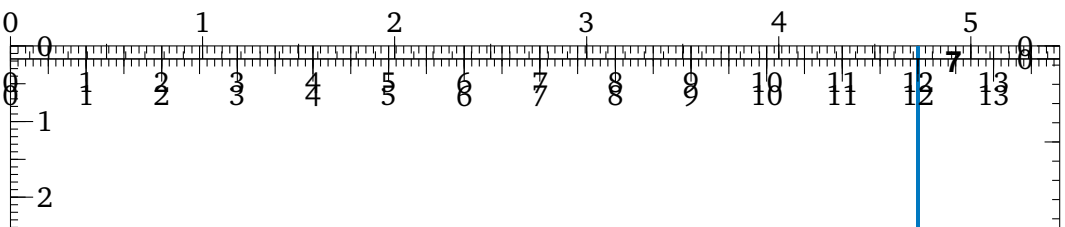
## 4.3 Theoretical Analysis

Proof of correctness and optimality. Complexity analysis of the CALock algorithm.

## 4.4 Implementation Details

Practical aspects of implementing CALock. Tools, libraries, and datasets used.

Don’t forget to put the source code link ;-)



# 5 Experimental Evaluation

## 5.1 Experimental Setup

Description of the experimental environment. Datasets, benchmarks, and metrics used for evaluation.

## 5.2 Performance Analysis

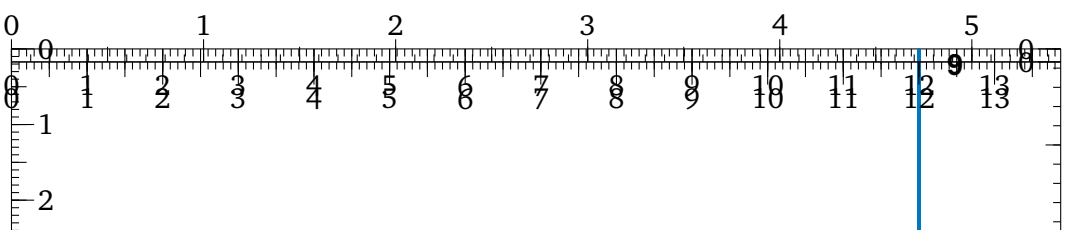
Results of CALock compared with existing algorithms. Analysis of performance improvements in terms of efficiency and resource utilization.

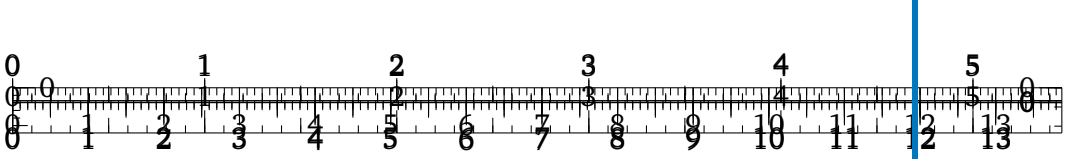
## 5.3 Sensitivity and Robustness Analysis

Evaluation of CALock under various conditions and parameters.

## 5.4 Discussion of Results

Interpretation of experimental results. Strengths and potential limitations of CALock.





# 6 Conclusion

## 6.1 Summary of Findings

Recap of the key findings and achievements of the research.

## 6.2 Contributions

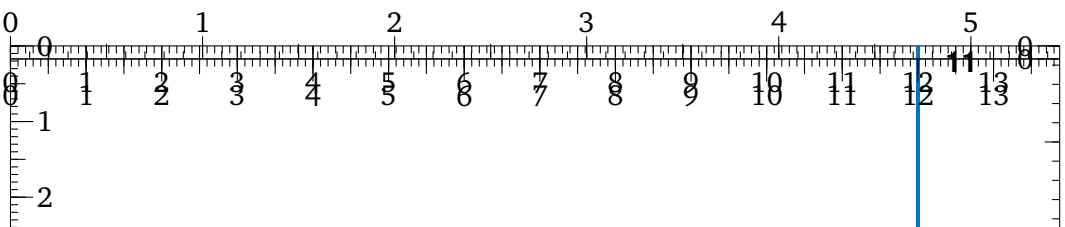
Reiteration of the thesis contributions in the context of the broader field.

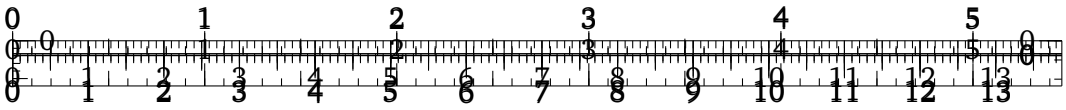
## 6.3 Future Work

Suggestions for further research and potential enhancements of CALock.

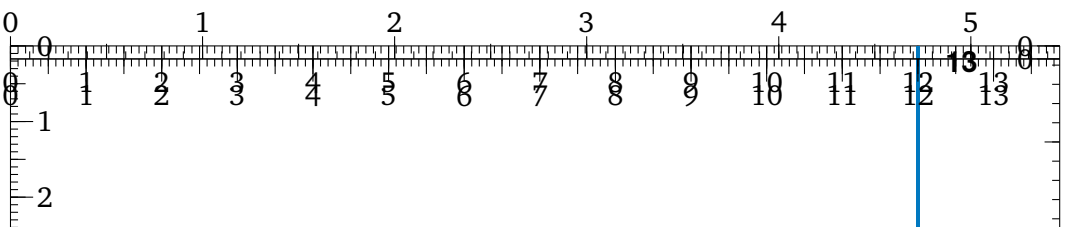
## 6.4 Final Remarks

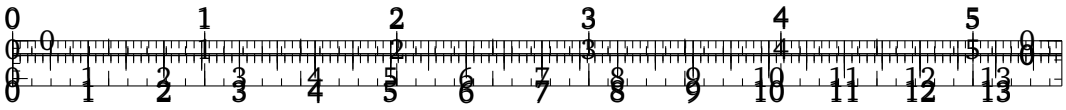
Reflection on the research journey and its impact.



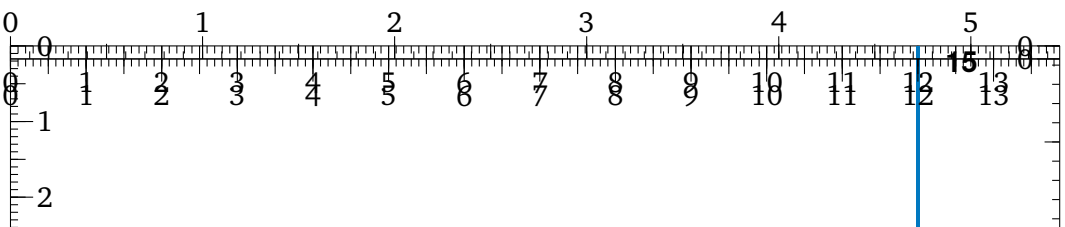


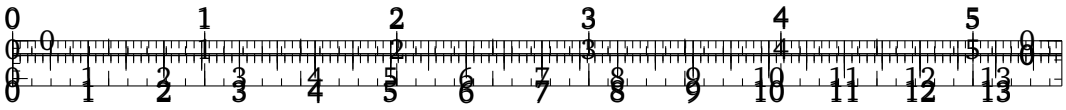
# List of Figures



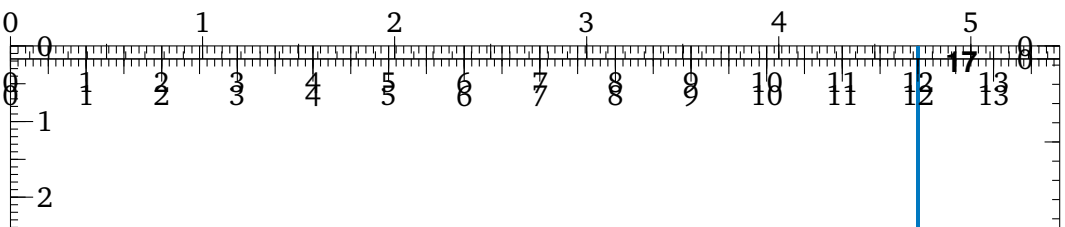


# List of Tables





# List of Listings



# Résumé


The doctoral schools of Sorbonne Université require at least one page of summary in French. (Even if the website says that you need to provide more than one page, one is enough, which I did).

Résumons donc en français. Tout bon résumé commence par une description générale du problème de la thèse. Le problème étant que les écoles doctorales de Sorbonne Université exigent au moins une page de résumé en français. Voici donc un résumé garanti non traduit sur un DeepL.

Le deuxième paragraphe du résumé doit répondre à la question *pourquoi le(s) problème(s) présenté(s) est un vrai problème ?*.

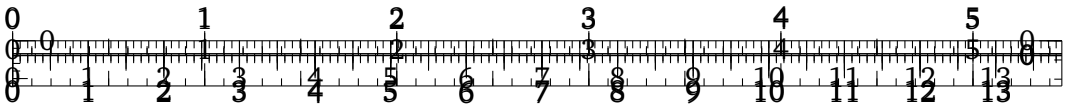
Ensuite, le troisième paragraphe doit répondre à la question *quelles sont les solutions apportées par la thèse à ces problèmes ?* Cette thèse explore ces problèmes en profondeur, en étudiant l'état de l'art lié ... et présente la solution *SystèmeCool2Ouf*, conçu pour répondre aux problématiques exposées. L'une des principales exigences est une approche *DurACuir* qui *EstVraimentVraimentDureCar...* Cependant, cela rend difficile la satisfaction des attentes en matière de *Toute solution à des petits défauts et compromis, on va pas se le cacher*.

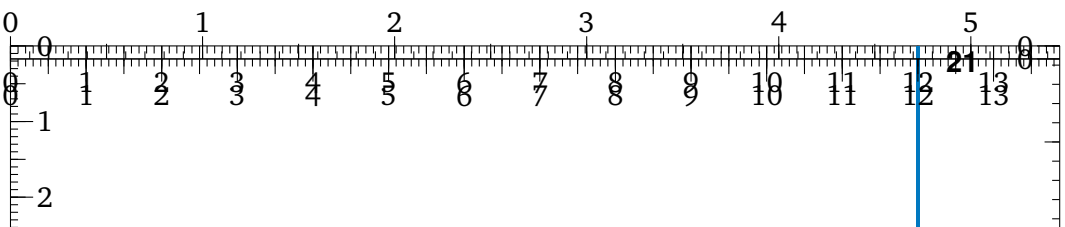
Paragraphe 4, *En quoi la solution est meilleure que l'état de l'art ?* Pour répondre à ces défis, nous avons fait le choix d'adopter une approche ... en fournissant les plus fortes garanties de .... Un défi connexe est ..., que nous avons limité grâce à ...

Enfin, traditionnelement on liste une référence aux contribution, Les contributions de cette thèse peuvent être résumées comme suit:

- ...;
- ...;
- ...;

Notre évaluation expérimentale montre que ...




21