SCIENCES & BIOENGINEERING SCIENCES

The Research Group
Software Languages Lab

has the honor to invite you to the public defence of the PhD thesis of

Jim Bauwens

to obtain the degree of Doctor of Sciences

Title of the PhD thesis:

FLEXIBLE CRDTS FOR A DEMANDING WORLD
An open and Systematic Implementation Approach For CRDT Development

Promotor:
Prof. dr. Elisa Gonzalez Boix

The defence will take place on

Friday, August 23, 2024 at 5 p.m. in auditorium I.2.01

Members of the jury

Prof. dr. Bas Ketsman (VUB, chair)
Prof. dr. Paul Van Eecke (VUB, secretary)
Prof. dr. Bruno da Silva Gomes (VUB)
Prof. dr. Carlos Baquero (University of Porto, Portugal)
Prof. dr. Annette Bieniusa (University of Kaiserslautern-Landau, Germany)

Curriculum vitae

Jim Bauwens obtained his BSc at the Artesis Plantijn Hogeschool in Antwerpen in 2015, and his master at the Vrije Universiteit Brussel (VUB) in 2019. He then started a PhD at the Software Languages Lab (SOFT) supported by an FWO-SB fellowship.

His research has mainly focused on software engineering techniques for replicated data types to help developers implement available systems. Jim's work resulted in three publications in international peer-reviewed journals and conferences, and six contributions in international peer-reviewed workshops.

Abstract of the PhD research

The increase in internet-connected devices, including smartphones and IoT systems, has changed the landscape of distributed systems. To ensure availability, data is often replicated between servers or systems at the edge using a local-first approach. This improves performance but introduces challenges as different copies of data need to be kept up-to-date. Conflict-free Replicated Data Types (CRDTs) can be used to ensure the convergence of replicated data, without the need for manual conflict resolution strategies. Replicas can diverge temporarily, but once updates stop they will eventually converge.

During our research, we identified four main challenges related to the use of CRDTs. First, memory management is complicated due to the accumulation of metadata tracking causality between operations. Secondly, there is limited support for dynamic networks, which is essential for local networks when peers can join and leave. Third, the reliance on Reliable Causal Broadcasting in many CRDT designs can lead to unnecessary delays. Operations will be buffered until all causal dependencies are met, even if those dependencies are not logically related. Fourth, existing support for the composition and nesting of CRDTs is inflexible and limited, forcing ad-hoc implementations.

This dissertation investigates and develops solutions to these concerns using Flec — a framework we developed for building eventually consistent applications — as a laboratory for experimenting with novel CRDT designs and techniques. Flec offers a modular approach for developing CRDTs through an open implementation that reifies the replication and convergence process.

To address the unbounded accumulation of metadata, we first introduce a technique that leverages communication acknowledgements to eagerly determine causal stability, enabling early metadata removal. We then adapt this mechanism to support networks where new peers can dynamically join, and show how peers can obtain correct states asynchronously. Secondly, we propose an extension to pure operation-based CRDTs that improves their responsiveness. By allowing pending operations stored in the RCB buffer to be partially applied before all causal dependencies have arrived, the reactivity of replicas is enhanced. Finally, we propose Nested Pure Operation-Based CRDTs, a novel framework for building nested replicated data structures. Our approach allows CRDTs to support composition without semantic changes or structural limitations.

We validate our contributions by implementing a range of novel structured RDTs within Flec. Each chapter includes a validation section that substantiates our proposed techniques and extensions. The results demonstrate the effectiveness of our approach, which we believe forms a step forward in enabling the usage of CRDTs in a wide variety of applications.