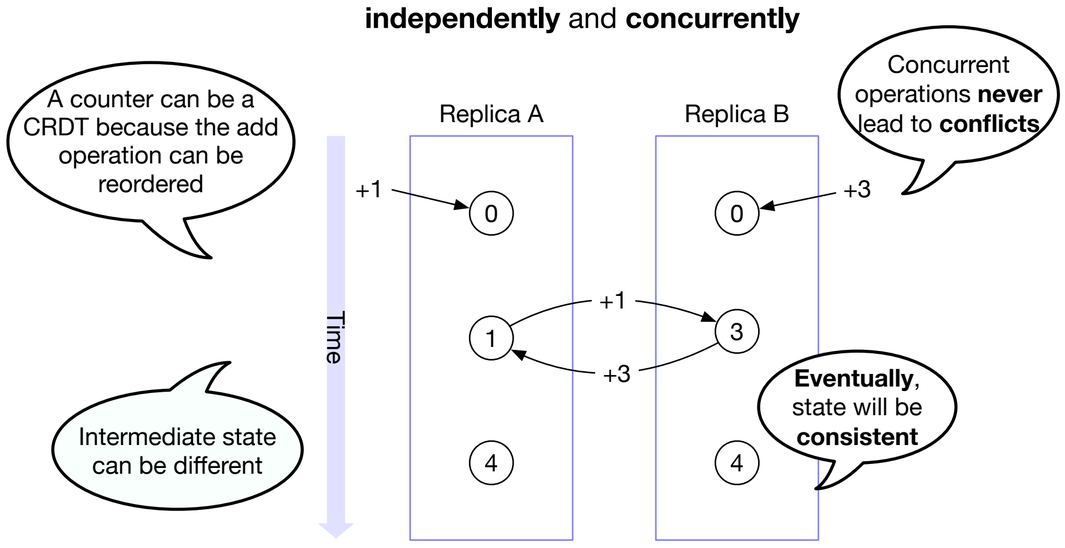
Memory Efficient CRDTs in Dynamic Environments

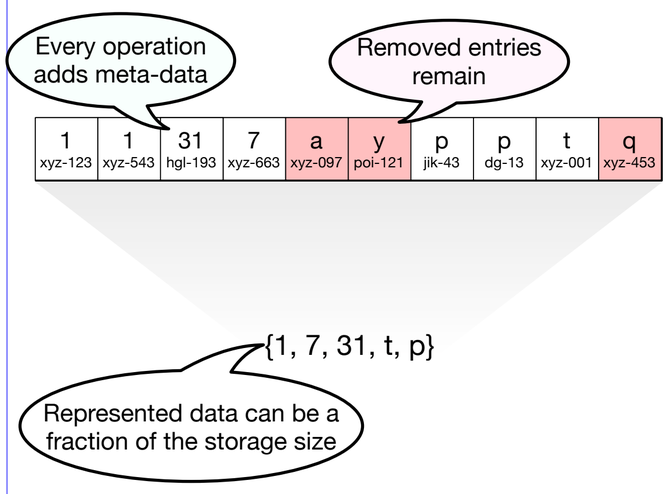
Jim Bauwens, Elisa Gonzalez Boix
Vrije Universiteit Brussel

Context: **C**onflict-free **R**eplicated **D**ata **T**ypes (CRDTs)

A family of **replicated data structures** where replicas can be updated **independently** and **concurrently**



CRDTs are not memory efficient



With every operation unique meta-data is used to ensure that concurrent operations always remain commutative

Causal Stability, a solution?

An operation is causally stable when it has been applied to all replicas (Baquero et al. 2017), i.e. no concurrent updates are possible.

When this is the case, meta-data is useless and can be removed.

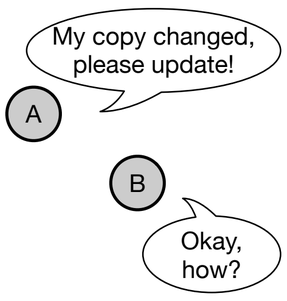
Current approaches assume a fixed number of replicas

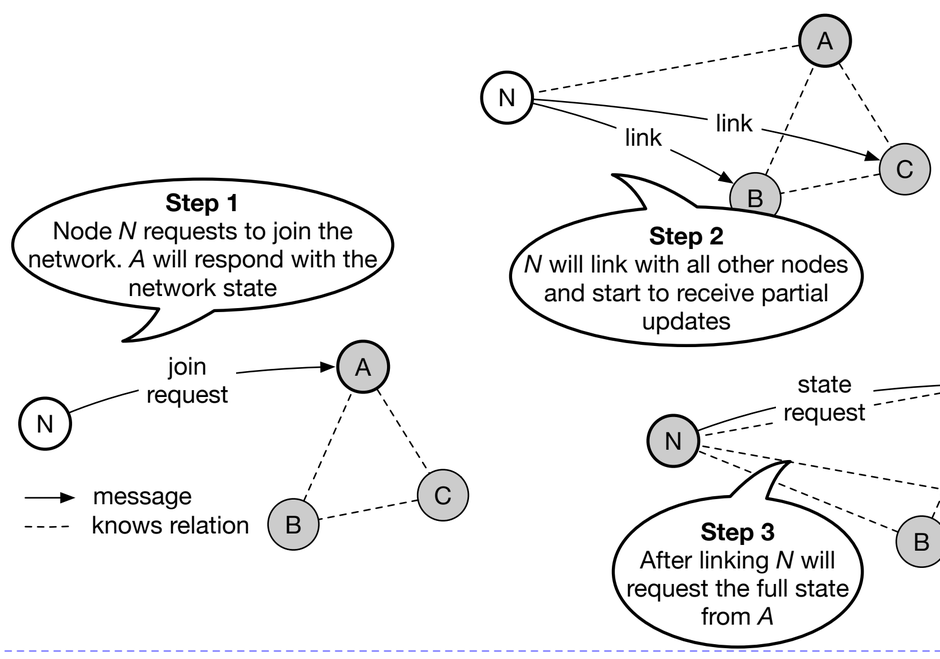
Our approach for efficient CRDTs in a dynamic environment

① **Join model**: dynamically adding replicas to a system

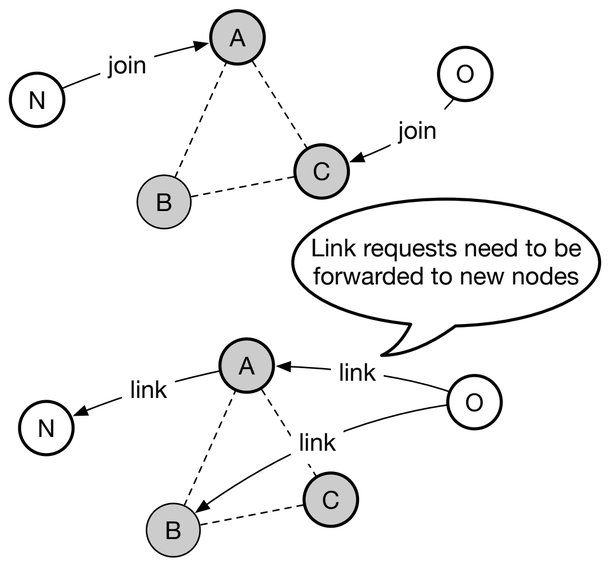
Why a join model?

Management of replicas and propagation of updates is currently left as a job for application developers instead of CRDT designers





Handling concurrent joins



② **Garbage collection**: determining causal stability in a dynamic environment

Our approach

In order to determine causal stability in a dynamic network we need to handle **joins concurrent with operations**. We highlight the three relevant cases needed to enable garbage collection in such environments.

The concurrent operation sources from A

- operation is already stable: no problem
- operation is not stable: add node *N* to the list of nodes that *A* needs a response from for determining stability

The concurrent operation sources from N

- buffer all operations from *N* until it has received a full state

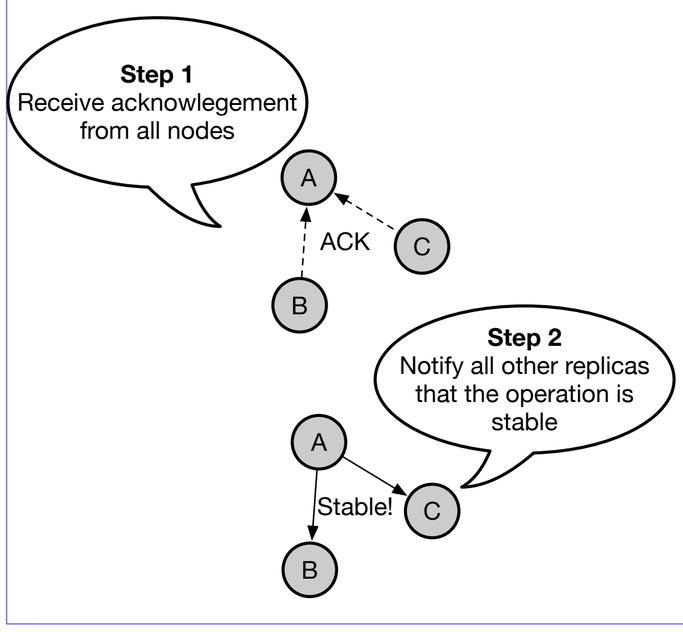
The concurrent operation sources from C

- *C* has already been linked to *N*: no problem
- *C* has not yet been linked to *N*: add node *N* to the list of nodes that *C* needs a response from for determining stability

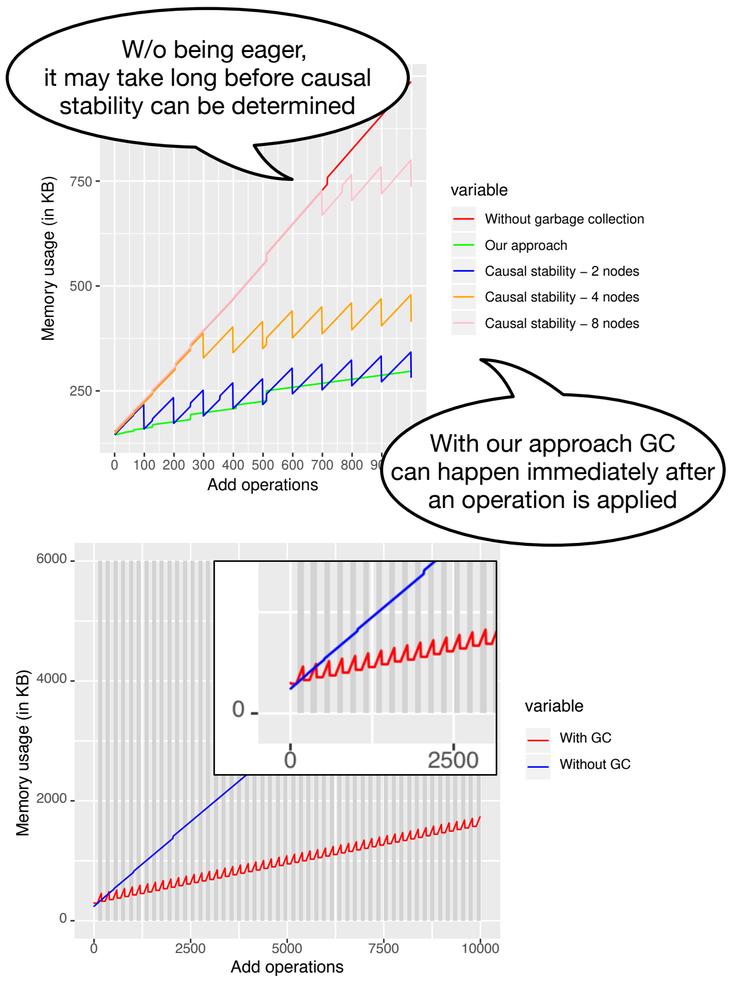
Be more eager!

Garbage collection based on causal stability can only occur in a replica when it has received a update from all nodes, i.e. if a replica issues no operations, causal stability can never be determined.

In our approach we piggyback on the acknowledgments received by the reliable causal broadcasting mechanism to enable more eager garbage collect



Experiments



Memory Efficient CRDTs in Dynamic Environments. J. Bauwens and E. Gonzalez Boix. Proceedings of the 11th ACM SIGPLAN International Workshop on Virtual Machines and Intermediate Languages - VMIL '19 (2019)

jim.bauwens,egonzale]@vub.be
https://soft.vub.ac.be/~jibauwen