« Evaluating CRDTs for Real-time Document Editing »

Summary

Ronan-Alexandre Cherrueau
Adrien Bougouin
Alban Ménager

20 mars 2012

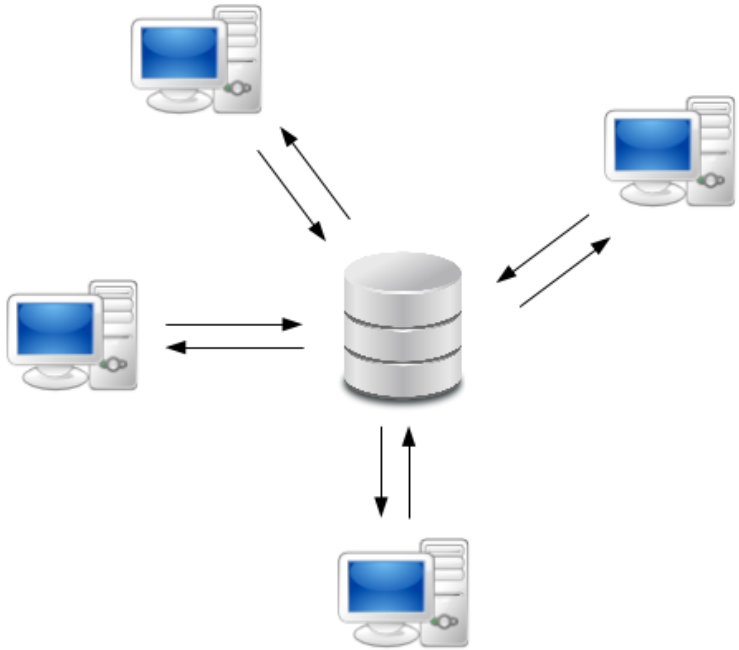
Plan

Introduction

- Increasing of collaboring work and *real-time* editing systems
- A good example : Google Docs
  - Allows editing on the same document at the same time by multiple authors.

Replication mechanism

- Real-time editing systems use replication mechanism to ensure consistency.
- Optimistic replication gives to users a low time of latency.



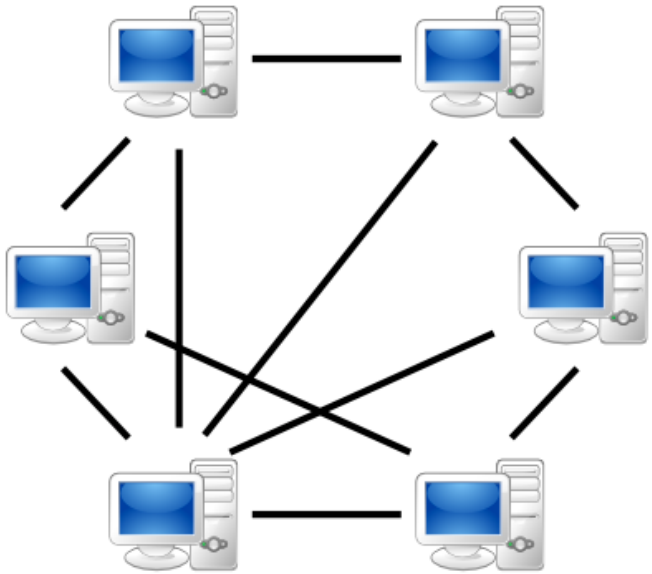
Problems

Centralized approach may cause problems :

- Personal datas are stored during the edition.
- It may be a privacy threat if they are used by corporations.

Solution

- Use decentralized mechanisms : Peer-to-peer.
- The main factor for suitable solutions is to respond to the users' actions in a reasonable time (about 50ms).



Goal

- Select algorithms based on optimistic replication.
- Evaluate them on a decentralized real-time collaborative editing system.
- Evaluations based on real context on the same conditions and using the same data flow.

First approach : Operation transformation

- Locally executed.
- Sent to others sites.
- Received by the centralized site.
- Transformed according to concurrent operations
- Executed on local copy.

New approach : CRDT

Commutative Replicated Data Types (CRDT)

- New class of replication mechanisms to preserve consistency.
- For peer-to-peer environment.
- The concurrent operations are natively commutative.
- The document is a linear sequence of elements.
- A single position identifier.

Selected Algorithms

- Logoot
- RGA
- WOOT
- WOOTO
- WOOTH

Theoretical evaluation

Algorithm	Local		Remote	
	INS	DEL	INS	DEL
WOOT	$O(H^3)$	$O(H)$	$O(H^3)$	$O(H)$
WOOTO	$O(H^2)$	$O(H)$	$O(H^2)$	$O(H)$
WOOTH	$O(H^2)$	$O(H)$	$O(H^2)$	$O(log(H))$
Logoot	$O(H)$	$O(1)$	$O(H.log(H))$	$O(H.log(H))$
RGA	$O(H)$	$O(H)$	$O(H)$	$O(log(H))$
SOCT2/TTF	$O(H+R)$	$O(H+R)$	$O(H^2)$	$O(H^2)$

Figure: Worst-case time-complexity analysis

We see that RGA and Logoot have the bests results.

Peer-to-Peer collaboration

- The team designed a real-time peer-to-peer collaborations application.
- In order to obtain real logs.
- And apply logs to the algorithms.

Groups for the experiments

- 3 groups have to do their semester report by only using the collaborating editor for one hour and a half :
  - 2 groups of 4 students.
  - 1 group of 5 students.
- 9 groups of 2 students have to translate an episode of *The Big Bang Theory*

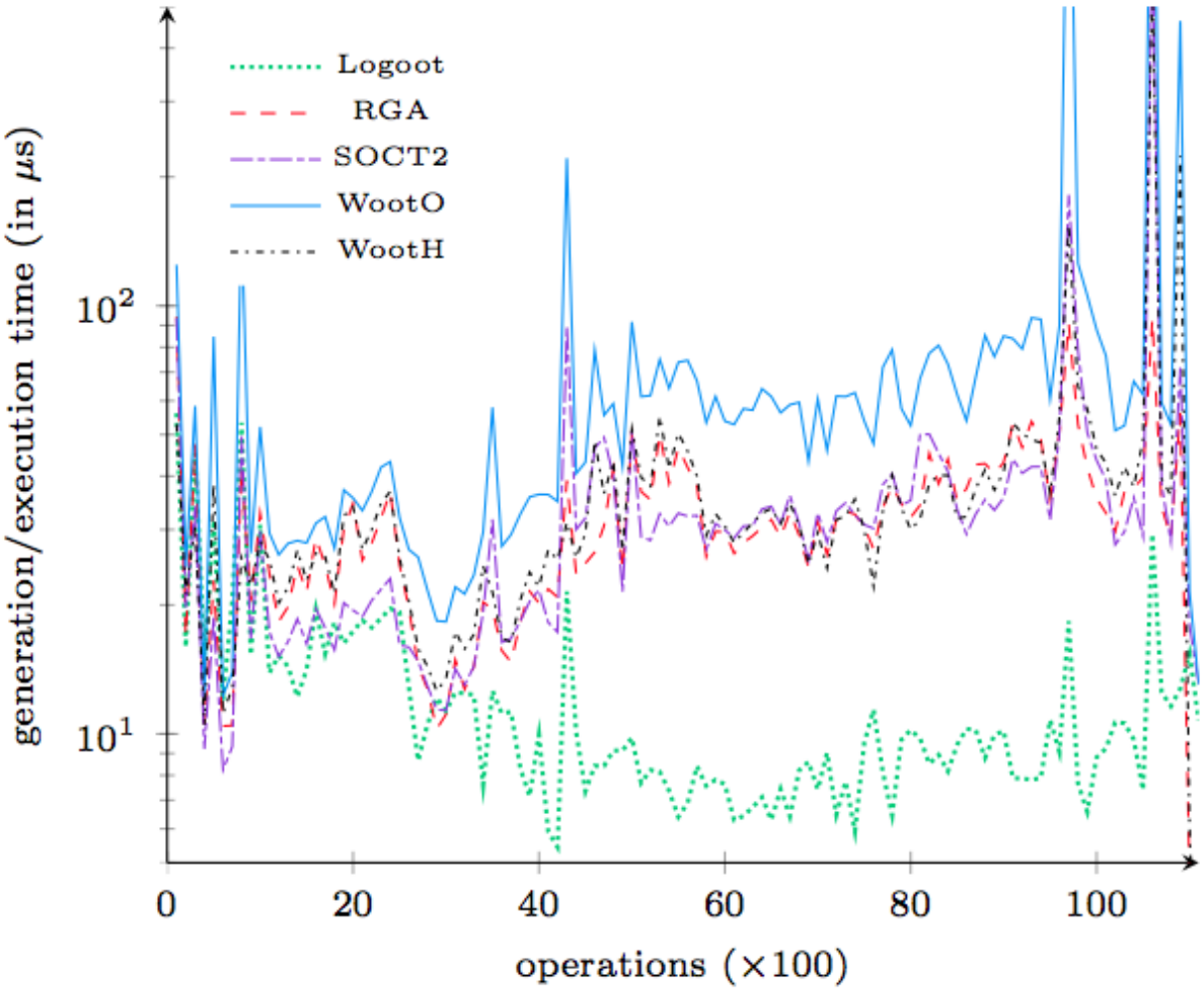
1H30 for each experiment.

Logs

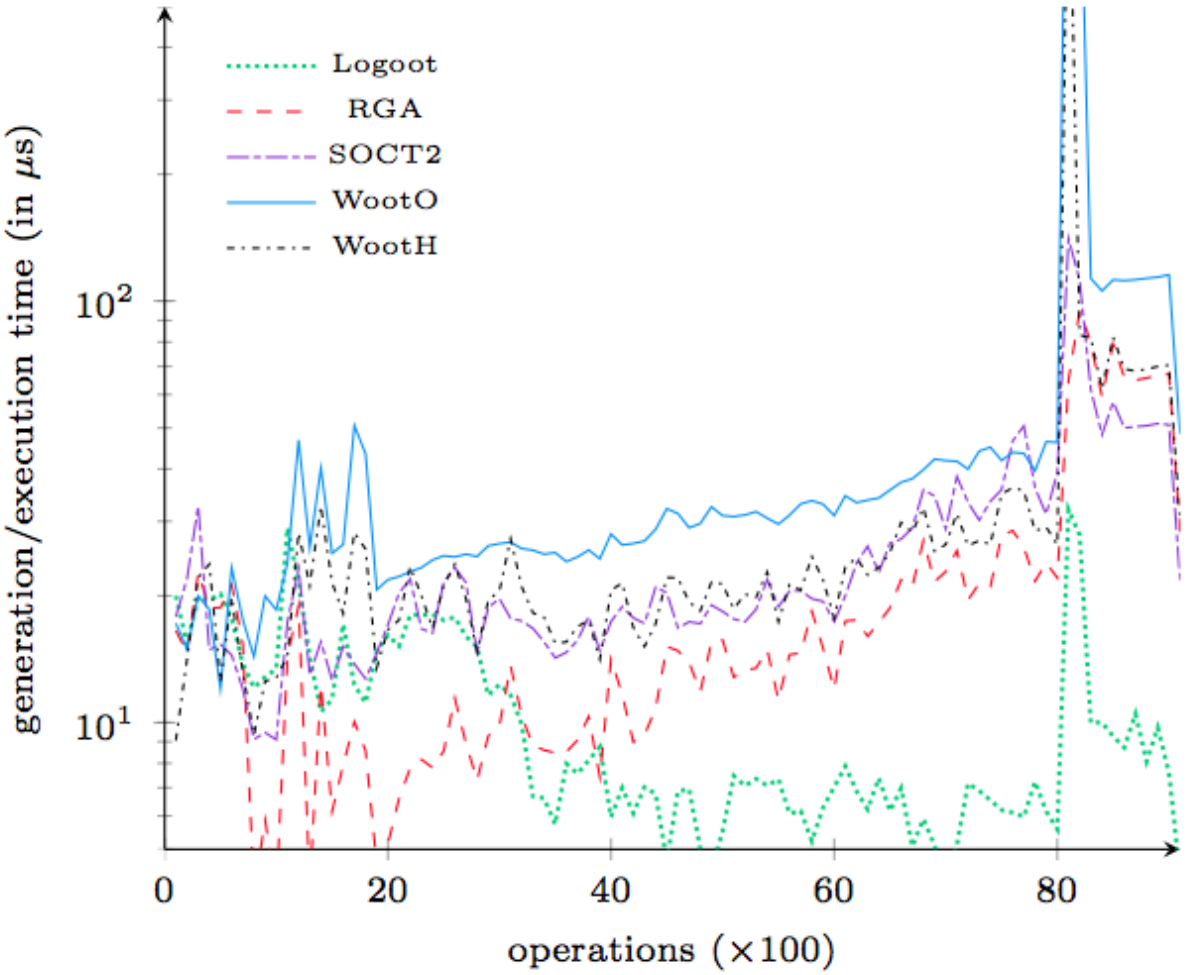
	Report			Series	
	Group 1	Group 2	Group 3	Doc 1	Doc 2
No. user operations	11 211	11 066	13 702	9 042	9 828
No. char. operations	26 956	47 992	42 443	29 882	10 268
% of del	12	12	12	9	5

Figure: Total number of user/character operations

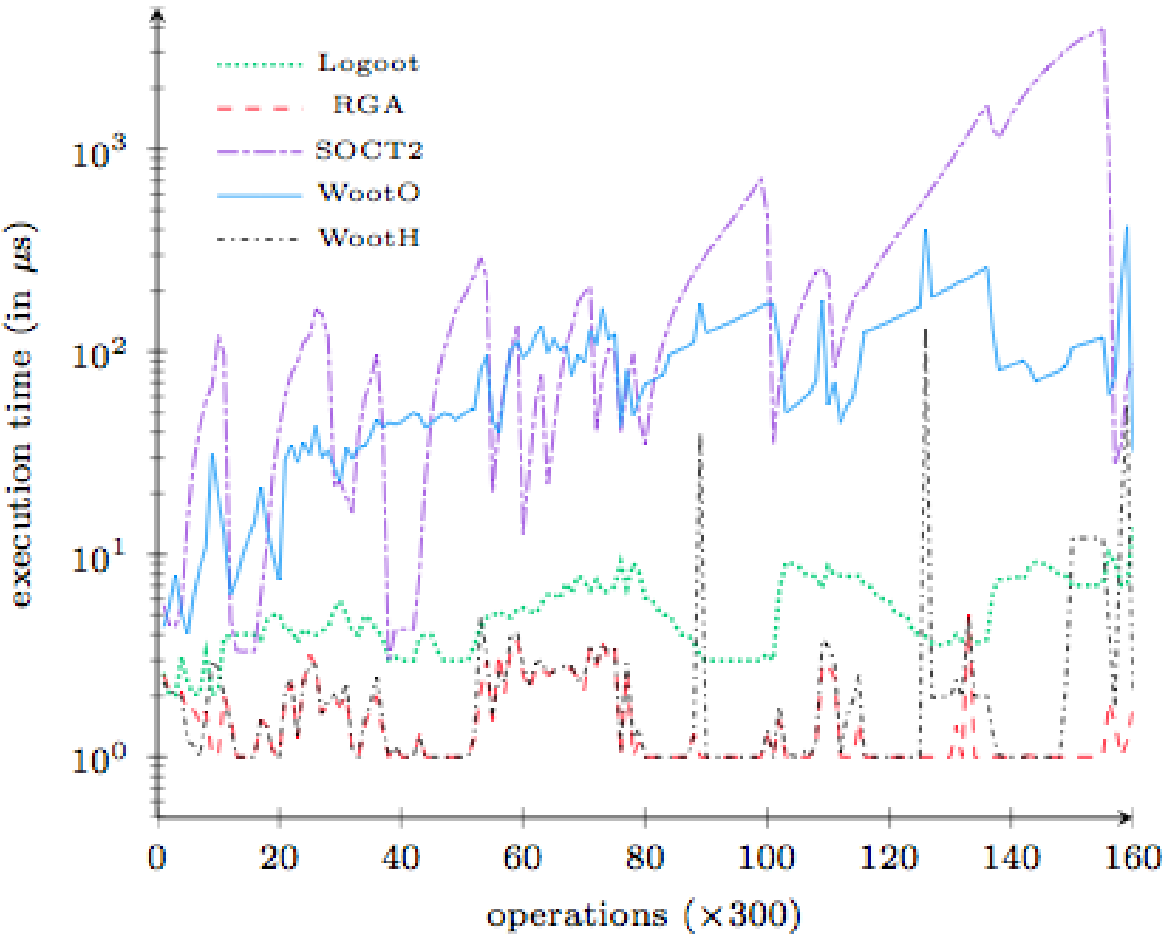
Users Operations : execution times - 2nd Group report



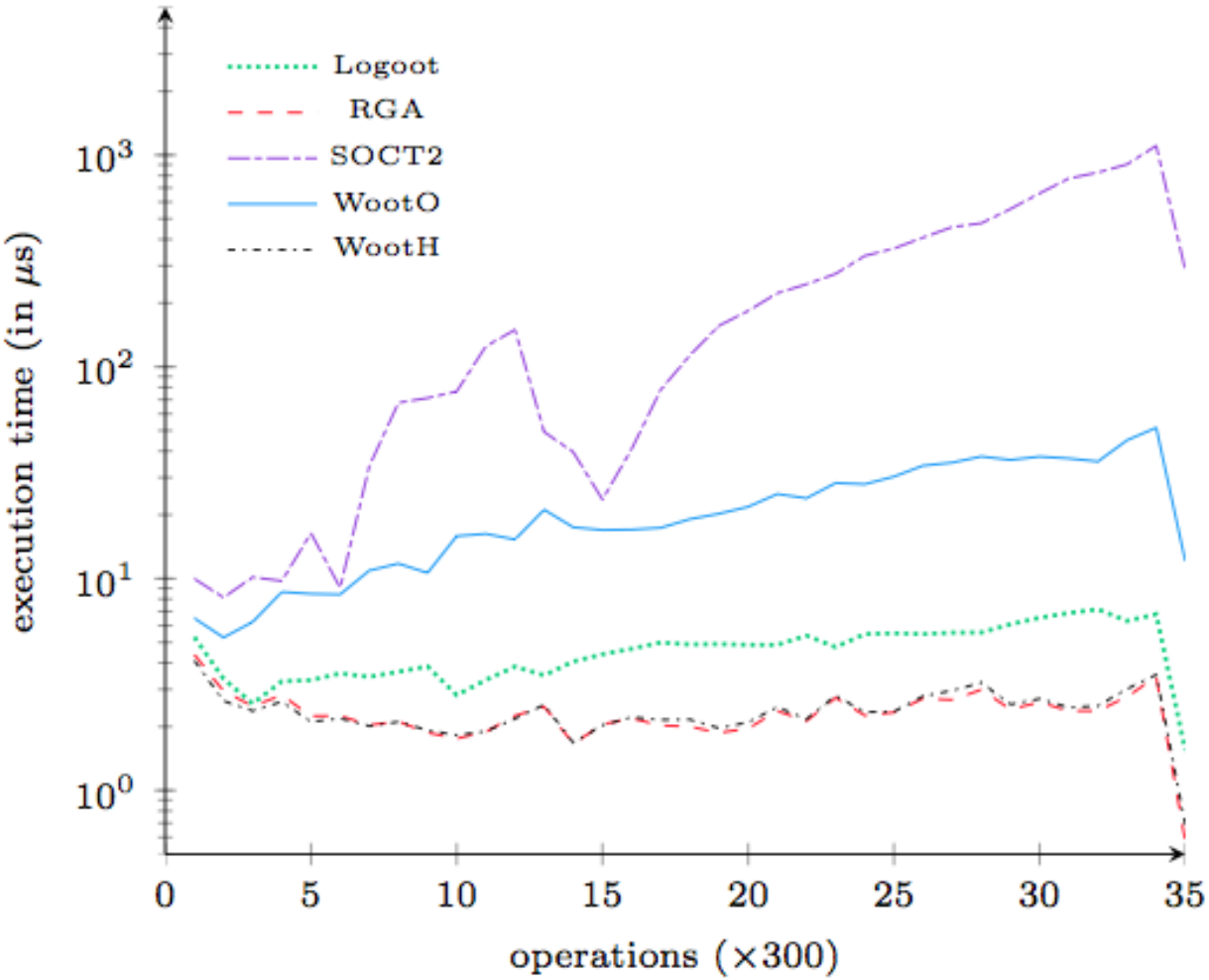
Users Operations : execution times - 1st series



Characters Operations : execution times - 2nd group report



Characters Operations : execution times - 1 time series



Conclusion

- First performance evaluation of algorithms with real collaboration traces including concurrency.
- Proves the suitability of CRDT algorithms in real-time collaboration.
- Outperform some representative operational transformation approaches.
- Well established for real-time collaboration in terms of local generation time and remote integration time.