SCert: Speculative Certification in Replicated Software Transactional Memories

Nuno Carvalho Paolo Romano Luís Rodrigues

INESC-ID/IST

May 31, 2011

Roadmap

Motivation

Related Work

SCert

Examples

Results

Conclusions

Roadmap

Motivation

Related Work

SCert

Examples

Results

Conclusions

Transactional Memory

- Set of mechanisms for shared memory access
- Uses de concept of *Transaction*

Programmers only indicate the set of operations that must be performed atomically: simpler than using Locks explicitly

Distributed Transactional Memory

Provides fault tolerance and increased performance

DSTM vs Distributed Shared Memory

- *Similar:* Hides the distribution from the programmers
- *Different:* Synchronization is only performed at the transaction boundaries

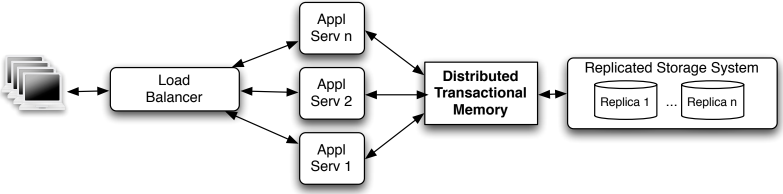
DSTM vs Replicated Databases

- *Similar:* Atomic Broadcast can be used to achieve a global serialization order
- *Different:* The relative overhead of the Atomic Broadcast is bigger

Example Application

Distributed transactional cache for multi-tier applications

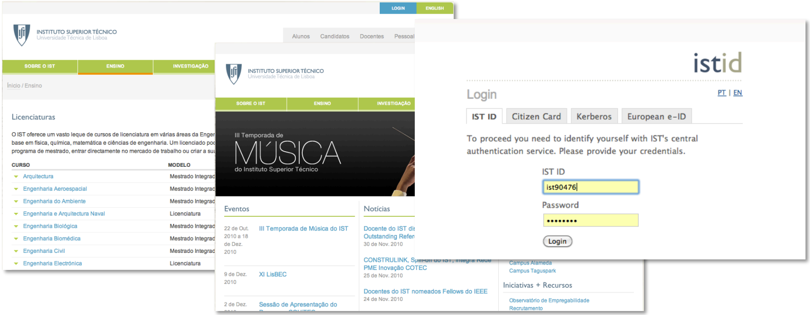
- Allows local processing of requests
- Detects both local and remote conflicts
- Alleviates pressure on back-end persistent storage



FenixEDU

University campus management system

- Used in an engineering school in Portugal
- Real system with real scalability and reliability issues



Início / Ensino

Licenciaturas

O IST oferece um vasto leque de cursos de licenciatura em várias áreas da Enger
base em física, química, matemática e ciências de engenharia. Um licenciado poc
programa de mestrado, entrar directamente no mercado de trabalho ou criar a sua

CURSO	MODELO
Arquitectura	Mestrado Integrad
Engenharia Aeroespacial	Mestrado Integrad
Engenharia do Ambiente	Mestrado Integrad
Engenharia e Arquitectura Naval	Licenciatura
Engenharia Biológica	Mestrado Integrad
Engenharia Biomédica	Mestrado Integrad
Engenharia Civil	Mestrado Integrad
Engenharia Electrónica	Licenciatura

Eventos

22 de Out. 2010 a 18 de Dez. 2010 — III Temporada de Música do IST

9 de Dez. 2010 — XI LisBEC

2 de Dez. — Sessão de Apresentação do

Notícias

Docente do IST di
Outstanding Refer
30 de Nov. 2010

CONSTRULINK, Spin-off do IST, integra Rede PME Inovação COTEC
25 de Nov. 2010

Docentes do IST nomeados Fellows do IEEE
24 de Nov. 2010

Campus Alameda
Campus Taguspark

Iniciativas + Recursos

Observatório de Empregabilidade
Recrutamento

Login

PT | EN

IST ID | Citizen Card | Kerberos | European e-ID

To proceed you need to identify yourself with IST's central authentication service. Please provide your credentials.

IST ID: ist90476

Password: ••••••••

Goals

Fault tolerance

Using replication schemes, already studied in other transactional systems (Databases)

Scaling up

Scale up in the number of STM instances to increase performance

Key Idea

- Use optimistic message deliveries to estimate the final transaction certification order
- Expose fresh (although possibly erroneous) data to new transactions
- Reduce the abort rate and detect conflicts earlier

Roadmap

Motivation

Related Work

SCert

Examples

Results

Conclusions

Distributed STMs

- Distributed Multi Versioning (Manassiev et al.)
- DiSTM (Kotselidis et al.)
- Cluster-STM (Bocchino et al.)

Fault tolerance is not the focus of previous work

Replicated STMs/DBMS

- Active replication without speculation:
  - (Kemme et al.) – uses optimistic total order to speedup commit but does not make speculative results visible
- Active replication with speculation
  - AGGRO (Palmieri et al.) – good for light weight transactions, as all nodes have to execute all transactions
- Certification without speculation
  - D^{2}STM (Couceiro et al.) and ALC (Carvalho et al.)

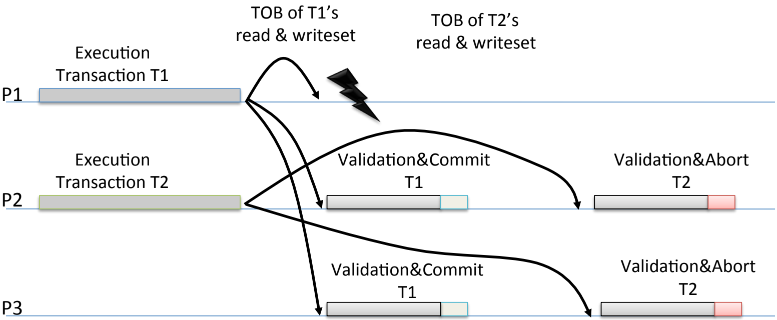
Related Work

	Active Replication	**Certification**
Non-Speculative	(Kemme et al.)	D^{2}STM and ALC
Speculative	AGGRO	SCert

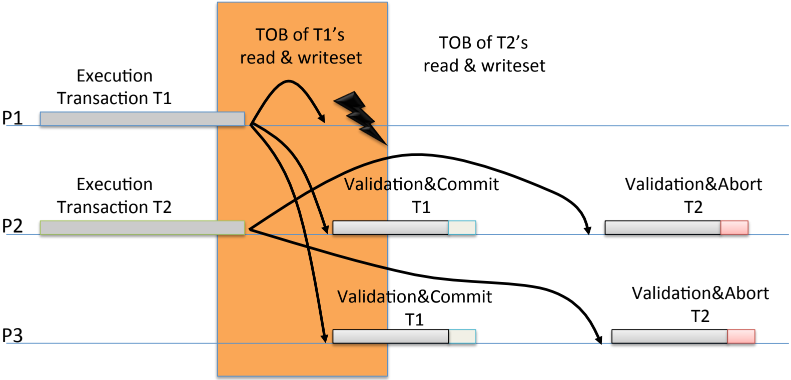
Replication Protocol Based on Certification

- Executes transactions in a single machine optimistically
- Transactions are certified only at commit time
- Exploits *Atomic Broadcast* to ensure replica consistency

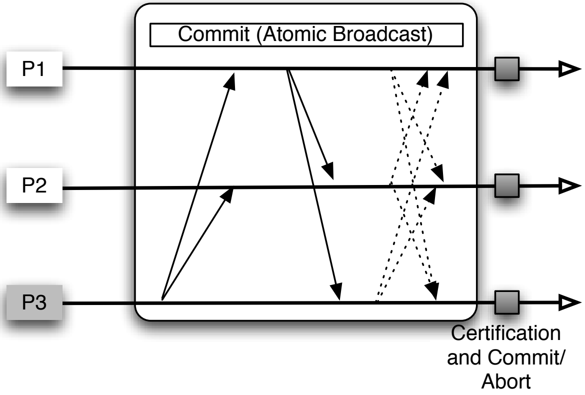
Baseline Replication Protocol



Baseline Replication Protocol



Certification Based Protocol

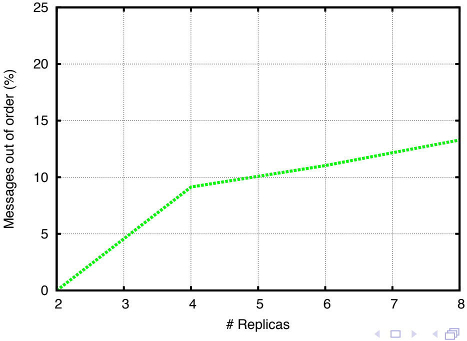


Problems of this Approach

- Loss of efficiency in high conflict scenarios
- Uses a heavy communication procedure (Atomic Broadcast)

Optimistic Atomic Broadcast (OAB)

- Delivers the message twice: an early estimate of the final order and the final order it self
- The estimated order matches the final order with high probability, on LANs



Roadmap

Motivation

Related Work

SCert

Examples

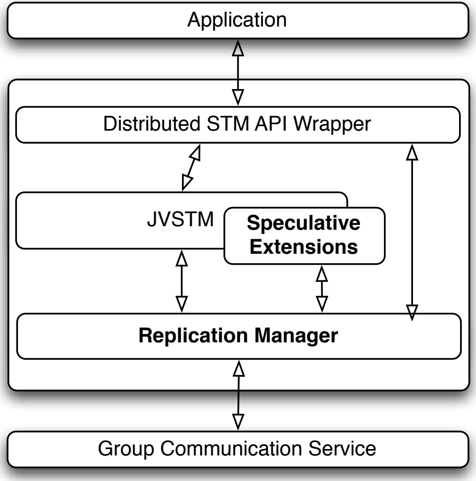
Results

Conclusions

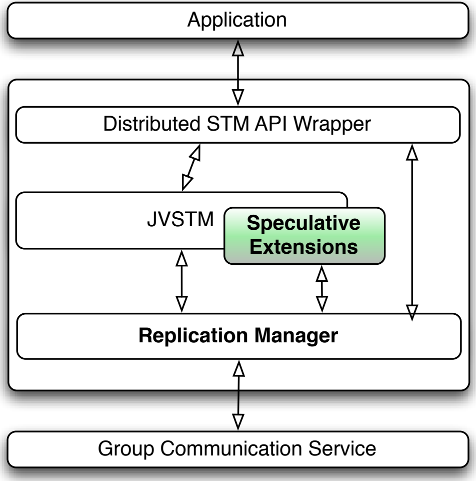
Speculative Certification (SCert)

- Certification based replication protocol
- Exploits *Optimistic deliveries of OAB* to generate fresh (but possibly erroneous) data
- New transactions read the optimistic data snapshots:
    - Provide executing transactions with fresher snapshots, reducing the probability of aborts
    - Detect conflicts earlier, reducing the amount of wasted computation and waiting time

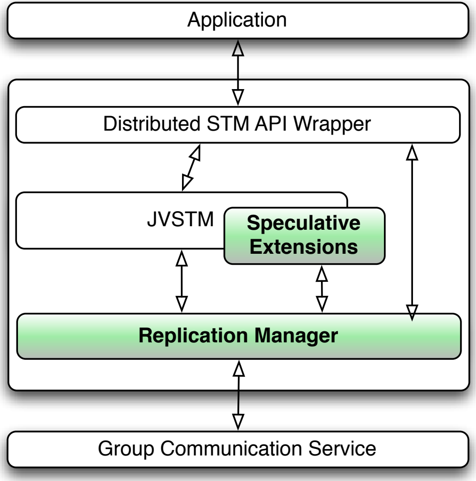
SCert: Architecture



SCert: Architecture



SCert: Architecture



Application

Distributed STM API Wrapper

JVSTM

Speculative Extensions

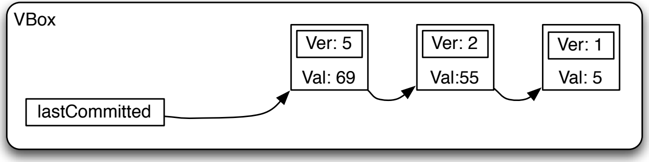
Replication Manager

Group Communication Service

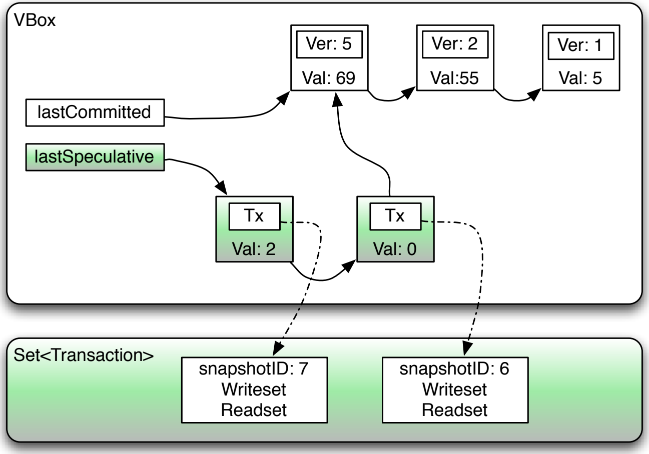
Speculative Extensions

- Provide the appropriate tools to expose speculative committed memory snapshots
- Speculative versions must be maintained
- The API must support speculative commits

JVSTM: Regular VBox



JVSTM Extensions for Speculative Transactions



SCert Replication Protocol (I)

- Transaction executes locally
- Upon Commit, the thread (locally) certifies the transaction and sends OAB
- Upon Optimistic Delivery, the transaction is certified and optimistically committed

SCert Replication Protocol (II)

- Upon Begin of new transactions, the new threads read the most fresh data (committed optimistically or finally)
- Upon Final Delivery:
    - Order matches: the transaction is marked as committed and the thread is unblocked
    - Order does not match: the optimistically committed snapshot is discarded and pending transactions must be re-certified

Roadmap

Motivation

Related Work

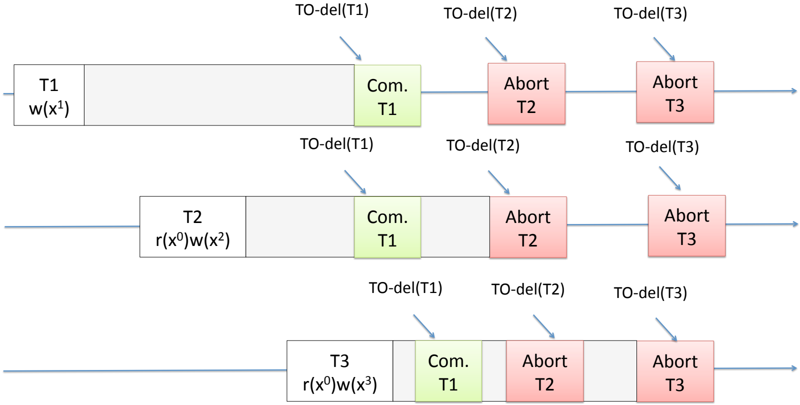
SCert

Examples

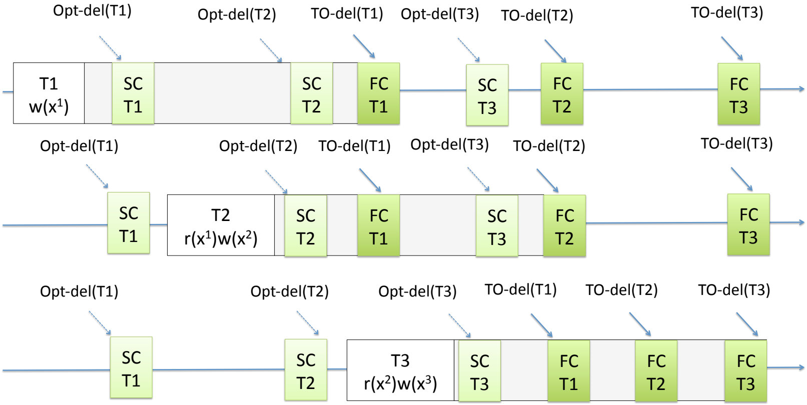
Results

Conclusions

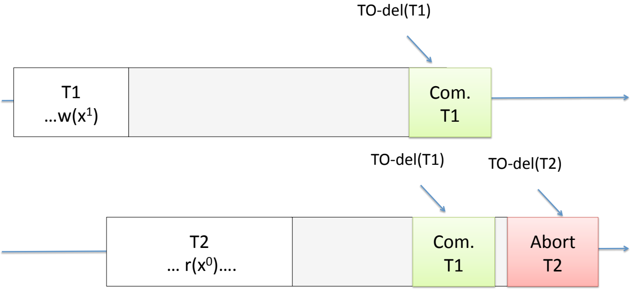
Regular Certification: Cascading Aborts Due to Conflicts



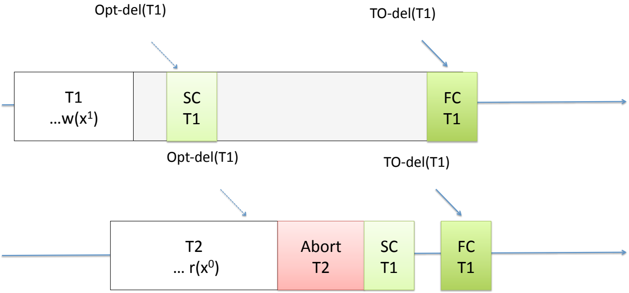
SCert: Cascading Commits



Regular Certification: Wasted Time



SCert: Early Notification



Roadmap

Motivation

Related Work

SCert

Examples

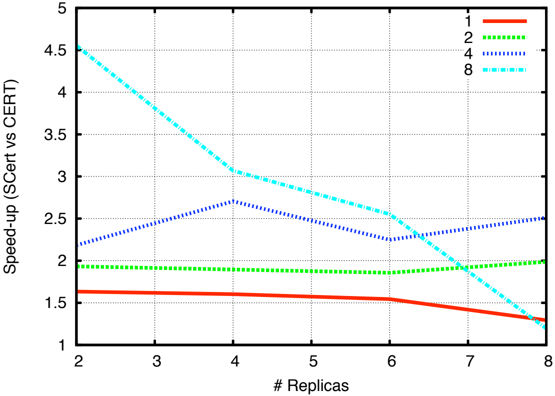
Results

Conclusions

Bank Benchmark: Full Conflict Scenario

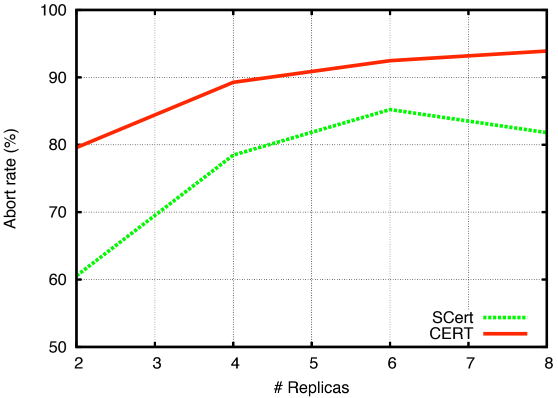
- Goal: worst case
- Replicas accessing the same memory region

Bank Benchmark: Throughput in Worst Scenario



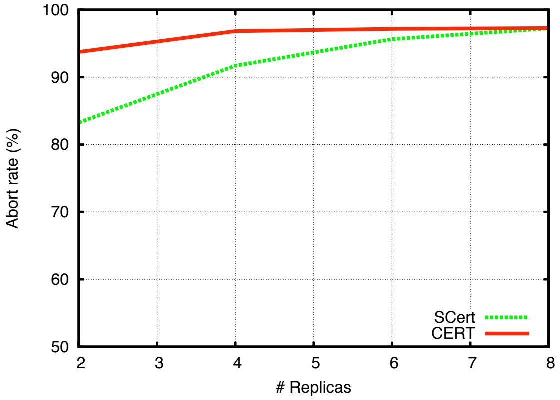
On avg. 1.5x speedup with one thread and up to 4.5x speedup with 8 threads per replica

Bank Benchmark: Abort Rate



1 thread per replica

Bank Benchmark: Abort Rate

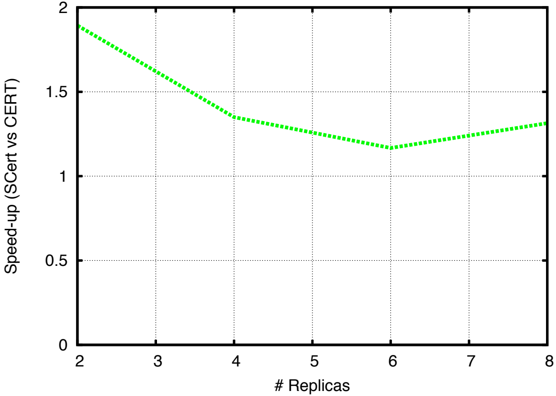


8 threads per replica

STMBench7 Benchmark: Scenario

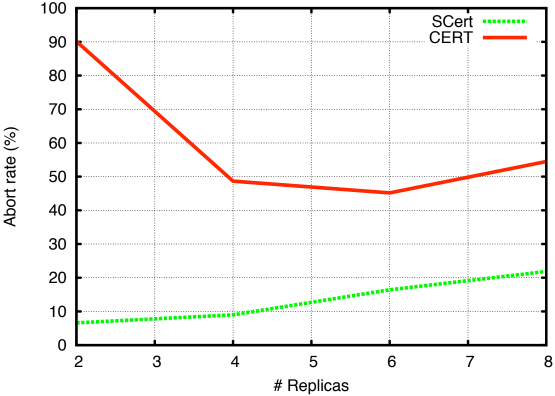
- Goal: more complex benchmark
- Richer benchmark featuring a number of operations with different levels of complexity over an object-graph with millions of objects
- Number of machines between 2 and 8
- Number of threads fixed to 2

STMBench7 Benchmark: Speedup



Almost twice speedup with a low number of replicas

STMBench7 Benchmark: Abort Rate



Roadmap

Motivation

Related Work

SCert

Examples

Results

Conclusions

Conclusions

- Reduce the number of transactions that read stale data
- Allows early detection of conflicts among transactions
- Performance improvements are achieved by exploiting optimistic deliveries of OAB
    - Up to 4.5x speed-ups

Thank you!

Questions?