Concurrences breaks State Machine Replication

SMR safety
Correct replicas produce matching replies

Today’s SMR implements safety through:
Agreement + deterministic execution
on request order	 typically sequential execution

Concurrency works against determinism
• replicas may diverge

Traditional SMR

Clients

Agree on order

Execute (sequential)

Replies must match

EV: Replicating Multithreaded Servers
Manos Kapritsos*, Yang Wang*, Vivien Quema†, Allen Clement‡, Lorenzo Alvisi* and Mike Dahlin*
The University of Texas at Austin* Grenoble INP† MPI-SWS‡

Flipping a 30-year-old design

SMR safety
without requiring determinism

Execute without agreeing on order
but verify that replicas did not diverge by checking replicas’ states and responses

Requires that:
• correct replicas seldom diverge
  Use heuristic to avoid executing obviously conflicting requests in parallel
• divergent replicas can be efficiently reconciled
  • efficient state comparison (Merkle tree)
  • efficient state rollback
  • efficient state transfer
  Copy-on-Write and versioning

10x throughput improvement

Allows multithreaded execution
while maintaining consistent replication

TPC-W benchmark
7.2x throughput improvement

Microbenchmark
12x throughput improvement

Fast recovery
through incremental state transfer

Masking Heisenbugs

Idea 1
• Leverage redundancy to mask concurrency bugs
• A bug is not masked only if it manifests in the same way on all replicas

Idea 2
• Use heuristic to avoid executing obviously conflicting requests in parallel → prevents most concurrency bugs from manifesting

Concurrency bugs
Bugs that may manifest during multithreaded execution but not during sequential execution

Masking a real bug

<table>
<thead>
<tr>
<th></th>
<th>Without heuristic</th>
<th>With heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Times bug manifest</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>Fixed with rollback</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>All identical</td>
<td>13</td>
<td>0</td>
</tr>
</tbody>
</table>

Results from running 750k TPC-W requests on a 2-way replicated H2 Database Server. Even when allowing all requests to execute in parallel, EV masks 82% of the bug occurrences. Application of a simple heuristic in request batching leads to the bug not manifesting at all.

Presenter: Allen Clement