BAYOU

Original Paper By: Douglas B. Terry, Marvin M. Theimer, Karin Petersen, Alan J. Demers, Mike J. Spreitzer and Carl H. Hauser

Presentation By: Patrick Coyle
WHAT IF?

- We’ve talked about systems that stay connected,
  - LAN, strong consistency, etc.
- What if we wanted our replicas in a distributed system to be mobile?
  - Cool! But how?
  - Won’t spotty connectivity be an issue?
**Solution: Bayou**

- Replicated, weakly consistent storage system
- Guaranteed to eventually be consistent
- Clients can execute writes immediately
  - Users can see all writes, even non-committed ones! (more about this later)
APPLICATION EXAMPLES

- Meeting room scheduler
  - Users want to schedule times in meeting rooms
  - Tentative reservations, must provide alternate times
  - Two users can try to book at the same time
- Bibliographic database
  - Database entries provide key-lookup
  - Keys created by author’s name, published year, plus any additional differentiating characters if needed
  - Must avoid two different papers being assigned the same key and duplicate copies of the same paper being assigned different keys.
  - May operate while disconnected to system
HOW DOES IT WORK?
HOW DOES IT WORK?

- Clients can read/write to any replica
- Clients can own a replica
- Writes are executed immediately on replica
- Replica can operate offline for extended periods of time
- Replicas communicate with each other when connected
WHAT ABOUT CONFLICTS?

- Let’s ask the application!
  - Application-specific detection and resolution of conflicts
  - Allows for optimizations over traditional, specifically-defined conflict detection
EXAMPLE

- I want to transfer $100 from my account
- Initial balance is $150
- A valid concurrent transaction deducts $50 from my account. This executes first
- Traditional conflict detection may detect that balance
- Application’s conflict detection method simply requires $100 to be left in my account
WE WANT CONSISTENCY!

- Reminder: Bayou is write-anywhere
  - Updates are written immediately to single server (After you checked for and resolved any conflicts of course, 😊)
  - Other servers wouldn’t have this new data!
- Bayou asynchronously propagates updates to other servers, when these other servers are connected to the network
COOL, BUT AM I THE ONLY USER?

- No! This might be a problem...
- Write Stability
  - Tentative Writes
  - Committed Writes
  - Writes given a local, monotonically increasing timestamp for ordering
COOL, BUT AM I THE ONLY USER?

- Who calls the shots (for final order of writes)?
  - Bayou chooses one replica as a primary
  - Primary determines commit ordering
  - Logs on each replica allow for necessary rollbacks of tentative writes to reach consistency
Figure 4. Bayou Database Organization
STORAGE ORGANIZATION

- Stable storage
  - Tuple Store Checkpoint – Current committed values written to stable storage
  - Write Log – Tail contains committed writes, followed by tentative. Tuple checkpoints included for recovery use

- Memory
  - Write log (same as above)
  - Current Tuple Store
  - Undo Log – Contains writes that need to be undone when a preceding write arrives during anti entropy
SECURITY/ACCESS CONTROL

- Public key cryptography
- Certificate Authority used
- 3 types of certificates
  - Privilege
  - Grant
  - Revoke
Performance

- Dependent on Implementation
  - Data store schema
  - Client server location
  - Running platforms
Table 1: Size of Bayou Storage System for the Bibliographic Database with 1550 Entries
(sizes in Kilobytes)

<table>
<thead>
<tr>
<th>Number of Tentative Writes</th>
<th>0 (none)</th>
<th>50</th>
<th>100</th>
<th>500</th>
<th>1550 (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Write Log</td>
<td>9</td>
<td>129</td>
<td>259</td>
<td>1302</td>
<td>4028</td>
</tr>
<tr>
<td>Tuple Store Ckpt</td>
<td>396</td>
<td>384</td>
<td>371</td>
<td>269</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>405</strong></td>
<td><strong>513</strong></td>
<td><strong>630</strong></td>
<td><strong>1571</strong></td>
<td><strong>4029</strong></td>
</tr>
<tr>
<td>Factor to 368K bibtex source</td>
<td>1.1</td>
<td>1.39</td>
<td>1.71</td>
<td>4.27</td>
<td>10.95</td>
</tr>
</tbody>
</table>
**Table 2: Performance of the Bayou Storage System for Operations on Tentative Writes in the Write Log**
(times in milliseconds with standard deviations in parentheses)

<table>
<thead>
<tr>
<th>Tentative Writes</th>
<th>0</th>
<th>50</th>
<th>100</th>
<th>500</th>
<th>1550</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Server running on a Sun SPARC/20 with Sunos</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undo all (avg. per Write)</td>
<td>0</td>
<td>31</td>
<td>(6)</td>
<td>70</td>
<td>(20)</td>
</tr>
<tr>
<td>Redo all (avg. per Write)</td>
<td>0</td>
<td>237</td>
<td>(85)</td>
<td>611</td>
<td>(302)</td>
</tr>
<tr>
<td></td>
<td>4.74</td>
<td>6.11</td>
<td>5.59</td>
<td>5.05</td>
<td></td>
</tr>
<tr>
<td><strong>Server running on a Gateway Liberty Laptop with Linux</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undo all (avg. per Write)</td>
<td>0</td>
<td>47</td>
<td>(3)</td>
<td>104</td>
<td>(7)</td>
</tr>
<tr>
<td>Redo all (avg. per Write)</td>
<td>0</td>
<td>302</td>
<td>(91)</td>
<td>705</td>
<td>(134)</td>
</tr>
<tr>
<td></td>
<td>6.04</td>
<td>7.05</td>
<td>7.01</td>
<td>6.4</td>
<td></td>
</tr>
</tbody>
</table>
## Performance

**Table 3: Performance of the Bayou Client Operations**  
(times in milliseconds with standard deviations in parentheses)

<table>
<thead>
<tr>
<th>Server Client</th>
<th>Sun SPARC/20 same as server</th>
<th>Gateway Liberty same as server</th>
<th>Sun SPARC/20 Gateway Liberty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read: 1 tuple 100 tuples</td>
<td>27 (19)</td>
<td>38 (5)</td>
<td>23 (4)</td>
</tr>
<tr>
<td></td>
<td>206 (20)</td>
<td>358 (28)</td>
<td>244 (10)</td>
</tr>
<tr>
<td>Write: no conflict with conflict</td>
<td>159 (32)</td>
<td>212 (29)</td>
<td>177 (22)</td>
</tr>
<tr>
<td></td>
<td>207 (37)</td>
<td>372 (17)</td>
<td>223 (40)</td>
</tr>
</tbody>
</table>
WRAPPING UP

- Designed for mobile, with spotty connectivity
- Write/Reads performed immediately and are visible to client
- Application-defined conflict resolution
- Guarantees eventual consistency
- Secure system, makes use of public key crypto
QUESTIONS?