Prophecy: Using History for High-Throughput Fault Tolerance

Siddhartha Sen, SOSP 2009

Joint work with: Wyatt Lloyd and Michael J. Freedman

Princeton University
Background & Goals

• Replication techniques that tolerate Byzantine (arbitrary) faults have poor throughput:
  – $3f + 1$ replicas needed to tolerate $f$ faulty replicas
  – Every replica participates in every operation

• **Goal**: Leverage properties of Internet services to improve throughput
  – Focus on read-mostly workloads
Traditional BFT

Clients

Agree?

Replica Group

application
D-Prophecy: A distributed sketcher

Clients

Replica Group

Agree?

sketcher

application
## D-Prophecy: A distributed sketcher

<table>
<thead>
<tr>
<th>Requests</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>sketch(req1)</td>
<td>sketch(resp1)</td>
</tr>
<tr>
<td>sketch(req2)</td>
<td>sketch(resp2)</td>
</tr>
<tr>
<td>sketch(req3)</td>
<td>sketch(resp3)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

### Replica Group

- sketcher
- application
D-Prophecy: A distributed sketcher

Clients → Replica Group

Agree? → sketch

full response
D-Prophecy: A distributed sketcher

Clients

Agree?

Replica Group

sketch

full response

sketch
D-Prophecy: A distributed sketcher

Clients

Load balancing

Replica Group

Sketch

Full response
D-Prophecy: A distributed sketcher

Trades off consistency for performance

- Traditional BFT: executes read at every replica, linearizability
- D-Prophecy: in-memory lookup at most, delay-once linearizability
D-Prophecy: A distributed sketcher

Delay-once linearizability

- Faulty replicas can return only stale (not arbitrary) data
- Load balancing limits repeated stale results
Internet services

• Unmodified clients
• Short-lived sessions
Internet services

• Unmodified clients
• Short-lived sessions
Internet services

- Unmodified clients
- Short-lived sessions
Internet services

Clients

Replica Group
Internet services

Clients

Proxy

Replica Group
Internet services

Consolidate sketch tables

Clients

Sketcher

Replica Group
Internet services

Consolidate sketch tables

Clients

Sketcher

Replica Group
Prophecy: A trusted proxy

Sketcher must be fail-stop, but...

Clients

Sketcher

Replica Group

Trusted
Prophecy: A trusted proxy

Already trust middleboxes for availability
Sketcher is *small* and *simple*
Prophecy: A trusted proxy

![Graph showing throughput vs write ratio]

Throughput (reqs/s) vs Write ratio:
- Prophecy
- Proxied-PBFT
Prophecy: A trusted proxy

![Graph showing throughput vs. write ratio for Prophecy and Proxied-PBFT.]

Throughput (reqs/s)

Write ratio

all reads
Summary

• Prophecy’s performance on reads approaches that of unreplicated service

• Relaxes consistency to delay-once semantics
Summary

• Prophecy’s performance on reads approaches that of unreplicated service

• Relaxes consistency to delay-once semantics
  – Not specific to BFT!
  – Can apply to Paxos, quorums, etc.