Agile Application-Aware Adaptation for Mobility

Brian D. Noble, M. Satyanarayanan, Dushyanth Narayanan, James Eric Tilton, Jason Flinn, Kevin R. Walker
School of Computer Science
Carnegie Mellon University
Outline

- Introduction
- Design Rationale
- Application-Aware Adaptation
- Design and Implementation
- Example Applications
- Evaluation
- Conclusion
Introduction

- Adaptation is the key to mobility.
  - Offer acceptable service in spite of the many problems that plague its existence.

- Odyssey
  - A set of extensions to the NetBSD to support adaptation.
Design Rationale

- **Motivation**
- **Fidelity**
  - The degree to which data presented at a client matches the reference copy at the server.
- **Concurrency**
  - Needs to be centralized monitoring and coordinated resource management on a mobile client.
Design Rationale

- Agility
  - The speed and accuracy with which it detects and responds to changes in resource availability.
  - Changed demand

- Minimalism
Application-Aware Adaptation

- Model of Adaptation
  - Application-aware adaptation.
  - Between system and individual applications.
  - System monitors resource levels, notifies applications of relevant changes.
Application-Aware Adaptation

- Realizing the model
  - In such an architecture, system code would treat data generically.
  - But it is impossible to optimize without some system-level knowledge of type.
- Wardens
  - Encapsulates the system-level support at a client necessary to effectively manage a data type.
- Viceroy
  - Responsible for centralized resource management.
Design and implementation

![Diagram showing Design and Implementation with Application, NetBSD Kernel, Odyssey, Viceroy, Warden3, Warden2, Warden1, and Interceptor. Diagram describes the flow of system calls and upcall between Application, NetBSD Kernel, and Odyssey layers.]
Design and implementation

(a) Resource Negotiation Operations

request(in path, in resource-descriptor, out request-id)
cancel(in request-id)

(b) Resource Descriptor Fields

resource-id
lower bound
upper bound
name of upcall handler
Design and implementation

handler(in request-id, in resource-id, in resource-level)

(d) Upcall Handler

tsoap(in path, in opcode, in insize, in inbuf, inout outsize, out outbuf)

(e) Type-Specific Operations

(tsoap(in path, in opcode, in insize, in inbuf, inout outsize, out outbuf)

(e) Type-Specific Operations
Example Applications

- **Video Player**
- Each movie is stored in multiple tracks at the server.
- **tsops**
  - Read a movie’s meta-data
  - Get a particular frame from a specified track.
Example Applications

- Web Browser
- Cellophane
  - makes use of the Odyssey API
  - selects fidelity levels.
- Distillation server
  - Fetches requested objects from the appropriate server.
  - Distills objects to the requested fidelity level.
Example Applications

- Speech Recognizer
- Warden decides whether it is faster to perform on the local slower CPU.
- Or to ship the larger raw utterance to the server.
Evaluation

(a) Step-Up

(b) Step-Down

(c) Impulse-Up

(d) Impulse-Down
Evaluation

- How Agile is Odyssey?
  - new = α(measured) + (1- α)(old)

\[
B = \frac{D}{T_{win} - (T_{rtt} / 2)}
\]
Evaluation

(a) Step-Up Waveform

(c) Impulse-Up Waveform

(b) Step-Down Waveform

(d) Impulse-Down Waveform
Evaluation

(a) 10% utilization/stream

(b) 45% utilization/stream

(c) 100% (attempted) utilization/stream
Evaluation

- How Beneficial is Adaptation?
- Video player
  - JPEG(99) 1.0
  - JPEG(50) 0.5
  - Black-and-white 0.01

<table>
<thead>
<tr>
<th>Waveform</th>
<th>B/W Fidelity = 0.01 Drops</th>
<th>JPEG(50) Fidelity = 0.5 Drops</th>
<th>JPEG(99) Fidelity = 1.0 Drops</th>
<th>Odyssey Drops</th>
<th>Fidelity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step-Up</td>
<td>0 (9.0)</td>
<td>3 (1.8)</td>
<td>169 (0.8)</td>
<td>7 (2.2)</td>
<td>0.73 (0.01)</td>
</tr>
<tr>
<td>Step-Down</td>
<td>0 (9.0)</td>
<td>5 (11.2)</td>
<td>169 (2.4)</td>
<td>25 (8.9)</td>
<td>0.76 (0.01)</td>
</tr>
<tr>
<td>Impulse-Up</td>
<td>0 (9.0)</td>
<td>3 (0.7)</td>
<td>325 (4.3)</td>
<td>23 (7.4)</td>
<td>0.50 (0.01)</td>
</tr>
<tr>
<td>Impulse-Down</td>
<td>0 (9.0)</td>
<td>0 (0.0)</td>
<td>12 (5.7)</td>
<td>14 (6.5)</td>
<td>0.98 (0.01)</td>
</tr>
</tbody>
</table>
## Evaluation

- **Web Browser**

<table>
<thead>
<tr>
<th>Waveform</th>
<th>JPEG(5) Fidelity = 0.05 Time (s)</th>
<th>JPEG(25) Fidelity = 0.25 Time (s)</th>
<th>JPEG(50) Fidelity = 0.5 Time (s)</th>
<th>Full Quality Fidelity = 1.0 Time (s)</th>
<th>Odyssey Time (s)</th>
<th>Fidelity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethernet</td>
<td>__</td>
<td>__</td>
<td>__</td>
<td>0.20 (0.00)</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Step-Up</td>
<td>0.25 (0.01)</td>
<td>0.30 (0.01)</td>
<td>0.29 (0.01)</td>
<td>0.46 (0.01)</td>
<td>0.35 (0.05)</td>
<td>0.78 (0.08)</td>
</tr>
<tr>
<td>Step-Down</td>
<td>0.25 (0.01)</td>
<td>0.30 (0.01)</td>
<td>0.29 (0.01)</td>
<td>0.46 (0.00)</td>
<td>0.35 (0.03)</td>
<td>0.77 (0.04)</td>
</tr>
<tr>
<td>Impulse-Up</td>
<td>0.27 (0.01)</td>
<td>0.33 (0.01)</td>
<td>0.34 (0.00)</td>
<td>0.71 (0.00)</td>
<td>0.42 (0.06)</td>
<td>0.63 (0.08)</td>
</tr>
<tr>
<td>Impulse-Down</td>
<td>0.24 (0.01)</td>
<td>0.27 (0.02)</td>
<td>0.29 (0.01)</td>
<td>0.34 (0.01)</td>
<td>0.36 (0.02)</td>
<td>0.99 (0.01)</td>
</tr>
</tbody>
</table>
Evaluation

- Speech Recognizer

<table>
<thead>
<tr>
<th>Waveform</th>
<th>Recognition Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Always Hybrid</td>
</tr>
<tr>
<td>Step-Up</td>
<td>0.80 (0.00)</td>
</tr>
<tr>
<td>Step-Down</td>
<td>0.80 (0.00)</td>
</tr>
<tr>
<td>Impulse-Up</td>
<td>0.85 (0.00)</td>
</tr>
<tr>
<td>Impulse-Down</td>
<td>0.76 (0.00)</td>
</tr>
</tbody>
</table>
Evaluation

- How important is Centralized Resource Management?

<table>
<thead>
<tr>
<th></th>
<th>Video</th>
<th>Web</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Drops</td>
<td>Fidelity</td>
<td>Seconds</td>
</tr>
<tr>
<td>Odyssey</td>
<td>1018</td>
<td>0.25 (0.00)</td>
<td>0.54 (0.02)</td>
</tr>
<tr>
<td>Laissez-Faire</td>
<td>2249</td>
<td>0.39 (0.01)</td>
<td>0.95 (0.03)</td>
</tr>
<tr>
<td>Blind-Optimism</td>
<td>5320</td>
<td>0.80 (0.00)</td>
<td>1.20 (0.00)</td>
</tr>
</tbody>
</table>

Diagram: 3 vertical bars labeled 1, 1, 1, 2, with numbers 1, 1, 1, 4 at the bottom.
Conclusion

- Odyssey architecture supports application-aware adaptation
- Confirms the feasibility of realizing this architecture
- Ability to support a wide range of applications