TierStore and DTN

Distributed Storage in Cambodia

Michael Demmer

This material is based in part upon work supported by the National Science Foundation under Grant No. 0326582.
Introduction / Outline

- Cambodia and the CIC Project
- The TierStore / DTN Technology
- Tier Cambodia Project Plan
Cambodia Overview

- Population: ~13 Million
- Land: ~175,000 sq. km
  - Roughly the size of Oklahoma
- GDP per capita (2004 est.): ~$2000
  - (measured in purchasing power parity)
  - Comparison: US: ~$40,000, India: ~$3100
- Population below poverty line: 40%

(Source: CIA - The World Factbook)
Cambodia Infrastructure

- Phones (2002 est.):
  - ~35,000 land, ~380,000 cell

- Internet access (2003 est.):
  - ~818 hosts, ~30,000 users

- Internet costs:
  - 64k wireless: $55/month
  - 56k dialup: $150/month
  - 9.6k mobile: $100-$300/month

(Source: CIA - The World Factbook)
Community Information Centers

- Asia Foundation project
- USAID Supported
- 22 centers
- (some closing)
- Basic Training / Internet Access
- All provinces
CIC Internet Access

Varied connectivity based on availability

- Broadband @ ~300kbps
- Cell-phone based dialup @ 9.6kbps
Introduction / Outline

Cambodia and the CIC Project

The TierStore / DTN Technology

Tier Cambodia Project Plan
DTN Overview

- Store-and-forward network
- Varied underlying technologies
- Novel routing, classes of service, reliability

This material is based in part upon work supported by the National Science Foundation under Grant No. 0326582.
TierStore Overview

- Persistent, replicated, storage using DTN as a transport
- Hierarchical storage model
  - Filesystem interface
- Eventually consistent
- Application-specific plugins
  - Email, http, log collection

```
/objects/<guid>
/web
/web/mirrors/...
/web/forms/...
/mail/<user>/inbox
/sensors/<field1>...
```

```
Tier
```

```
Technology and infrastructure for emerging regions
```

This material is based in part upon work supported by the National Science Foundation under Grant No. 0326582.
Why this approach?

- Fundamentally hard and expensive to get reliable network links
- Common infrastructure to deal with challenging network environment
  - Easy app integration
- Benefits from local caching, more efficient network usage
Introduction / Outline

- Cambodia and the CIC Project
- The TierStore / DTN Technology
- Tier Cambodia Project Plan
Tier Cambodia: Phase 1

- July 2005 Visit
  - Tour of centers, meetings with Asia Foundation
- Installed network monitoring software
  - Packet logger onto 1GB USB drive
  - Five centers w/ varied connectivity
- Obtained traffic logs from web proxy
Early Challenges

- Language, timezone
  - Debugging from the other side of the world
- Unexpected issues
  - ~1 week after we left, windows reinstall erased monitoring software
  - Funding shortage cut some internet access
  - Still optimistic...
**CIC Traffic analysis**

<table>
<thead>
<tr>
<th>HTTP Requests</th>
<th>Download</th>
<th>Upload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>7,917,754</td>
<td>60.9 GB</td>
</tr>
<tr>
<td>All Webmail</td>
<td>217,748 (3%)</td>
<td>7.5 GB (13%)</td>
</tr>
<tr>
<td>Yahoo CheckMail</td>
<td>33,130</td>
<td>2.6 GB</td>
</tr>
</tbody>
</table>

⚠️ Inefficiency of web-based email:
- Yahoo CheckMail avg download 78KB vs. mail spool avg message 12KB
- Doesn’t even count images, other URLs
Tier Cambodia: Phase 2

- January 2006 Visit
  - Prototype TierStore / DTN deployment

- Goals:
  - Demonstrate viability of the solution
  - Better access in poorly connected centers
  - Show cost benefit of asynchronous connectivity, store-and-forward
Results / Conclusions

Stay tuned...
Backup
CIC Usage Statistics

- Over 500,000 visits since opening the first center in 2003
- ~37,000 users

(April-September 2004 Statistics, courtesy of Pauline Tweedie)
Conflict Management

- All distributed storage systems must worry about conflicting operations
  - Intermittency makes this even harder

- How to deal with conflicts?
  - Subscription granularity
  - Conflict-free naming
  - “View” splits, app handlers