Threats to Preservation

David S. H. Rosenthal

LOCKSS Program
Stanford University Libraries
http://www.lockss.org/

© 2006 David S. H. Rosenthal
Optimism vs. Pessimism

- Two kinds of engineering
  - Optimistic – making good things happen
    - e.g. turbochargers
  - Pessimistic – preventing bad things happening
    - e.g. air-bags
- Preservation is 100% pessimistic
  - Goal is that nothing bad happen to content
- Pessimistic engineering = applied paranoia
Overview

- No system is perfect
  - How good does preservation need to be?
  - How good is preservation?

- What are the threats to preserved content?
  - How can we model & address them?
  - How can we measure how well we're doing?

- How can we set performance goals?
  - And improve cost-performance through trade-offs

- Preservation service level agreements?
  - Can they actually transfer responsibility?
“It is essential that ERA design proposals be analyzed against a threat model in order to gain an understanding of the degree to which alternative designs are vulnerable to attack. ...This initial threat modeling would be only the first step of a larger, iterative threat-countering process that involved designing against expected threats, observing failures that occur, and designing new countermeasures.”
Threats Not Isolated

• “Close examination of 6 case studies ... indicate that latent rather than active failures now pose the greatest threat to the safety of high-technology systems.” Reason *Human Error* (1990)

• Errors are correlated – for example:
  – Between drives in storage array (Talagala 1999)
  – Human error & hardware failure (e.g. TMI)

• Correlations make threat modeling difficult
A Start on Modeling

- Baker *et al.*, Eurosys '06
- Archive data are infrequently accessed
  - Can't depend on user access to detect errors
  - Must audit or *scrub* replicas against each other
  - Errors at any time, some *latent* until next audit
- Errors have correlation parameter $> 0$
- We ask: “How likely is a double failure?”
  - Second failure *after* first occurs
  - *Before* first failure detected and repaired

Lots of copies keep stuff safe
Using Our Model

• Model 2 replicas of part of Internet Archive
  – Using IA data on hashes of files over time
  – 30K hrs, 1.5M 50MB files, 1336 hash changes

• Auditing improves Mean Time To Data Loss
  – No audit, MTTDL 64 days
  – 4 month audit, MTTDL 3.4 years
  – 2 week audit, MTTDL 12.3 years

• Key is not to let latent errors fester
  – But auditing can be costly – IA turned it off
Well, Duh!

• Getting analytic model this far is hard
  – Need more replicas, threats, correlation
  – Thus need simulation not analytic model

• Getting good data to drive models is hard
  – IA data set noisy, short, old.
  – Others (NetApp, MSFT, ...) unavailable

• Better models could answer basic questions
  – For target reliability, *how much replication?*
    • Answer controls economics, thus sustainability
  – For target replication, *how to arrange replicas?*
    • Answer controls system architecture
Our Threat Model

- Media failure
- Hardware failure
- Software failure
- Network failure
- Obsolescence
- Natural Disaster
- Operator error
- Internal Attack
- External Attack
- Organization Failure
- Economic Failure
Media Failure

- No affordable media reliable enough
  - Both bit rot and catastrophic failure inevitable
- Need many independent replicas
  - Geographically, administratively, technologically
- Replicas must be audited frequently
  - Otherwise latent errors fester
- Routine access to, migration of replicas
  - Otherwise they likely won't work when needed
Hardware Failure

• Useful life of hardware < useful life of media
• Hardware must *flow through* the system
  – Rolling, desynchronized upgrade of replicas
  – Encourage diverse (=independent) hardware
• Better to add and delete replicas separately
  – Upgrade in place likely to synchronize errors
Software Failure

- Diversity & Randomization are keys
- Replicas with diverse implementations
  - down to operating systems => very expensive
  - protocols not software – replica interoperability
  - don't rule it out for the future
- Version skew is a start on diversity
  - Replicas spread across 3 versions
- Randomization is a form of diversity

Lots of copies keep stuff safe
Network Failure

- Both communication & services can fail
- $10^{-7}$ packets have undetected errors
  - End-to-end closed-loop checks essential
- Preservation systems use network services
  - Routing? DNS? NTP? Resolvers? ...
  - All have temporary or permanent failures
- High correlation with other failures
  - e.g. natural disaster, economic failure
Obsolescence

• Obsolescence isn't just for formats, software
  - although that's what's had all the attention
  - see our Nov 2005 D-Lib paper

• Format obsolescence is like prostate cancer
  - It's a serious, potentially fatal problem
  - If you live long enough you *will* suffer from it
  - No certain cure, no effective prophylactics
  - Odds are something else will kill you first
  - Watchful waiting is normally the best Rx
Natural Disaster

- Geographic distribution with fail-over
- Recovery should be automatic
  - The people will have better things to do
- Load-sharing much better than fail-over
  - Nothing special happens in a disaster
  - No-one needs to do anything
  - Much more likely to work (Patterson 2002)
Operator Error, Internal Attack

- High prevalence, massive under-reporting
  - [link](http://www.secretservice.gov/ntac/its_report_050516.pdf)

- Administrative independence essential
  - Replicas must be *peers* not masters & slaves
  - No central control => cooperating organizations

- Dual-key administration ineffective
  - Group-think, social engineering, ... => not independent

- Logs must be *tamper-proof*
  - Hard to ensure this
External Attack

• Diversity
  – of administration – social engineering
  – of jurisdiction – legal attacks
  – of software - vulnerabilities

• Paranoia
  – Constant security review – learn from OpenBSD

• Isolation
  – Dedicated hardware, aggressive packet filters
  – Off-line replicas? They can't be kept off-line ...
Organization Failure

- Succession planning
  - Fall-back sustainability?
  - Accepting custody of content is never free
- Open Source software, open formats are key
  - Without them, transfer may be too expensive
- SIP=DIP capability
  - Get out exactly what you put in
Economic Failure

- Sustainability is the fundamental problem
  - Bits vulnerable to interruptions in money supply
- Economic triage is inevitable
  - No-one has budget to keep all they want to keep
- Cost-performance trade-offs minimize triage
  - No-one has cost or performance data or models
- Cost-insensitive design is all too common
  - E.g. metadata quality vs. cost of acquisition vs. benefit
Measuring Performance

• Long-term storage is a big market
  – Without a performance benchmark!
  – Benchmarks drive mature tech markets

• My suggested benchmark: bit half-life
  – Look at a bit in a storage system
  – How long until 50% chance it has flipped?

• Technology cost/performance axes
  – Cost: $/bit/yr
  – Performance: bit half-life
A Reasonable Goal?

- How long do we need to keep data?
  - Libraries routinely keep paper for 100 years
  - Copyright is life + 70 years
  - SNIA “100-year Archive Task Force”

- 1PB, 100 years, 50% probability no damage
  - 1PB is a lot of data now ...
  - But in 100 years it will be $10^{-9}$ of a hard drive
How Hard Can It Be?

- 1PB, 100 years, 50% probability no damage
  - Sounds reasonable, doesn't it?
- That's a bit half-life of $10^{18}$ years
  - One hundred million times age of universe
  - Must measure really, *really* small effects
- Say the half-life of a bit on a disk is 10 years
  - That's a long service life for a drive
- Must amplify drive bit half-life by $10^{17}$
  - Even improbable events will have a big effect
Read the Fine Print

- Example from Amazon S3 license:
  - "AMAZON DOES NOT WARRANT THAT AMAZON WEB SERVICES ... WILL BE ACCESSIBLE ON A PERMANENT BASIS OR WITHOUT INTERRUPTION OR THAT THE DATA YOU STORE IN ANY SERVICE ACCOUNT WILL NOT BE LOST OR DAMAGED."

- All services disclaim liability the same way
  - So do all software components of preservation systems
  - Which is why the lawyers insist on adding them

- No players have any skin in the game
  - If things go wrong, its not their problem
## LOCKSS Monitoring

<table>
<thead>
<tr>
<th>Archival Units</th>
<th>Content Size</th>
<th>Disk Usage (MB)</th>
<th>Peers</th>
<th>Polled</th>
<th>Status</th>
<th>Last Poll</th>
<th>Last Crawl</th>
<th>Last TreeWalk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Volume</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 1</td>
<td>309077</td>
<td>3.0 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>17:09:54</td>
<td>13:13:17</td>
<td>16:56:58</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 3</td>
<td>774857</td>
<td>7.1 peers</td>
<td>0.0001</td>
<td>96% Agreement</td>
<td>12:01:06</td>
<td>11:29:06</td>
<td>12:01:06</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 4</td>
<td>5,218,273</td>
<td>18.2 peers</td>
<td>0.0001</td>
<td>92% Agreement</td>
<td>12:01:06</td>
<td>11:29:06</td>
<td>12:01:06</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 5</td>
<td>5,995,846</td>
<td>4.4 peers</td>
<td>0.0001</td>
<td>Waiting for Poll</td>
<td>09:27:06</td>
<td>11:20:06</td>
<td>12:01:06</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 6-7</td>
<td>1,957,169</td>
<td>12.4 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>12:01:06</td>
<td>08:39:56</td>
<td>17:51:26</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 8</td>
<td>865,254</td>
<td>1.5 peers</td>
<td>0.0001</td>
<td>99% Agreement</td>
<td>12:01:06</td>
<td>14:51:00</td>
<td>16:48:51</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 9</td>
<td>565,216</td>
<td>1.5 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>12:01:06</td>
<td>14:51:49</td>
<td>16:20:58</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 10</td>
<td>494,278</td>
<td>0.7 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>12:01:06</td>
<td>14:54:52</td>
<td>17:06:15</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 11-12</td>
<td>1,710,078</td>
<td>2.0 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>12:01:06</td>
<td>13:11:35</td>
<td>17:42:18</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 13</td>
<td>790,924</td>
<td>1.1 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>12:01:06</td>
<td>13:14:10</td>
<td>16:28:48</td>
<td></td>
</tr>
<tr>
<td>Applied Semiotics / Sémiotique appliquées Volume 14</td>
<td>814,204</td>
<td>1.1 peers</td>
<td>0.0001</td>
<td>100% Agreement</td>
<td>12:01:06</td>
<td>13:08:39</td>
<td>17:24:04</td>
<td></td>
</tr>
</tbody>
</table>

**LOT OF COPIES KEEP STUFF SAFE**
Where Are We?

- Sustainability is the fundamental problem
  - Adequate bit half-life @ affordable $/bit/yr
  - Adequate bit half-life is a very aggressive target

- Cost & performance models unrealistic
  - Dynamic costs, multiple correlated threats, ...
  - Many hard-to-quantify threats poorly understood
  - Very hard to benchmark system performance

- Not a good place to be
  - Better models + better data is the place to start
Work Done By

- **LOCKSS Research Team (since 2001)**
  - Mary Baker, Mehul Shah & colleagues @ HP Labs
  - Mema Roussopoulos & students @ Harvard CS
  - Petros Maniatis & interns @ Intel Research Berkeley
  - Support: NSF, HP, Intel, Sun

- **LOCKSS Engineering Team (since 1998)**
  - Tom Lipkis, Tom Robertson, Seth Morabito, Thib G.
  - Special thanks to Mark Seiden
  - Support: LOCKSS Alliance, Mellon, Library of Congress