Using Hard Disks for Digital Preservation

David S. H. Rosenthal
Stanford University Libraries, Stanford, California, USA

Mema Roussopoulos
Computer Science, Harvard University, Cambridge, Massachusetts, USA

TJ Giuli
Computer Science Department, Stanford University, Stanford, California, USA

Petros Maniatis
Intel Research, Berkeley, California, USA

Mary Baker
HP Labs, Palo Alto, California, USA

Abstract

The LOCKSS system is a tool librarians can use to preserve long-term access to content published on the web. It has three main functions. It *collects* the content by crawling the pub-lisher's web sites, it *distributes* the content by acting as a proxy for reader's browsers, and it *preserves* the content through a cooperative process of damage detection and repair. The system uses the hard disk holding the copy used for access as a preservation medium; the cooperative damage detection and repair mechanism eliminates the need for off-line backups on removable media. We describe the LOCKSS system as an example of the techniques needed to use hard disks as a medium for long-term preservation.

1. Introduction

The digital information normally considered to be a candidate for preservation is initially stored on hard disks. It is needed there to provide the instant access that makes digital information so much more useful in practice than information on paper. With good reason, digital preservation systems typically consider this too risky a medium for long-term storage. They create backup copies of the information on media that are thought to be more persistent, and store them off-line where they are thought to be less at risk.

Unfortunately, the one thing everyone agrees on in the field of digital preservation is that there is not enough money to do the job. Backing up computer systems is expensive, and notoriously hard to do correctly for the long term. Adding the costs of creating, storing, auditing, copying and recovering off-line media to the costs of maintaining the on-line access copy may place the whole process of digital preservation beyond the budgets of the institutions involved.

Since early 1999, Stanford University has been developing the LOCKSS* system for preserving access to academic journals published on the Web; the system is currently being tested at libraries around the world. The primary goal of the LOCKSS program is to make it affordable for many librarians to preserve access not merely to the major scientific, technical and medical journals, but also to those transient and less-formal but critical journals in the humanities, by providing an e-journal preservation tool that is very cheap to operate.

The LOCKSS design emulates the system by which libraries preserve access to academic journals on paper, scattering large numbers of relatively vulnerable copies around the world and using peer-to-peer cooperation among libraries to make the system as a whole much more reliable than any individual component.

Each LOCKSS peer serves three main functions. It *collects* the content by crawling the publisher's web sites, it *distributes* the content by acting as a proxy for readers' browsers, and it *preserves* the content through a cooperative process of damage detection and repair. In particular, there is no need in the LOCKSS system for a library to back up their copy of the preserved content. Each copy takes part in a continual, slow, autonomous audit process that detects any

damaged or missing copies, and repairs them automatically from other copies. Even if an entire disk is lost, its contents can be recovered via this form of inter-library copying.

The LOCKSS system thus provides an example of the techniques that are needed to use hard disks as a long-term preservation medium. In this paper we provide a brief overview of the economics of digital preservation, the characteristics of hard disks, their advantages and disadvantages for this purpose, and describe the techniques we have developed to exploit the advantages and overcome the disadvantages while keeping the system affordable.

2. Economics of Digital Preservation

A digital preservation system performs three major functions, and each has costs:

- *Acquiring* the material to be preserved can be costly, not just in subscription costs but also first in the costs of negotiating with the publisher to get the rights to preserve it,[†] and second in the technical process of ingesting the content.
- *Distributing* the material to readers on request can be costly, especially if the access copy is no longer available and the requested item must be retrieved from archival offline storage.
- *Preserving* the material can be costly, especially when each off-line backup copy must be regularly retrieved from archival storage, audited, and if necessary migrated to a new storage medium or a new format.

Another way of looking at the economics of preservation is to set a goal that the preservation costs that will accumulate through the entire history of preserving a year of a journal should amount to no more than say 10% of the cost of acquiring it. A journal costing say $1000/year and staff costing $50/hour mean that everything done to that year of that journal through the entire preservation process must amount to no more than 120 minutes of staff time.

Suppose there is a reader access every 5 years, an audit every 3 years, a media migration every 7 years and a format migration every 30 years. In the first 100 years of preservation there would be 70 of these accesses. If we assume everything but staff time is free, these operations have to average less than 2 minutes of staff time each if the 10% budget is to last the first 100 years.

3. Threats to Digital Preservation

Discussion of digital preservation typically focusses on the technical challenges of ingest, collection management, access and format migration. These are indeed formidable, but a practical digital preservation system must guard against a much broader spectrum of threats:

- *Economic* threats are perhaps the most serious. No institution has an adequate budget for digital preservation, and most institutions have no budget at all.[2] A digital preservation system that is controlled and administered centrally or lacks a robust business model can fail catastrophically with a single budget cut.
- *Human Error* is a risk that increases as budget constraints reduce the quality of both system administration and its supervision.
- *Disasters* such as floods and earthquakes must be anticipated. Surviving them requires storing multiple replicas of the content at geographically dispersed locations.
- *Attacks* on preservation systems are inevitable. In the absence of skilled system administration they are difficult to thwart and expensive to recover from.

Unfortunately, guarding against most other foreseeable risks raises the cost of preservation, increasing the risk of economic failure.

4. Hard Disk as a Preservation Medium

The ideal preservation medium would be write-once, last forever and need no power. Hard disks have none of these characteristics. They are inherently writable, they last about 5 years in service, and they need power. They do, however, have a number of advantages:

- They package the read/write technology together with the information it accesses, obviating the need for the preservation of obsolete tape drives, etc.
- They achieve remarkable storage density in terms of bytes per unit volume.
- They are surprisingly robust. For example, almost no data was lost from the disks submerged when the basement of Stanford's Green Library flooded.[3]
- They are a mass-market technology, with a decades-long history of rapid cost-per-byte reduction and capacity growth.[4]

The key requirements of a digital preservation system using hard disk as the preservation medium are:

- *Replication* to survive the inevitable failures.
- *Audit* to detect the inevitable failures.
- *Copying* to create the replicas and repair the damage.
- *Automation* to reduce the per-replica cost and minimize economic threats.
- *Diversity* to prevent epidemic failure.

The LOCKSS system is designed to provide each of these, and all but diversity have been implemented in the current system (see Section 9).

5. Techniques

The techniques used by the LOCKSS system to preserve e-journals on hard disks include:

- Each peer preserves the same copy that it uses for access, in the same way that libraries keep copies of academic journals in the stacks and let readers use them. Librarians running the system cooperate informally to ensure that material of interest is held by large numbers

of peers; we expect popular material to have hundreds of replicas.
- Each peer continually audits its copy against others' to detect damaged or missing data, using a peer-to-peer process of voting on the hash of the stored files.[5]
- Each peer automatically repairs any damaged or missing data discovered by the auditing process, including total loss caused by catastrophic disk failure.
- Eventually, typically after 4-5 years, a peer's disk capacity fills up. When this happens, the institution supplies a new, empty peer and the full one clones itself over the network to the new one. Disk drive technology progress means that the contents of the old peer will occupy some 3-9% of the new peer.[6] The old peer can be left running until it fails, increasing the number of copies and thus the system's reliability, or it can be turned off.

None of these operations require significant attention from a system administrator, contributing greatly to keeping the system affordable. In a more conventional system, each of them involve costly attention from a system administrator:
- The system administrator has to create several backup copies on removable media and transfer them to suitable storage.
- On a regular schedule, the removable media must be audited; this involves removing media from storage, reading them, and verifying that the data read back are correct. In addition, the access copy must be verified at intervals.
- In either case, if damage is detected it must be repaired either from the access copy or from another of the backup copies on removable media.
- Eventually, the usable lifetime of the removable media expires or (more likely) the technology becomes obsolete. When this happens, the entire stock of backup copies must be retrieved from storage and migrated to a new medium and technology.

6. Implementation

An individual LOCKSS peer is implemented in three functional layers:
- The *platform* layer provides a robust, secure Java virtual machine. Our current implementation is as a "network appliance"[7]: a single-function computer system that can be installed as part of an institution's network and left to function with little administrative attention. It is distributed as a bootable CD image that runs a specially modified version of the free, Open Source OpenBSD operating system on generic, low-cost PC hardware.
- The *daemon* is a free, Open Source Java application that performs the three functions of the LOCKSS system:
  - It *collects* the journal pages as they are published by crawling the publisher's web site.
  - It *distributes* these pages by acting as a web proxy for readers' browsers, supplying the publisher's copy if it is available and the local copy if it is not.
  - It *preserves* the collected data by implementing the peer-to-peer auditing and repair mechanism outlined above.
- *Plugins* are small, downloadable Java programs that adapt these generic functions to the particular ways publishers organize their web-sites. For example, the plugin for an e-journal may know about publishing frequency, restrictions on the times and ways the site may be crawled, and the vagaries of the system that inserts advertisements into the pages.

Four years into the LOCKSS program, the system is in preproduction testing at over 80 libraries worldwide. About 50 publishers representing over 1000 titles are supporting the program.[8] We expect the transition to full production use this year.

7. Genres of Content

The LOCKSS system was originally designed to allow librarians at Universities and other institutions to preserve the web editions of the scientific, technical and medical journals to which they subscribe. These subscriptions are expensive, and growing more so,[2] leading librarians to fear that when they are forced to cancel a subscription they will lose access to the back content for which they paid. The fear that the content will be lost to society as a whole is less pressing; librarians expect major publishers and academic societies to survive. However, as librarians came to understand the capabilities of the LOCKSS system they added two genres that are at more immediate risk of loss:
- *Humanities Journals*. Much of the more interesting work in the humanities now appears only on-line in small, informal web "journals". The finances of these web publishers are typically parlous, and unless librarians take action to preserve them, the materials are at great and immediate risk. The LOCKSS team is testing the use of the system on several such journals (a list is at Ref. [9]). A group of humanities specialists from major US Universities is choosing a set of the highest-priority publishers of these journals with whom to work and arrange for the preservation of their materials.
- *Government Documents*. Material governments publish on paper is distributed and preserved by a network of government documents libraries, such as the US Federal Depository Library Program (FDLP). These not merely provide access to the public, and a way of preserving the material that curators deem important, but also a mechanism that keeps Governments honest. Altering or with drawing material once published leaves a paper trail. Governments are rapidly switching the bulk of their publishing to the Web. There are no corresponding mechanisms for preservation, and material can be changed or removed at the whim of the publishing agency. The LOCKSS team is working with the US Government Printing Office and the FDLP librarians to

apply the LOCKSS technology as a way of migrating the FDLP into the electronic age.[10]

In both cases, the low cost of the technology and its inherently distributed and collaborative nature are attractive features.

8. Related Work

The Rosetta Project[11] is preserving information using extremely reliable storage (a micro-engraved nickel disk projected to last at least 2000 years) and scattering large numbers of replicas. Information is stored in analog form to ensure long-term readability. The project also publishes the material on-line to provide access.

RAID (Redundant Arrays of Inexpensive Disks)[12] is a popular technique using small numbers of tightly coupled local replicas to increase the reliability of disk storage. For our purposes, RAID would increase the per-peer costs substantially (increasing the economic threat) but would not provide protection against many real threats such as attack, human error and disaster. The extra replicas would be more effective remotely, at additional peers, where being loosely coupled they assist in defending against a broader range of threats.

Peer-to-peer systems have been proposed for large-scale, persistent storage. Some of these systems (e.g., Intermemory[13]) use RAID-like cryptographic sharing techniques to distribute n partial replicas from any $m<n$ of which the original can be reconstituted. The goal of the LOCKSS system is to allow librarians to take custody of the material to which they subscribe; the relative costs of staff time and hardware mean that the administrative, legal, and operational advantages of each library having its own complete copy far outweigh the hardware costs of the additional replicas. Others (e.g., PAST[14]) distribute many complete copies but do not allow control over where the copies are located. This implies a level of trust hard to achieve and maintain for the long term among independent institutions.

The Mellon Foundation (which funds the LOCKSS system) is also funding JSTOR to investigate establishing a centralized, subscription-based archive for e-journals.[15] In the context of developing a business plan for this effort a study[16] was conducted that demonstrated the potential the transition from paper to electronic journals offers libraries in terms of reducing the non-subscription costs of maintaining their collection. In the absence of a widely-adopted system for preserving e-journals, the study had to exclude these costs, while noting that some of the potential cost savings in other areas could be redirected to preservation costs. The 25-year non-subscription costs they identify for both paper and electronic formats are typically smaller than our assumed 10% of subscription, reinforcing our point that minimizing the costs of digital preservation is essential for success.

9. Future Work

The immediate next step for the LOCKSS system is to transition from testing to actual production use, and ramp up its use both in terms of numbers of copies and journals. We expect this to happen before the middle of 2004. As deployment proceeds, we plan to gather data on the costs of running the system in production, to validate our approach to making the system affordable.

In the near future, we will replace the current protocol[5] that peers use to communicate with each other with a new protocol[17] that has better scaling and attack-resistance properties.

The next priority is to evolve away from our vulnerable monoculture by introducing diversity (see Section 4) into the layers of the system (see Section 6):

- The platform layer need only provide a Java virtual machine to run the daemon. We have already demonstrated that the daemon can run on various Unix-like operating systems; packaging it for other operating systems should not be an onerous task.
- The daemon needs to implement the peer-to-peer protocol and support the necessary plugins. An independent, clean-room implementation of the daemon in Java would be easy and would provide some diversity; an implementation in some other language would provide more diversity at the much higher cost of cross-language support for a Java enviroment in which the plugins could run.
- Providing diversity at the plugin layer is a more complex problem requiring further study. A more productive medium-term plan is to provide a highly restrictive Java security manager (or sandbox) in which to run plugins.

In the longer term, we will investigate the possible use of network attached storage (NAS) in the LOCKSS system. The advent of low-cost NAS disks may allow a single low-cost PC to manage a much larger amount of low-cost disk storage than with existing ATA disks, and thus improve the system's economics further.

10. Conclusions

The LOCKSS system is demonstrating that the hard disk can be an effective and affordable medium for long-term digital preservation, provided that enough institutions are motivated to pay a small premium in money and effort to preserve access to some information cooperatively. Key to this is the very high level of automation the system achieves, and thus the very low level of staff costs involved. Both are made possible by preserving the access copy itself, rather than making and preserving a separate backup copy. This in turn is made possible by a set of techniques that overcome the disadvantages of hard disk as a preservation medium, primarily mutual audit and repair among a large number of independent replicas.

11. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 9907296, however any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The LOCKSS program is grateful for support from the National Science Foundation, the Andrew W. Mellon Foundation, Sun Microsystems Laboratories, the Open Society Institute and Stanford Libraries.

References

1. Electronic Frontier Foundation, Digital Millennium Copyright Act (DMCA) Archive, http://www.eff.org/IP/DMCA/.
2. Association of Research Libraries, ARL Statistics 2000-01, http:// www.arl.org/stats/arlstat/01pub/intro.html(2001).
3. Sami Menafee, Drivesavers cuts price for lost data recovery, http://www.findarticles.com/cf_dls/m0NEW/n39/ 20324409/p1/article.jhtml(1998).
4. Clayton M. Christensen, The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail, Harvard Business School Press (1997).
5. David S. H. Rosenthal and Vicky Reich, Permanent Web Publishing, in Proceedings of the USENIX Annual Technical Conference, Freenix Track (Freenix 2000), San Diego, CA, USA, pp. 129–140 (2000).
6. Edward Grochowski, Emerging Trends in Data Storage on Magnetic Hard Disk Drives, Datatech, pp. 11–16 (1998).
7. David S. H. Rosenthal, A Digital Preservation Network Appliance Based on OpenBSD, in Proceedings of BSDcon 2003, San Mateo, CA, USA (2003).
8. LOCKSS program, Project Status, http://lockss.stanford.edu/projectstatus.htm.
9. LOCKSS, Lockss title registry, http://lockss.stanford.edu/ titleregistry.html (2004).
10. LOCKSS-DOCS, Lockss-docs: Exploring distributed access to web-based us government information, http://lockss-docs.stanford.edu/ (2004).
11. The Long Now Foundation, The Rosetta Project, http://www.rosetta.org.
12. David A. Patterson, Garth Gibson and Randy H. Katz, A Case for Redundant Arrays of Inexpensive Disks (RAID), in Proceedings of the ACM SIGMOD International Conference on Management of Data, Chicago, IL, USA, pp. 109–116 (1988).
13. Yuan Chen, Jan Edler, Andrew Goldberg, Allan Gottlieb, Sumeet Sobti and Peter Yianilos, A Prototype Implementation of Archival Intermemory, in International Conference on Digital Libraries, Berkeley, CA, USA, pp. 28–37 (1999).
14. Antony Rowstron and Peter Druschel, Storage Management and Caching in PAST, A Large-scale, Persistent Peer-to-peer Storage Utility, in Proceedings of the Eighteenth ACM Symposium on Operating Systems Principles, Chateau Lake Louise, Banff, AB, Canada, pp. 188–201 (2001).
15. JSTOR, The challenge of digital preservation and jstor's electronic-archiving initiative, http://www.jstor.org/about/ earchive.html (2004).
16. Roger C. Schonfeld, Donald W. King, Ann Okerson and Eileen Gifford Fenton, Library periodicals expenses: Comparison of non-subscription costs of print and electronic formats on a life-cycle basis, D-Lib Magazine, 10(1) (2004).
17. Petros Maniatis, Mema Roussopoulos, TJ Giuli, David S. H. Rosenthal, Mary Baker and Yanto Muliadi, Preserving Peer Replicas By Rate-Limited Sampled Voting, in Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, Bolton Landing, NY, USA, pp. 44–59 (2003).

* LOCKSS is a Trademark of Stanford University. It stands for "Lots Of Copies Keep Stuff Safe."

† Unlike paper, a consequence of the DMCA[1] is that an electronic subscription doesn't provide a right to preserve the content