Toward an Economic Model of Long-Term Storage

Daniel C. Rosenthal
UC Santa Cruz

David S. H. Rosenthal
Stanford University Libraries

Ethan L. Miller
UC Santa Cruz

Ian F. Adams
UC Santa Cruz

Mark W. Storer
NetApp

Erez Zadok
Stony Brook University

Motivation

Society's memory used to be stored on paper, which survives benign neglect well. It is now stored on digital media, which are less tolerant. A Blue Ribbon panel [2] has identified the economic sustainability of digital archives as a major problem. Economic threats are likely to be highly correlated among different archives, limiting the effectiveness of replicating data across archives. The lifetime of preserved data is much longer than the service life of the media storing it, thus cost/reliability trade-offs are not made once, but *continuously through the life of the data*. The lifetime cost and reliability are *path-dependent*; predicting them requires Monte Carlo simulation. We are in the early stages of developing a simulation model which we hope will illuminate the impact of real-world situations on preservation economics. Such situations include sudden spikes in media cost (e.g., from floods in Thailand), possible future slowing of the exponential drop in media cost [6], and the competition between storage with high capital cost but low running cost (flash), lower capital cost but higher running cost (disk), and zero capital cost but higher running cost (cloud).

Previous work in this area has either treated cost/reliability trade-offs as a one-time decision (e.g., [9, 1]), or has extrapolated historical cost data (e.g., [7]). We present example results from two of our early models showing that they exhibit plausible behavior as they model investment decisions through time.

Models and Examples

Our first model follows a unit of hardware in a data center which stores data growing exponentially. Disks are added to match the growth; their capacities grow over time according to Kryder's law. They consume power and labor, and are replaced as they fail or end their service life.

Figure 1 shows an example analysis. Parameters are set to plausible values, such as 57% annual growth in data [4] and 5% probability of failure in service (estimated from [8]). The graph shows how total cost of ownership and its components vary with the service life of the disks. It demonstrates the well-known observation that disks (and tapes) are replaced when their density becomes too low to justify the space and power they use, not when their life expires. With our chosen parameters, the model predicts optimum replacement in under 3 years.

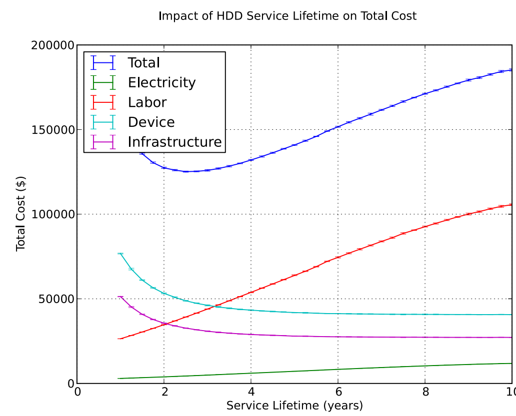

Figure 1: The effect of hard drive service lifetime on archive TCO at the end of a ten-year period.

This model's definition of total cost of ownership is the sum of expenditures; it ignores the time value of money. Over short periods this is justifiable, over decades it is not. Economists' standard technique for comparing expenditures through time is *Discounted Cash Flow* (DCF), which assumes a constant interest rate. Recent work shows that DCF is flawed both in practice [5] and in theory [3], and that Monte Carlo simulations are necessary.

Comparable accounting for costs through time is the focus of our second model, which follows the life of a unit of data as it migrates between storage media. It com-

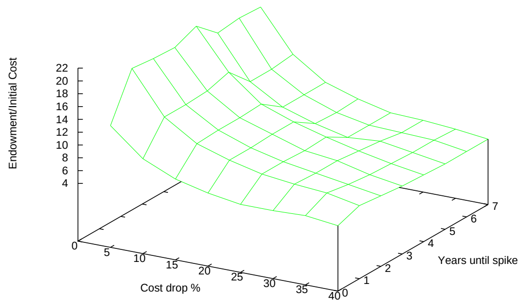

Figure 2: The impact of spikes in media cost on the endowment required for 95% survival at various rates of media cost decrease. Zero delay has no spike for comparison. Note the impact of a spike at the 4-yr media life.

putes the *endowment* needed to preserve the data, that is the sum of money deposited with the data at the beginning of time that, invested at the prevailing interest rate as it varies through time, will suffice to pay for the expenses the data incurs.

This model includes storage media, with purchase and running costs. They are replaced with successor media when their service life expires, or when new media become available whose costs are lower enough to justify the cost of migrating out of the old medium into the newer one. The endowment earns interest, and pays for the purchase, running and migration costs.

Figure 2 shows an example analysis. Interest rates are modeled on the past 20 years, media costs drop exponentially at various rates, and the service life of the media is 4 years. After a variable delay media costs double for a year then resume their exponential decrease. The graph shows the endowment that provides 95% probability of surviving 100 years without running out of money, as a multiple of the initial cost. Plausibly, if storage costs drop rapidly spikes have little effect but if they drop slowly the effect is large. Also, if costs drop slowly enough that media are replaced at their service life, and the spike happens at that time, the effect is amplified.

Future Work

Our plans for future work include developing similar simple models, attempting to calibrate them against the available historical data, and integrating them into a single comprehensive model capable of "what-if" explorations of alternative scenarios.

Notes

Daniel Rosenthal will present this work as a poster and a WIP. He and Ian Adams are students. David Rosenthal is grateful for support from the Library of Congress. We would also like to thank our colleagues in the Storage Systems Research Center (SSRC) at UC Santa Cruz, as well as the industry sponsors of the SSRC.

References

[1] ADAMS, I., MILLER, E. L., AND ROSENTHAL, D. S. H. Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes. Tech. Rep. UCSC-SSRC-11-07, University of California, Santa Cruz, October 2011.

[2] ANON. Blue Ribbon Task Force on Sustainable Digital Preservation and Access. `http://brtf.sdsc.edu/`, 2008.

[3] FARMER, J. D., AND GEANAKOPLOS, J. Hyperbolic Discounting is Rational: Valuing the Far Future With Uncertain Discount Rates. Tech. Rep. 1719, Cowles Foundation, Yale University, Aug 2009.

[4] GANTZ, J. F., REINSEL, D., CHUTE, C., SCHLICHTING, W., MCARTHUR, J., MINTON, S., XHENETI, I., TONCHEVA, A., AND MANFREDIZ, A. The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010. White Paper, IDC, March 2007.

[5] HALDANE, A. G., AND DAVIES, R. The Short Long. In *29th Societe Universitaire Europeene de Recherches Financieres Colloquium: New Paradigms in Money and Finance?* (May 2011).

[6] KRYDER, M., AND KIM, C. S. After hard drives - what comes next? *IEEE Transactions on Magnetics* (2009).

[7] MOORE, R. L., D'AOUST, J., MCDONALD, R. H., AND MINOR, D. Disk and Tape Storage Cost Models. In *Archiving* (2007).

[8] PINHEIRO, E., WEBER, W.-D., AND BARROSO, L. A. Failure trends in a large disk drive population. In *Proceedings of the 5th USENIX Conference on File and Storage Technologies* (2007).

[9] STRUNK, J. D., THERESKA, E., FALOUTSOS, C., AND GANGER, G. R. Using utility to provision storage systems. In *Proceedings of the 6th USENIX Conference on File and Storage Technologies* (2008).