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

A significant obstacle facing archival storage is how to finance the storage of digital data over the long term. Little is understood about the economic implications of various trade-offs involved in designing, implementing, and managing a digital archive.

Goals

- Understand how various trade-offs affect storage system longevity
- Design a cost-accurate model of an archive
- Implement a simulator which reflects this model
- Estimate the probability of not running out of money with a given cash flow

Economic Model Overview

- Monte Carlo simulations used to model storage system evolution over time
- Simulations account for initial and operating storage system costs
- Simulations model single-replica storage with no redundancy (assumes failed data can be recovered from off-site)
- Tunable simulator parameters for:
  - Various components of initial costs (storage devices, infrastructure)
  - Various components of operating costs (electricity, labor)
  - Data size and storage density growth rates
  - Device power draw, service lifetime, and failure probability
  - Storage system utilization

Results

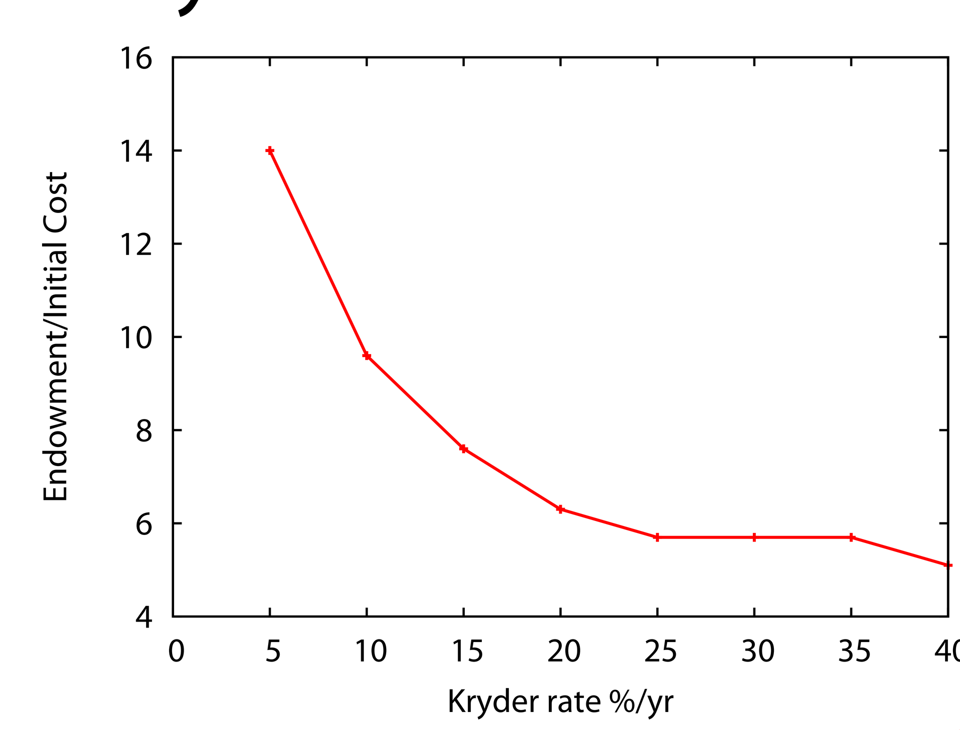
- Simulations use an endowment model of paying for storage
- Infrastructure, labor, and electricity costs are modeled as perfectly scalable
- For models which consider time value of money, interest rates are modeled using the past 20 years

Simulator Parameters	
General	
Simulation Length	10 years
Time Step	30 days
Data	
Initial Data	100 TB
Growth Rate	57% annual growth
Infrastructure	
Space Cost	\$800 / m^3
Power Capacity Cost	\$20,000 / kW
Electricity Cost	12.78 cents / kWh
Labor Cost	\$50 / hr
Device	
Storage Medium	Hard Drive
Capacity Growth Rate	Kryder's Law*
Power Draw	5 W
Service Lifetime	7 years
Failure Probability	0.05
Management Time Cost	1 hr / year
Purchase Price	\$150

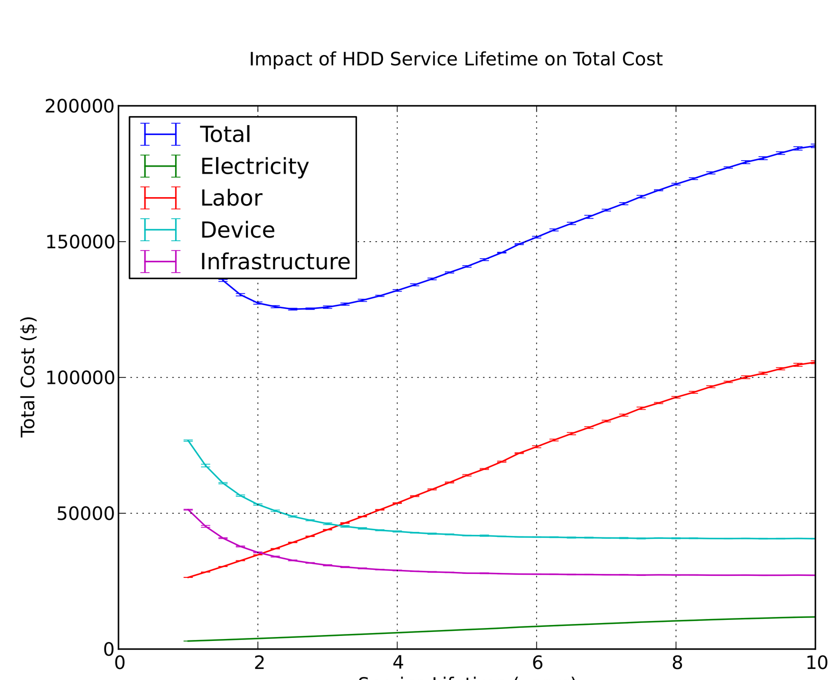
* Kryder's Law: Annual doubling in device capacity

Effect of Storage Density Growth Rate

- Storage density growth rate has a non-linear influence on data storage costs
- Graph shows endowment-to-initial-cost ratio required for 98% chance of not running out of money for 100 years

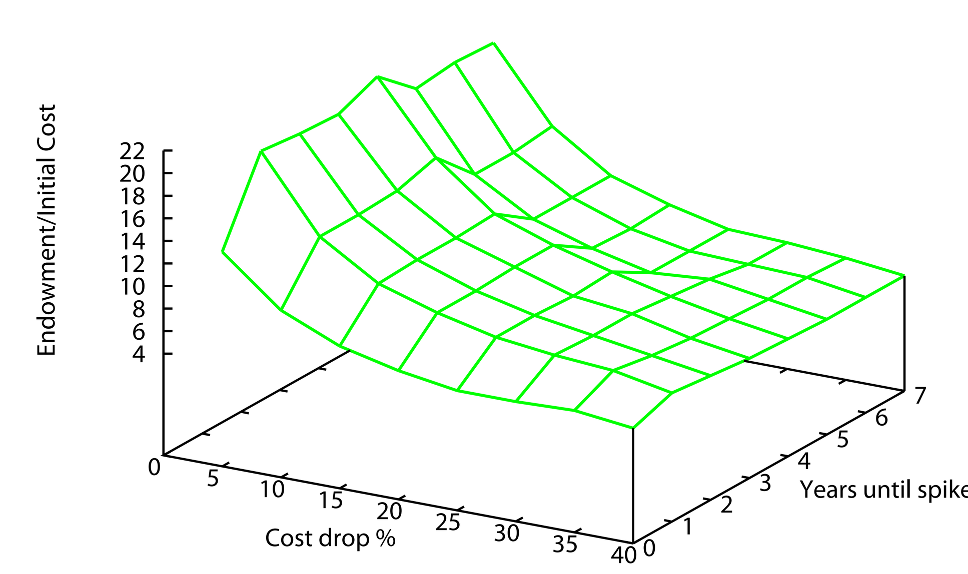


Effect of HDD Service Lifetime on TCO



- Graph shows TCO breakdown as a function of disk service lifetime
- Devices have a constant failure probability during their service lifetime.
- ``Leave in-service until failure'' is a suboptimal device replacement policy

Modeling Real-World Events: Thailand Floods



- Floods modeled by a temporary spike in storage media cost
- Graph shows endowment required for 95% chance of not running out of money for 100 years

Ongoing and Future Work

- Study the impact of disruptive technologies on archival storage
- Study trade-offs between endowment size, protection level, and survivability
- Compare various storage media (disk, flash, cloud, etc.) for suitability in archival storage
- Experiment with various data and media capacity growth rates
- Examine the impact of financial events on archive survivability
- Test various forecasting strategies
- Explore ``what if?'' scenarios

Sponsors

SSRC Storage Systems Research Center; University of California Santa Cruz; LOCKSS; Stony Brook State University of New York; NSF; CRSS NSF Center for Research in Intelligent Storage; CITRIS Center for Information Technology Research in the Interest of Society