Pergamum: Replacing Tape with Energy Efficient, Reliable Disk-Based Archival Storage

Mark W. Storer, Kevin M. Greenan, Ethan L. Miller, Kaladhar Voruganti
Storage System Research Center, University of California, Santa Cruz; *Network Appliance

The Need for Energy-Efficient, Archival Storage

With both private and personal data, there is a need for cost-effective, archival storage solutions. In the businesses arena, data protection is now considered a business strategy. For individuals, personal histories, photos, movies and personal documents are now being backed up digitally. In both scenarios, it is critical to identify and quantify how much electricity will be useful at a future date. The effectiveness of any archival solution is dramatically improved by providing a cost-effective, energy-efficient way that allows the preservation of anything that might prove useful one day.

Many commonly used power supplies operate at only 65-75% efficiency, representing one of the primary culprits of excess heat production and contributing to cooling demands that account for up to 60% of data-center energy usage.

Pergamum, a distributed network of intelligent, disk-based, storage appliances that stores data reliably and energy-efficiently. While existing MDS (Massive Arrays of Idle Disks) systems keep disks idle to save energy, Pergamum adds NVRAM at each node to store data signatures, metadata, and other small items, allowing deferred writes, metadata requests and interdisk data verification to be performed while the disk is powered off. Pergamum uses both intra-disk and inter-disk redundancy to guard against data loss, relying on hash tree-like structures of algebraic signatures to efficiently verify the correctness of stored data.

Reliability in Pergamum

Intra-Tome Redundancy

Pergamum uses hash trees built from algebraic signatures. Using signatures of blocks as a shows that signatures of sets of signatures follow the same relationships as the underlying data; this property is maintained all the way up to the root of the tree. Thus, the signature at the root of each disk’s hash tree for the region should yield a valid erasure code word when combined together. If they do not, some block in the redundancy group is invalid, and the disks recurse down the hash tree to find the bad block, exchanging the contents at each level to narrow the location of the “bad” block.

Integrity Checking using Algebraic Signatures

Historical Background

Greek library built by Eumenes II (rule: 197–158 BC)

Credited with the transition from papyrus to parchment

Tome and inter-disk redundancy

Reliability Settings

Redundancy Group Particle Region

What has changed to enable Pergamum to replace tape with disks?

- Hard drives have dropped in price relative to tape
- Availability of high-performance, low-power CPUs consuming 2-3W when active (300 mW inactive)
- Increased speed of processors consumed 30X. W.
- Availability of inexpensive, high-speed networks
- Work on MDSs has demonstrated that considerable energy-based cost savings can be realized while maintaining high levels of performance.