Computer scientists develop solutions for long-term storage of digital data

April 21 2008

Although the digital age is well under way, one crucial detail remains to be worked out--how to store vast amounts of digital information in a way that allows future generations to recover it.

"The problem is how to build a large-scale data storage system to last 50 to 100 years," said Ethan Miller, associate professor of computer science in the Baskin School of Engineering at the University of California, Santa Cruz.

Tape libraries are widely used for data storage, but digital tape has many shortcomings as an archival medium. Miller's group has come up with a new approach, called Pergamum, which uses hard disk drives to provide energy-efficient, cost-effective storage. The declining cost of hard drives has made them more competitive with tape, and they offer numerous advantages for searching and retreiving data. "It's like the difference between a VCR and TiVo," Miller said.

Pergamum, named after the ancient Greek library that made the transition from fragile papyrus to more durable parchment, is a distributed network of intelligent, disk-based storage devices. The team that developed it includes UCSC graduate students Mark Storer and Kevin Greenan, along with researcher Kaladhar Voruganti of NetApp (formerly Network Appliance), a company that focuses on storage and data management solutions.

Archival storage is a big issue for businesses, partly due to legal

requirements for the preservation of financial and business records, and also because data mining strategies can turn stored data into a valuable resource. Long-term storage is also a growing issue for individuals who are filling their personal computers with digital photos, movies, and documents.

"There is a risk that an entire generation's cultural history could be lost if people aren't able to retrieve that data," Storer said. "Everyone is switching to digital cameras, but we've never demonstrated that digital data can be reliably preserved for a long time."

Pergamum has attracted a lot of attention from industry since Storer presented it at a leading conference in the field, the USENIX Conference on File and Storage Technologies (FAST '08), held in San Jose in February. Robin Harris, an industry consultant who writes an influential blog called StorageMojo, called the Pergamum paper his "favorite FAST '08 paper".

The researchers designed the system to provide reliable, energy-efficient data storage using off-the-shelf components. It also has the ability to evolve over time as storage technologies change. "You want to avoid 'forklift upgrades,' where you have to get rid of the old system and transfer all your data to a whole new system," Miller said.

According to Storer, businesses are beginning to recognize that archival storage is very different from simply backing up their data. "A backup is a safety net--you hope you won't need it. Archival data you do want to use--it's a valuable resource and you want to be able to mine it for information," he said.

Tapes work well for backups, in which data are written once, rarely read, and not kept indefinitely. But archival data should be easy to read, query, browse, and search, and tape has inherent weaknesses in these areas.

Existing disk-based systems offer excellent performance, but rely on power-hungry central controllers.

"Energy usage is a big issue, so a lot of our effort in designing Pergamum focused on dramatically reducing power use," Miller said.

Pergamum uses individual building blocks consisting of a hard drive; a small, low-power processor (like the chip in an iPhone); a flash memory card; and an ethernet port. These units, called "tomes," are connected using relatively inexpensive ethernet switches.

"Each tome is like a minicomputer, but with very low power demands," Miller said. "When not in use, it can shut down almost completely."

Even when active, the devices use very little power (less than 13 watts), which can be delivered over the network using Power over Ethernet technology. As a result, each unit is essentially a self-contained box with a network connection. The flash memory provides low-power, persistent storage so that many operations can be performed without activating the hard drive.

For reliability, Pergamum uses two levels of redundancy--within and between disks--to protect from both disk failures and errors in writing data to a disk (so-called "latent sector errors"). Tomes can be easily added to expand the system or to replace failed disks. And if hard disk drives become obsolete in 10 years, Pergamum won't suffer the same fate. The system doesn't care what the actual storage medium is, as long as the device can implement the simple protocol that will allow it to function as part of the network.

"In 50 years, the devices might use holographic storage," Storer said. "As long as you can wrap the new storage medium in this intelligent layer that speaks the protocol, it can participate in the network."

Pergamum is one of several related projects being developed by researchers in the Storage Systems Research Center (SSRC) at UCSC's Baskin School of Engineering. The center's other archival storage projects include Deep Store, which dramatically reduces the amount of space required to store data, and POTSHARDS, which provides long-term secure storage using "secret splitting" instead of traditional encryption. Both of these projects would be compatible with Pergamum, Miller said.

Source: University of California - Santa Cruz

Citation: Computer scientists develop solutions for long-term storage of digital data (2008, April 21) retrieved 2 May 2024 from https://phys.org/news/2008-04-scientists-solutions-long-term-storage-digital.html