DIGITAL TECHNOLOGY CENTER • A center of the Institute of Technology

UNIVERSITY OF MINNESOTA

DISC

MEMBER RESEARCH UPDATE

Winter 2007

DISC CALENDAR

May 15-16 | ISW5: Fifth Annual Intelligent Storage Workshop
Leading experts from academia and industry provide detailed information on the latest topics in intelligent storage technology

May 17 | Member's Day
DISC members and sponsors will gather for a full day of storage technology discussions and project reviews

For more information, visit www.dtc.umn.edu/disc/Events

Object-Based Storage (OSD) Applications

DISC explores a unified framework for storing and querying unstructured data

The Mayo Clinic weekly generates about three million images, or what amounts to a terabyte of new data every nine days. The exploding volume of computerized information poses storage and access challenges. One DISC research initiative involves improving access to the growing volumes of patient information at Mayo Clinic.

Professors David Du and Mohamed Mokbel, Aravindan Raghuveer, Ph.D. student, and graduate student contributors Biplob Debnath and Meera Jindal are collaborating on developing a unified framework for storing and querying unstructured and structured data (SQUAD).

In addition to the volume issue, Mayo Clinic also faces a heterogeneity problem. Most of the information that scientific experiments produce is a mix of structured and unstructured data. Such data includes regular patient registration and demographic information, as well as x-rays, laboratory tests, genetic profiles, and scanned images. In addition, Mayo Clinic maintains 18 million clinical notes that are annotated and searched, and revises 5,000 notes every day.

The SQUAD research project attempts to unify the access to all types of data and improve efficiency and productivity in medical treatment. Figure 1 shows the architecture under development to address these issues. It uniquely blends traditional database technology and Object-Based Storage Devices (OSD) Intelligent Storage Nodes (ISN) through a Meta-Data Server (MDS).

The SQUAD system stores data where it best can be managed. ISNs manage unstructured data, while a database stores structured data, exploiting the rich semantics and relationships between structured and unstructured data to guide data storage and retrieval. The underlying OSD infrastructure allows SQUAD to support strong scalability.

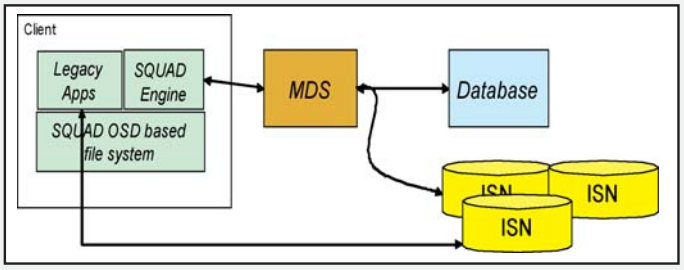


Fig. 1 **Storing and Querying Unstructured Data (SQUAD) Architecture**

JOIN US

DISC membership brings with it many advantages, including the opportunity to help shape and participate in cutting-edge research that improves the capabilities of storage devices. In addition, consortium members gain access to a pool of the most talented Ph.D. students, and many of those students are making excellent contributions as interns at companies such as Seagate, IBM, McData, Sun, and LSI Logic.

Current DISC members include:
LSI Logic
Sun Microsystems
Symantec
ITRI

DISC contributors include:
Los Alamos National Labs
Seagate
IBM
Brocade (McData)

Please contact Cory Devor at devor@dtc.umn.edu for more information about DISC membership.

This project also includes evaluation of the design, efficiency, and effectiveness of hierarchical indexes, possibly through a coarse index at the MDS and finer indexes at the ISNs. In addition, significant work will be directed at evaluating Layout Aware Exhaustive Search techniques. Since ISNs have the layout information of the storage device, DISC researchers propose to use a layout-aware scheduling scheme to perform exhaustive search for "unexpected queries"—queries that address non-indexed data or the index data itself.

Improving Operational Efficiency of Storage Systems

DISC researchers study the use of Massive Array of Idle Disks (MAID) to reduce energy consumption

Performance requirements of integrated queries, such as those in SQUAD, suggest an online storage solution, particularly because many queries may require access to older "archived" data. Tape-based archival solutions suffer from high latency and low throughput. Using large disk arrays offers an attractive option for large distributed sites such as the Mayo complex.

In this research initiative, Professor David Du and graduate student Pramod Mandagere are looking at a number of issues that impact the use of a Massive Array of Idle Disks (MAID) to reduce energy costs significantly while improving archive query response time.

The decreasing cost and increasing capacity of disk drives is rapidly changing the economics of online storage. In addition to their other advantages, large disk arrays will enable system scaling. With growth in clinical and research medical data predicted both in at-rest storage (TBs or more) and in delivered data (GBs/day or more), system scaling is an important property.

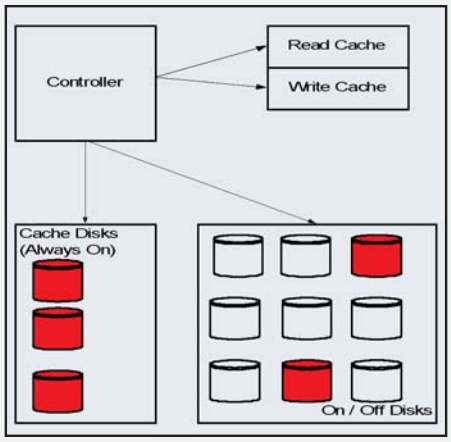


Fig. 2 **MAID Storage Configuration with Idle Disks (1 to n disk controllers)**

This enhanced performance for data queries comes at a price. Keeping huge disk arrays "spinning" has a hidden cost—energy. Industry surveys suggest that the cost of powering up the nation's data centers is growing at the rate of 25 percent every year.

Among various components of a data center, storage consumes 27 percent of total energy. To make matters worse, the demands of increasing performance have led to disks with higher power requirements, while at the same time, storage demands are continuously growing by 60-100 percent annually according to an industry reports. Given the well-known growth in healthcare costs, an optimal solution involves keeping data online yet mitigating the high cost of power.

Various studies of data access patterns in data centers suggest that on any given day the total amount of accessed data is less than 5 percent. While it's difficult to predict the future access rates for medical data once fully online, it's likely to exhibit similar behavior. Most energy conservation techniques make use of various optimizations to conserve energy, but these techniques usually come with a performance penalty and also may fail because of the random nature of the data access pattern. Access patterns to medical data may have more predictability, and researchers want to exploit hidden patterns and well-known patterns inherent in data access and query history to significantly reduce energy consumption.

In this initiative, DISC researchers plan to explore solutions using MAID system architecture *(see Figure 2)*. Issues under consideration will include the following:

- Ordering of requests may help maximize the number of serviced requests.
- Relying on disk scheduling algorithm may not lead to the optimal performance.
- When to invoke the decision making algorithm?
- What is the best way of handling cases where the workload is skewed and requires disks that are "close by" to be "turned on" concurrently?
- One answer to the above question may involve making data/LUN binding decisions.

File Systems

File System in User-Space Portability Study (FUSE) offers potential; portability of FUSE to Open Solaris environment investigated

File systems are one of the most complex pieces of software residing within the kernel of an operating system. Designing and implementing a file system within the kernel of an operating system is extremely hard and requires extensive knowledge of the kernel data structures. The file systems need to be ported to every new version of the kernel, which again requires significant

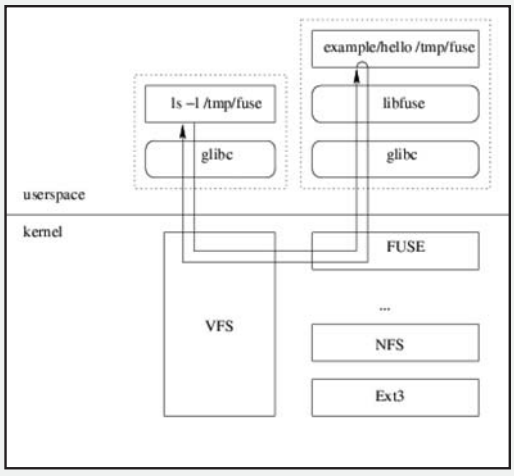


Fig. 3 **FUSE Architecture**

amount of effort and resources during every kernel release cycle, and kernel reboot times can hog a major portion of time spent on debugging and developing the file systems.

These reasons help make user level file systems highly desirable. The entire file system resides in user space, avoiding frequent kernel reboots during development and debugging. Very good debuggers exist for user space programs compared to limited alternatives for kernel code. The file system can remain stagnant across multiple versions of the kernel.

The trend of designing file systems that can be tuned to optimally store certain types of data also is supporting the increased popularity of user-space file systems.

While there are multiple methods to design and implement user-space file systems, several of the most popular ones include FUSE, NFS, Coda, and Stackable File systems. FUSE *(see Figure 3)* has proven the most popular choice to develop user level file systems; more than 70 file systems have been registered in SourceForge.

FUSE is a popular choice for a number of reasons: It was designed from scratch to help easily create user level file systems, unlike NFS and Coda, and, as a result, has a very simple user-space library interface that makes the file system writing task less arduous. With the thin layer in the kernel, indirection does not cause too much overhead. Unlike NFS, the communication is not through the network (loopback interface) and is highly optimized for local file systems. Also FUSE does not require the file system to be run in super-user mode.

DISC is investigating the portability of FUSE and other "user space" file system issues in preparation for more extensive file systems work to support OSD applications. The port of FUSE to Solaris will be done from the free BSD version of FUSE due to licensing issues. This will be completed in the following phases:

- Phase 1 will be to port existing free BSD implementation of FUSE to Solaris.
- Phase 2 will address outstanding gaps in FUSE to make it fully functional on Solaris.
- Phase 3 will involve experimentation and evaluation.

DISC

FACULTY PROFILE

Ahmed Tewfik explores new frontiers as part of DISC group

The E.F. Johnson Professor of Electronic Communications in the Department of Electrical and Computer Engineering, Ahmed Tewfik received his master's degree and Sc.D. from the Massachusetts Institute of Technology. A fellow of the IEEE, Tewfik has consulted with companies such as MTS Systems and Rosemount, worked with Texas Instruments and Computer Devices International, and served as president and CEO of a technology company that he founded.

As a DTC faculty member, Tewfik collaborates with faculty and students as part of several DTC Intelligent Storage Consortium (DISC) initiatives.

The Start: In 2002, Tewfik and colleague David Lilja were looking into the use of wireless communication in computing and storage. They connected with DISC Director David Du and other DTC faculty. "We started meeting as a group and talking about directions and ideas for collaborative work," says Tewfik.

Expertise: Tewfik works in a large number of areas, including bioinformatics, sensors, programmable wireless networks, ultrawideband communications for wireless personal area and storage networks, and smart wireless devices.

Emerging Issues: Increasing volumes of data in industries such as health care pose the practical challenges of storage and easy access. DISC research and the development of the Object-Based Storage Device (OSD) are paving the way for applications that seek to put the data to good use in a number of ways.

In his collaborations with DISC colleagues, Tewfik is researching ways to store data that is collected at different points and to track updates to that data from different sources, offering one place for users to review all relevant data and access the latest changes in the information.

Increasing the storage device efficiency is another key to applications, and DISC researchers are studying the massive array of idle disks (MAID) as a way to maximize the capabilities of storage devices. While storing significant volumes of data and information requires significant amounts of energy, MAID seeks to distribute energy to better match performance requirements.

"Say that you only need to access data on 5 percent of the storage device," says Tewfik. "It becomes inefficient to run huge disks, and we want to find ways for the device to distribute the data for minimal power expenditure."

In addition, the development of large networks of wireless sensors is a research area with rich applications. In health care, for example, sensors could gather and transmit data from patients to networks, with storage systems synthesizing the data, adding it to existing data, and making it available to physicians.

"The OSD platform seems to be a nice fit to that kind of synchronization," says Tewfik.

Future: "We are really interested in ways to efficiently store vast amounts of data, add to that data, and make it easily accessible," says Tewfik. "OSD lays the infrastructure for different types of applications and allows for connections to compatible wireless applications."

Benefits of Collaboration: Working with other researchers and industry in DISC collaborations has helped Tewfik apply his research expertise in new ways. "As a result of my involvement, I've been identified number of interesting and challenging problems that are at the intersection of multiple disciplines." The collaborations with industry help shape the direction of the research, he says. "We are better able to understand their needs."

DISC HIGHLIGHTS

DISC receives National Science Foundation grant

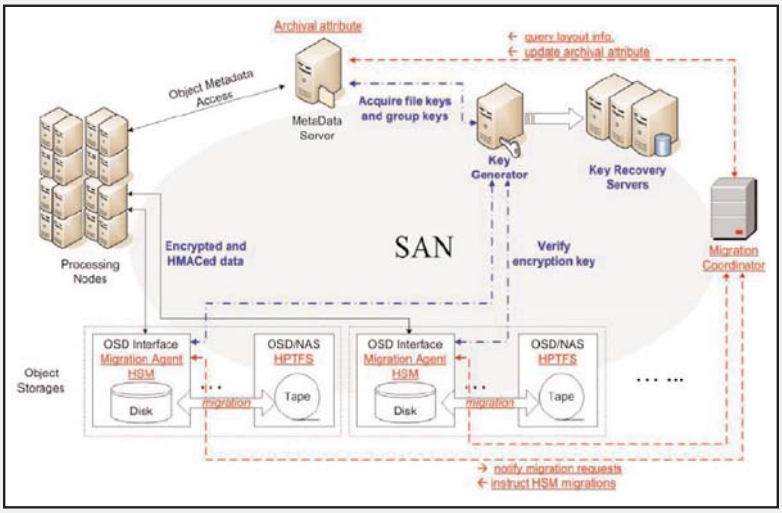
The National Science Foundation awarded DISC a three-year, $600,000 grant to study critical issues in high-performance computing and data security. DTC faculty Yongdae Kim and David Lilja and Andrew Odlyzko, DTC director, will serve as principal investigators for the project, which will explore an integrated infrastructure for secure and efficient long-term data management.

The project addresses the need for storage and data archiving technologies that can meet the increasing demands for bandwidth and long-term data security.

Archive systems have been identified as a key bottleneck in high-performance storage systems. According to estimates, typical HPC applications in 2003 required more than 35 TB/hr (10GB/sec) of archive bandwidth, while in 2007 required bandwidth is expected to grow to more than 350 TB/hr.

Protecting and preserving cryptographic keys in a long-term archive also presents a major challenge since many unforeseen changes can occur during the data's lifetime. For example, the user who originally encrypted the data may be unknown or unavailable when the data is to be decrypted. Keys and the cryptographic algorithms also can be compromised. All of these issues require fundamentally new approaches to develop a storage archive that is capable of securely preserving data for 30 years or more.

In this project, DTC researchers propose to develop a tape-based long-term archival data management system that will enable high-performance while maintaining appropriate levels of security throughout the data's lifecycle.



New article from DISC researchers published

"On Providing Reliability Guarantees in Live Video Streaming with Collaborative Clients," an article by DISC Director David Du and graduate students Aravindan Raghuveer and Yingfei Dong, will appear in *Proceedings of the Fourteenth Annual Multimedia Computing and Networking Conference* [MMCN] 2007.

DISC hosts OSD collaboration meeting

On Oct. 27, 2006, DISC hosted a working group meeting that included IBM, Sun, Symantec, LSI, Xyratex, and Seagate, as well as DISC faculty and students. The group explored current and future collaborations to advance research and development projects that support selected OSD applications. Presentations from the meeting can be founda t .

DISC

STUDENT PROFILE

Vishal Kher

Ph.D. student
Computer Science and Engineering

GRADUATION: Expected 2007
EDUCATION: BE, University of Pune, India; MS, University of California, Santa Barbara
DISC RESEARCH ADVISOR: Yongdae Kim and David Du

What are your current DISC projects?
I am working on several DISC projects: OSD reference implementation, secure cross-domain file sharing, secure accounting, long-term key management, and SIMON.

Have you collaborated with any DISC members?
I have collaborated with Sojeong Hong, Aravindan Raghuveer, Jaehoon Jeong, Dingshan He, and Yingping Lu. I am also collaborating with LSI logic and Symantec on OSD reference implementation.

Do you have a specific career path in mind?
I enjoy both the technical as well as the management tracks. I plan to spend several years in the technical field developing secure systems. After gathering enough technical experience, I would like to lead and manage projects and explore the management track.

Where would you like to work? Doing what?
I enjoy research and development in the area of system security and distributed file systems. I believe that prudent application of cryptography is important for building practical, usable, and manageable secure systems. I am interested in working with companies that have challenging projects and motivated people in these areas, especially, in companies that are interested in building secure and distributed storage systems.

What has been the most interesting thing about working with DISC?
I had the opportunity to work on interesting research problems that were not only of academic interests but also had practical applications. DISC gave me the opportunity to understand practical research challenges and design new secure storage systems. Workshops, seminars, and the wide variety of projects at DISC helped me to broaden my knowledge in the area of storage networks.

For more information on DISC membership and DISC projects visit our website at: www.dtc.umn.edu/disc/ or contact Cory Devor at devor@dtc.umn.edu or 612-625-1716.

The University of Minnesota is an equal opportunity educator and employer.