PARALLEL DATA LABORATORY
CARNEGIE MELLON UNIVERSITY

THE
PDL Packet

THE NEWSLETTER ON PARALLEL DATA SYSTEMS • FALL 1999
http://www.pdl.cs.cmu.edu

AN INFORMAL PUBLICATION FROM A UNIVERSITY RESEARCH COMMUNITY DEVOTED TO ADVANCING THE STATE OF THE ART IN STORAGE SYSTEMS AND TO EFFICIENTLY INTEGRATING STORAGE INTO PARALLEL AND DISTRIBUTED FILE SYSTEMS, HIGH BANDWIDTH NETWORKS AND COMPUTER CLUSTERS.

I ◆ N ◆ S ◆ I ◆ D ◆ E

CONSORTIUM MEMBERS

CLARiiON Array Development
EMC Corporation
Hewlett-Packard Laboratories
Hitachi, Limited
IBM Corporation
Infineon Technologies
(formerly Siemens Microelectronics)
Intel Corporation
LSI Logic
MTI Technology Corporation
Novell Corporation
ProCom Technology
Quantum Corporation
Seagate Technology
Wind River Corporation
3Com Corporation

NASD Prototype Released

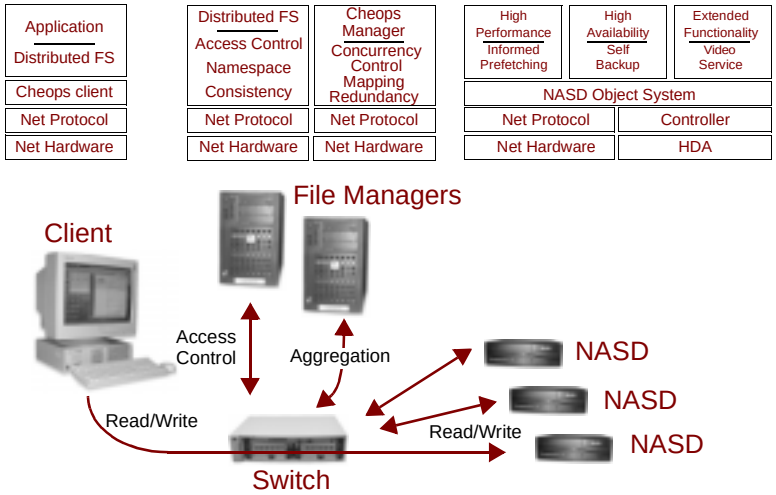
http://www.pdl.cs.cmu.edu/extreme/

David Nagle & Joan Digney

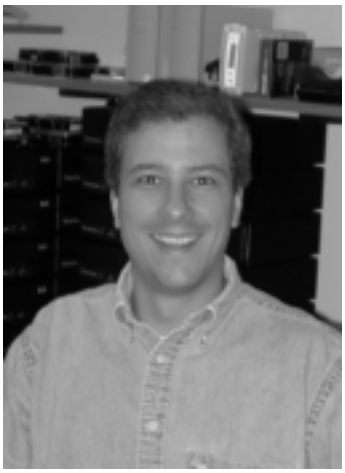
The Parallel Data Lab announced in July the public release of their next generation scalable storage system software: Network Attached Secure Disks (NASD). NASD answers today's requirements for high bandwidth, low latency accessible storage and is available for download from http://www.pdl.cs.cmu.edu/extreme/distrib.html. NASD implements a scalable, distributed parallel storage architecture, interface and protocol to comprehensively reduce access latency. These revolutionary changes enable commodity storage components to be the building blocks of high bandwidth, low-latency, scalable storage systems. (See figure below).

The first public release of the NASD code contains NASD drive object management code, user-level client interfaces with NFS-like semantics, a user-level aggregation library, and an NFS-like file manager. The NASD drive software contained in this release operates at user-level in Linux (v2.2) and at user-level and the kernel-level in Digital UNIX (v3.2g),using SRPC (a TCP-based remote procedure call package) for communication. The security system in this code protects integrity using keyed message digests (HMAC-SHA1)

... continued on pg. 11



1999: The Year of Network-Attached Storage



From the Director's Chair
DAVID NAGLE

Greetings from sunny Pittsburgh and the Parallel Data Lab. The last 12 months have been very productive for the lab. Two dissertations have been completed: Howard Gobioff's "Security for a High Performance Commodity Storage Subsystem," and Erik Riedel's "Active Disks." Six new papers have been accepted at Hot Interconnects99, Extreme Linux, USENIX99, FREENIX99, IEEE Computer, and the best student paper at OSDI. The students, staff and faculty have received numerous awards including: John Griffin (NFS Graduate Research Fellowship), Fay Chang (Intel Fellowship), Karen Lindenfelser (CIT Staff Award), Paul Mazaitis (SCS Staff Recognition Award), Todd Mowry (Sloan Research Fellow), Seth Goldstein (NFS CAREER Award), and Garth Gibson (IEEE Reynold B. Johnson Information Storage Award). We have several new members of PDL including Mor Harchol-Balter, who joins the SCS faculty after completing her NSF Postdoctoral Fellowship at MIT's Laboratory for Computer Science. Also, the PDL's computing infrastructure continues to grow. We've added over 40 PPRO I I I's from Intel, a 32-port Myrinet Gigabit network, and are currently receiving a 100-node network-attached storage testbed from Quantum, and an 8-node Netfinity cluster with FibreChannel and Gigabit Ethernet from IBM.

Turning to research, 1999 has been the year of Network-attached Storage. Our Network-attached Secure Disk (NASD) project focuses on creating scalable network-attached storage (NAS) interfaces and architectures with research in the areas of object-based drive interfaces, security models for NAS, distributed file systems for NAS, networking support for NAS, and drive-based optimizations that exploit the drive's object-interface architecture.

One of NASD's major challenges is to evolve the disk interface from a block to an object model. Using our first interface design as a starting point [see Sigmetrics98], CMU has worked closely with the National Storage Industry Consortium's network-attached storage devices working group to develop an object-based interface standard. I am pleased to announce that the working group has delivered an Object-based Storage Device (OBSD) proposal to the ANSI X3 T10 standards committee (responsible for SCSI). The OBSD proposal will be presented to the T10 working group meeting, Nov. 5-9, 1999 in Monterey, CA. You can obtain a copy of the OBSD proposal at: ftp://ftp.t10.org/t10/document.99/99-315r0.pdf.

We are also sharing our NASD software-base with other researchers and industry through our recently released Linux and Digital UNIX ports of the NASD software. Announced at the 1999 Extreme Linux conference and described in the article on page 1, this experimental platform demonstrates the

… continued on pg. 3

THE PDL PACKET

The Parallel Data Laboratory
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3891
VOICE 412•268•6716
FAX 412•268•3010
FAX ALTERNATE 412•268•5576

PUBLISHER
David Nagle

EDITOR
Joan Digney

The *PDL Packet* is published approximately once per year and mailed to members of the Parallel Data Consortium. Copies are given to other researchers in industry and academia as well. Postscript and pdf versions will reside in the Publications section of the PDL Web pages. Contributions are welcome.

PDL LOGO

Skibo Castle and the lands that comprise its estate are located in the Kyle of Sutherland in the northeastern part of Scotland. Both 'Skibo' and 'Sutherland' are names whose roots are from Old Norse, the language spoken by the Vikings who began washing ashore regularly in the late ninth century. The word '*Skibo*' fascinates etymologists, who are unable to agree on its original meaning. All agree that '*bo*' is the Old Norse for '*land*' or '*place*.' But they argue whether '*ski*' means '*ships*' or '*peace*' or '*fairy hill*.'

Although the earliest version of Skibo seems to be lost in the mists of time, it was most likely some kind of fortified building erected by the Norsemen. The present-day castle was built by a bishop of the Roman Catholic Church. Andrew Carnegie, after making his fortune, bought it in 1898 to serve as his summer home. In 1980, his daughter, Margaret, donated Skibo to a trust that later sold the estate. It is presently being run as a luxury hotel.

MISSION STATEMENT

To advance the state of the art in storage systems and the integration of storage into parallel and distributed file systems, high bandwidth networks, and computer clusters.

CONTACTING US

WEB PAGES

PDL Home: http://www.pdl.cs.cmu.edu/

Please see our web pages at http://www.pdl.cs.cmu.edu/PEOPLE/ for further contact information.

FACULTY

David Nagle (director)
412•268•3898
david.nagle@ece.cmu.edu

Christos Faloutsos
christos.faloutsos@cs.cmu.edu

Greg Ganger
greg.ganger@ece.cmu.edu

Garth Gibson
garth.gibson@cs.cmu.edu

Seth Goldstein
seth.goldstein@cs.cmu.edu

Mor Harchol-Balter
mor.harchol-balter@cs.cmu.edu

Todd Mowry
todd.mowry@cs.cmu.edu

Hui Zhang
hui.zhang@cs.cmu.edu

STAFF MEMBERS

Patty Mackiewicz (business administrator)
412•268•6716
patty.mackiewicz@cs.cmu.edu

Joan Digney

Jennifer Landefeld

Nat Lanza

Karen Lindenfelser

Paul Mazaitis

David Rochberg

GRADUATE STUDENTS

Khalil Amiri
Angela Demke Brown
Jeff Butler
Fay Chang
Shemin Chen
David Friedman
Garth Goodson
John Griffin
Charles Hardin
Stavros Harizopoulos
Ed Hogan
Andrew Klosterman
Chris Lumb
David Petrou
Erik Riedel
Chris Sabol
Mike Scheinholtz
Jiri Schindler
Steve Schlosser
John Strunk
Craig Soules
Neal Tibrewala
Tom Wagner
Mengzhi Wang
Ted Wong
Shuheng Zhou

Year of Network-Attached Storage

… continued from pg. 2

value of NASD's interface by providing both device and file system prototype software. You can obtain the current release at: http://www.pdl.cs.cmu.edu/extreme

A very important area for high-performance storage applications is high-bandwidth, low-overhead networking. To understand how networking can best support storage, we have been evaluating NASD over the VI Architecture (http://www.viarch.org). This work includes a user-level RPC that enables direct communication between a user-level NASD client and a user-level drive module. Currently delivering over 500 Mb/sec with our older Myrinet hardware, we expect to see Gb/sec speeds in our the new Myrinet testbed. We've also begun a network comparison between Fibre Channel and Gigabit Ethernet to understand how effectively each supports NAS traffic.

PDL's Active Disk project has been busy extending the basic NASD architecture to enable application functionality that executes directly at the storage device. Leveraging the inherent parallelism available in large numbers of storage devices, our Active Disk prototype achieves significant performance improvements across a range of applications including data mining, searching, sorting, and edge detection. Recently, we have integrated Active Disks with the Postgres database system. Our modifications to Postgres's query planner allow Postgres to automatically use Active Disks by constructing a plan based on available disk resources (e.g., number of disks, available memory, processing rates) and data characteristics. When executed, the plan automatically migrates the appropriate part of the database query to the Active Disks, providing significant performance improvements (some queries obtaining a 10X speed-up).

Extending the notion of remote function execution, our ABACUS project optimizes overall system performance by dynamically migrating functionality between clients and storage devices. Based on black-box monitoring of run-time resource utilization, ABACUS's cost-benefit performance model adapts the placement of each function (i.e., ABACUS module) as system resource utilization and application demands change over time. For example, migrating the caching module between clients (when writes are rare) and storage devices (when writes are common), improves performance by up to 2.7X. Further, dynamic directory management improves performance by up to a factor of 8X on a heavily loaded local-area network. Of course, the PDL's award winning aggressive prefetching projects continues to probe more deeply into automatically preplanning I/O resource usage. For example, please see Fay Chang's thesis proposal abstract on page 12.

There are also several new projects within PDL. The Automated Disk Characterization Project explores characterization and tuning (e.g., via aggressive scheduling) of disk drive performance (see page 14 of the PDL Packet). Another new project, Self-securing Storage, exploits rising disk capacities and new object interfaces for novel mechanisms to protect stored data from accidental or malicious destruction.

Finally, the PDL is actively involved in CMU's Center for Highly Integrated Information Processing and Storage (CHIPS). The goal of CHIPS's core technology (IC-based nonvolatile rewriteable mass storage) is to enable low-cost embedded computers with multiple gigabytes of IC-based mass storage. Our vision for CHIPS is the creation of a ubiquitous storage infrastructure that enables handheld devices that contain gigabytes of nonvolatile storage, data

… continued on pg. 12

YEAR IN REVIEW

November 1999
- PDS Career Day, Retreat & Workshop
- Final NSIC/NASD public meeting; first SNIA/OBSD meeting, Santa Clara, CA
- Jim Gray; distinguished lecturer, CMU

October 1999
- Erik Riedel defends Ph.D. Dissertation: Active Disks
- First SNIA/OBSD administrative meeting; Seattle, WA
- Garth speaks at OpenSIG99, CMU

September 1999
- DSSC-NSF Fall Review at CMU
- Garth attends SGPFS '99 in Santa Fe, NM
- Bill Joy; distinguished lecturer, CMU

August 1999
- Hot Interconnects 99 - David Nagle on "Network Support for Network Attached Storage".

July 1999
- Howard Gobioff defends Ph.D. Dissertation: Security for NASD
- SDI Seminar: Bennet Yee, University of California, San Diego

June 1999
- USENIX 99, Extreme Linux Workshop - Garth on NASD Scalable Storage Systems
- USENIX 99 - David Petrou on Implementing Lottery Scheduling
- SDI Seminar: David Petrou, CMU

May 1999
- Garth receives Reynold B. Johnson award at ISCA'99 in Atlanta, GA

April 1999
- Parallel Data Systems Spring Open House
- David Nagle attends Active Nets Workshop in San Diego, CA

March 1999
- Jeff Butler and Charles Hardin attend IEEE Infocom in NYC
- Dave Nagle attends Active Networks Team Reviews in Los Angeles, CA.
- DSSC-NSF Spring Review at CMU
- Fay Chang proposed Speculative Execution to Hide I/O Latency
- SDI Seminar: Arif Merchant, HP Labs

February 1999
- OSDI 99 - Fay Chang on Automatic I/O Hint Generation through Speculative Execution
- SDI Seminar: Mustafa Uysal, University of Maryland

January 1999
- SDI Seminar: Christos Faloutsos, CMU

December 1998
- NSIC/NASD quarterly meeting in Milpitas, CA

November 1998
- PDS Career Day, Retreat & Workshop
- ASPLOS VIII, San Jose, CA - Garth Gibson on "A Cost-Effective, High-Bandwidth Storage Architecture." Dave Nagle, Greg Ganger and Charles Hardin also attend.
- Active Nets Research Project Visit at Princeton University, NY
- Dave Nagle attends Active Nets Workshop in NYC with Mike Bigrigg
- Garth and Tara attend Supercomputing '98 in Orlando, FL
- SDI Seminar: Jim Hughes, StorageTek

NEW FACULTY

Mor Harchol-Balter

Mor Harchol-Balter received a BA in mathematics and computer science from Brandeis University in 1988. From 1988-1990, she worked in the machine intelligence technology group at GTE. She received her Ph.D. in computer science from the University of California at Berkeley in 1996. From 1996-1999, Mor was funded by the NSF Postdoctoral Fellowship in the Mathematical Sciences and worked at the Laboratory for Computer Science at MIT. In the Fall of 1999, Mor joined Carnegie Mellon University as an assistant professor.

Mor's research focuses on the design and performance analysis of distributed computer systems including distributed Web servers, distributed supercomputing servers, networks of workstations, and communication networks. She is looking for solutions to common questions involving migration policies, task assignment policies, scheduling policies, etc. Mor's approach involves integrating measured workload distributions into the analysis of systems, rather than the traditional analysis approach based on unrealistic Markovian workload assumptions. She has demonstrated that this method is a much better predictor of reality than conventional analysis, and the difference can be as great as several orders of magnitude. Her empirically driven analysis has resulted in new protocols for distributed systems which significantly outperform the current state-of-the-art and furthermore, defy conventional wisdom in this area.

Mor has filed two patents, both of which have been licensed by industry, and looks forward to more such fruitful research collaborations.

http://www.pdl.cs.cmu.edu/publications/publications.html

Network Support for Network-Attached Storage

Nagle, Ganger, Butler, Goodson & Sabol

Hot Interconnects 1999, August 18 to 20, 1999, Stanford University, Stanford, CA, July 1999.

ABSTRACT

Storage systems represent a vital market with storage densities growing at 60% per year, resulting in 35% to 50% per year decreases in the cost per byte. Unfortunately, current distributed file system architectures severely limit scalable storage. In current distributed file systems, all storage bytes are copied through file server machines between peripheral buses (typically SCSI) and client LANs. In essence, these file server machines act as application-level inter-network routers, converting name-spaces (disk block versus file range) and protocol layers (SCSI versus RPC/UDP/IP). Storage devices, however, are already effective network data transfer engines and traditional client-server protocol stacks and operating system software layers often copy data several times in delivering it to applications, significantly reducing network-attached storage performance. This paper examines networking requirements for storage and the integration of user-level networking with network-attached storage (NAS). To provide context for the networking demands of NAS, we begin by describing alternative network-attached storage architectures and CMU's network-attached storage system. Next, we survey storage's networking requirements and describe how one user-level networking architecture, the VI Architecture (VIA), can be effectively mapped onto our network-attached storage prototype.

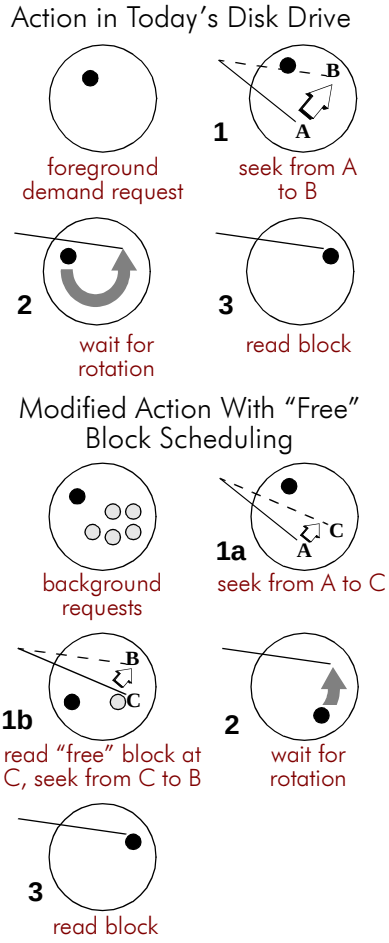
Data Mining on an OLTP System (Nearly) for Free

Riedel, Faloutsos, Ganger & Nagle

CMU SCS Technical Report, CMU-CS-99-151, July 1999.

ABSTRACT

This paper proposes a scheme for scheduling disk requests that takes advantage of the ability of high-level functions to operate directly at individual disk drives. We show that such a scheme makes it possible to support a Data Mining workload on an OLTP (on-line transaction processing) system almost for free: there is only a small impact on the throughput and response time of the existing workload. Specifically, we show that an OLTP system has the disk resources to provide a consistent one third of its sequential bandwidth to a background Data Mining task with close to zero impact on OLTP throughput and response time at high transaction loads. At low transaction loads, we show much lower impact than observed in previous work. This means that a production OLTP system can be used for Data Mining tasks without the expense of a second dedicated system. Our scheme takes advantage of close interaction with the on-disk scheduler by reading blocks for the Data Mining workload as the disk head "passes over" them while satisfying demand blocks from the OLTP request stream. We show that this scheme provides a consistent level of throughput for the background workload even at very high foreground loads. Such a scheme is of most benefit in combination with an Active Disk environment that allows the background Data Mining application to also take advantage of the processing power and memory available directly on the disk drives.



An illustration of 'free' block scheduling. In the original operation, a request to read or write a block causes the disk to seek from its current location (A) to the destination cylinder (B). It then waits for the requested block to rotate underneath the head. In the modified system, the disk has a set of potential blocks that it can read "at its convenience." When planning a seek from A to B, the disk will consider how long the rotational delay at the destination will be and, if there is sufficient time, will plan a shorter seek to C, read a block from the list of background requests, and then continue the seek to B. This additional read is completely 'free' because the time waiting for the rotation to complete at cylinder B is completely wasted in the original operation.

... continued on pg. 8

DISSERTATION ABSTRACT:

Security for a High Performance Commodity Storage Subsystem, CMU-CS-99-160, July 22, 1999.

http://www.ece.cmu.edu/Consortium/Papers/hbg_thesis.pdf

Howard Gobioff, S.C.S.

How do we incorporate security into a high performance commodity storage sub- system? Technology trends and the increasing importance of I/O bound workloads are driving the development of commodity network attached storage devices which deliver both increased functionality and increased performance to end-users. In the network attached world, storage devices co-exist on the network with their clients, application file-managers, and malicious adversaries who seek to bypass system security policies. As storage devices move from behind the protection of a server and become first-class network entities in their own right, they must become actively involved in protecting themselves from network attacks. They must do this while cooperating with higher level applications, such as distributed file systems or database systems, to enforce the application's security policies over storage resources. In this dissertation, I address this problem by proposing a cryptographic capability system which enables application filemanagers to asynchronously make policy decisions while the commodity storage devices synchronously enforce these decisions.

This dissertation analyzes a variety of access control schemata that exist in current distributed storage systems. Motivated by the analysis, I propose a basic cryptographic capability system that is flexible enough to efficiently meet the requirements of many distributed storage systems. Next, I explore how a variety of different mechanisms for describing a set of NASD objects can be used to improve the basic capability system. The result is a new design based on remote execution techniques. The new design places more access control processing at the drive in order to deliver increased performance and functional advantages. Based on the performance limitations of software cryptography demonstrated in a prototype implementation of a network attached storage device, I propose and evaluate an alternative to standard message authentication codes. This allows storage devices to precompute some security information and reduces the amount of request-time computation required to protect the integrity of read operations. Finally, I discuss the availability of cryptographic hardware, how much is required for a network attached storage device, and the implications of adding tamper-resistant hardware to a storage device.

DISSERTATION ABSTRACT:

Active Disks - Remote Execution for Network-Attached Storage, October 12, 1999.

Erik Riedel, E.C.E.

Today's commodity disk drives, the basic unit of storage for computer systems large and small, are actually small computers, with a processor, memory, and 'network' connection, along with the spinning magnetic material that permanently stores the data. As more and more of the information in the world becomes digitally available, and more and more of our daily activities are recorded and stored, people are increasingly finding value in analyzing, rather than simply storing and forgetting, these large masses of data. Sadly, advances in I/O performance have lagged the development of commodity processor and memory technology, putting pressure on systems to deliver data fast enough for these types of data-intensive analysis. This dissertation proposes a system called Active Disks that takes advantage of the processing power on individual disk drives to run application-level code. Moving portions of an application's processing directly to the disk drives can dramatically reduce data traffic and take advantage of the parallelism already present in large storage systems. It provides a new point of leverage to overcome the I/O bottleneck.

This dissertation presents the factors that will make Active Disks a reality in the not-so-distant future, the characteristics of applications that will benefit from this technology, an analysis of the improved performance and efficiency of systems built around Active Disks, and a discussion of some of the optimizations that are possible with more knowledge available directly at the devices. It also compares this work with previous work on database machines and examines the opportunities that allow us to take advantage of these promises today where previous approaches have not succeeded. The analysis is motivated by a set of applications from data mining, multimedia, and databases and is performed in the context of a prototype Active Disk system that shows dramatic speedups over a system with traditional, "dumb" disks.

… continued on pg. 12

Greg salutes the PDL on his birthday.

August 1999
Twins for the Zhangs
Jackie and Hui have announced the arrival of twin boys...Alex was born at 2:17 p.m. August 18, 1999, weighing 8 lbs. 1 oz. and Eric came along at 2:34 p.m., weighing 7 lbs. 6 oz. Congratulations!

July 1999
Mowry Among The Top 100!*
Todd Mowry has been selected as one of 100 young innovators, with "the potential to make significant technological contributions in the next century" by Technology Review. These folks are the ones to watch and promise to impact the world in the coming years. Todd and the other recipients will be profiled in a Special Issue of TR in November and honored at an awards gala on November 4. The program also includes an historic, one-day gathering of business, technology and academic leaders who will meet to celebrate these young visionaries. Watch for details at www.techreview.com/tr100/.

July 1999
PANASAS Launched
Garth has taken a part time leave from his duties at CMU to participate in the launching of PANASAS (which, by the way, doesn't stand for anything in particular), a storage solutions start up. Garth and several others have set up their office just off of Craig Street, here in Pittsburgh - close enough so we still see them fairly often.

Congratulations to Garth and Stephanie!

July 1999
John Griffin Awarded NSF Fellowship
In July, John Griffin was awarded a three year National Science Foundation Graduate Research Fellowship which will support his pursuit of an advanced degree. Though the award is not associated with a particular project, in his application John discussed his ongoing work on "robustness testing of communication systems" (the PIRANHA project directed by Phil Koopman and Greg Ganger). The award is presently supporting his work with Steve Schlosser on the architecture of MEMS-based storage systems.

June 1999
Garth Gibson and Stephanie Byram Wed
Congratulations to Garth and Stephanie who were married in a small private ceremony at Hienz Chapel on June 18. Following an outdoor dinner reception at the Frick House Gardens, Garth and Stephanie embarked on an ambitious honeymoon tour of Peru, climbing into the heights of the Andes Mountains.

May 1999
Fay Chang Receives Intel Graduate Fellowship*
Fay Chang has been awarded the 1999-2000 Intel Graduate Fellowship in Computer Science, considered among the company's most prestigious academic honors. The award, which recognizes outstanding academic achievement and scientific relevance of the recipient's research to Intel's industrial focus, includes funding and a microprocessor-based personal computer system from Intel.

May 1999
SCS Staff Recognition Awards*
One of our own was a winner of this year's SCS Staff Recognition Awards! Congratulations to Paul Mazaitis for exemplifying the very best in job performance, dedication, and winning attitude. Patty Mackiewicz and Jennifer Landefeld were also among the nominees for these awards. SCS is richer, wiser, and more productive for all their efforts.

May 1999
Clay Mowry Group Inc. Launches Baby 1.0
PITTSBURGH, May 12 (Reuters) - Shares of Clay Mowry Group Inc. (NASDAQ:CMGI) rose sharply today on announcement of a 3-for-2 split coinciding with the launch of the company's first product: Baby Version 1.0 ("frogs and snails and

And congratulations to Todd and Karen on the arrival of their son Connor!

... continued on pg. 13

… continued from pg. 5

NASD Scalable Storage Systems

Gibson, Nagle, Courtright, Lanza, Mazaitis, Unangst & Zelenka

USENIX99, Extreme Linux Workshop, Monterey, CA, June 1999.

ABSTRACT

The goal of CMU's Network-Attached Secure Disks (NASD) project is to define the next era of storage system interfaces and architectures. To encourage industry standardization of a compliant storage device/subsystem interface, we are working closely with the National Storage Industry Consortium working group on network-attached storage. Our experimental demonstration of the NASD interface's value is device and filesystem prototype software that delivers the scalability inherent in a NASD storage architecture. To engage the academic community and to provide a reference implementation for industry development, CMU is releasing its Linux and Digital UNIX ports of this software. In this paper, we overview the NASD scalable storage architecture and the codebase we are releasing for Linux.

Embedded Security for Network-Attached Storage

Gobioff, Nagle & Gibson

CMU SCS Technical Report, CMU-CS-99-154, June 1999.

ABSTRACT

As storage interconnects evolve from single-host small-scale systems, such as traditional SCSI, to the multi-host Internet-based systems of Network-attached Secure Disks (NASD), protecting the integrity of data transfers between client and storage becomes essential. However, it is also computationally expensive and can impose significant performance penalties on storage systems. This paper explores several techniques that can protect the communications integrity of storage requests and data transfers, imposing very little performance penalty and significantly reducing the amount of required cryptography.

Central to this work is an alternative cryptographic approach, called "Hash and MAC," that reduces the cost of protecting the integrity of read traffic in storage devices that are unable to generate a message authentication code at full data transfers rates. Hash and MAC does this by precomputing security information, using and reusing the precomputed information on subsequent read requests. We also present a refined Hash and MAC approach that uses incremental hash functions to improve the performance of small read and write operations as well as non-block-aligned operations.

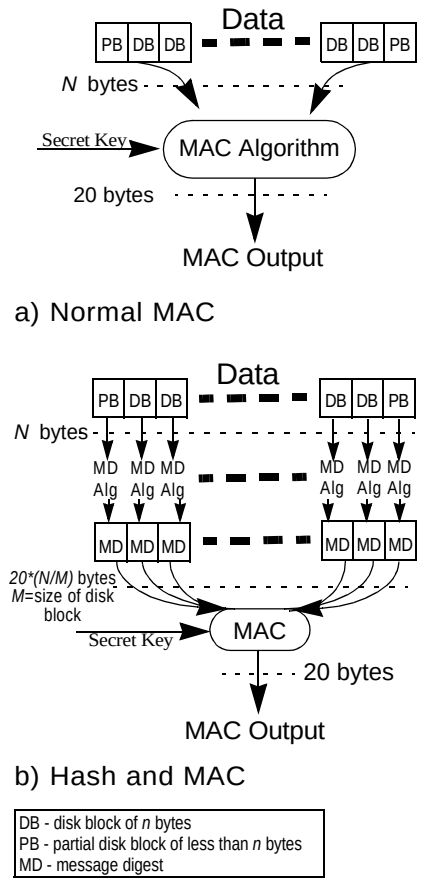
Dynamic Function Placement in Active Storage Clusters

Amiri, Petrou, Ganger & Gibson

CMU SCS Technical Report, CMU-CS-99-140, June 1999.

ABSTRACT

Traditional distributed filesystems statically partition their principle device-independent functions (applications, caching, reliability, directory service, and synchronization) between client and server machines. In this paper, we demonstrate that dynamic placement of these functions can improve performance. Moreover, we show that black-box monitoring of the resource consumption of these functional modules and

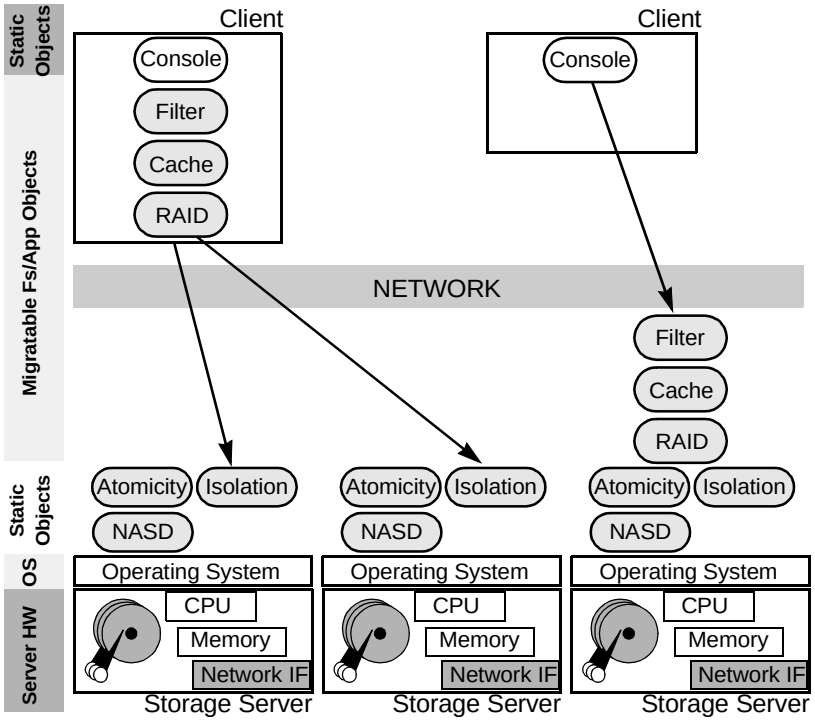


The Hash and MAC approach reduces the amount of computation involving the secret key. Each message consists of a sequence of full disk blocks which may be preceded and/or followed by a partial disk block. In a), most MAC algorithms involve the key in the computation over all the bytes of data and process the data linearly. In b), Hash and MAC does not involve the key until late in the computation. This enables parallelization and precomputation for increased performance. The dotted lines indicate the amount of data that passes in and out of the message digest or MAC algorithms at that different stages of computation. In the Hash and MAC approach, a calculation over only 20 bytes per disk block involves the key while the rest of the computation can potentially be precomputed without knowledge of the key.

… continued on pg. 9

… continued from pg. 8

http://www.pdl.cs.cmu.edu/publications/publications.html



Decomposition of a prototype cluster filesystem running on ABACUS. Active storage servers provide basic non-migratable modules for serving flat object stores, and ensuring concurrency control and recovery (the NASD, Isolation and Atomicity modules). The rest of the filesystem function is encapsulated in modules (RAID, Caching, etc.) which can be dynamically migrated at run-time between the servers and clients. Data intensive applications are similarly decomposed into modules which can be migrated between clients and storage servers.

their intercommunication, and of the resources available throughout the cluster is sufficient to describe the site-specific and load-specific context needed to make dynamic placement decisions. Our demonstration distributed filesystem is built on ABACUS, a distributed runtime system for automating the placement of functional modules among the cluster's machines.

ABACUS defines an explicitly-migratable programming model, allowing application and filesystem programmers to define acyclic graphs of migratable component modules. Driven by observations of per-module resource usage and per-machine resource availability, ABACUS employs a cost-benefit performance model to adapt the partitioning of each module graph among client and server machines. As workload and resource conditions change, ABACUS updates its performance model and re-migrates modules accordingly. Measurement of various filesystem and application scenarios show that ideal function placement can often improve performance by 2-10X, that ABACUS is almost always within 30% of the ideal, and that ABACUS adapts best when the potential performance impact is largest (exactly as one would desire).

Implementing Lottery Scheduling: Matching the Specializations in Traditional Schedulers

Petrou, Milford & Gibson

USENIX 1999, Monterey CA, June 9-11, 1999.

ABSTRACT

We describe extensions to lottery scheduling, a proportional-share resource management algorithm, to provide the performance assurances present in traditional non-real time process schedulers. Lottery scheduling enables flexible control over relative process execution rates with a ticket abstraction and provides load insulation among groups of processes using currencies. We show that a straightforward implementation of lottery scheduling does not provide the responsiveness for a mixed interactive and CPU-bound workload offered by the decay usage priority scheduler of the FreeBSD operating system. Standard lottery scheduling ignores kernel priorities used in the FreeBSD scheduler to reduce kernel lock contention. We show how to use dynamic ticket adjustments to incorporate into a lottery scheduler the specializations present in the FreeBSD scheduler to improve interactive response time and reduce kernel lock contention. We achieve this while maintaining lottery scheduling's flexible control over relative execution rates and load insulation. In spite of added scheduling overhead, the throughput of CPU-bound workloads under our scheduler is within one percent of the FreeBSD scheduler for all but one test. We describe our design, evaluate our implementation, and relate our experience in deploying our hybrid lottery scheduler on production machines.

… continued on pg. 10

… continued from pg. 9

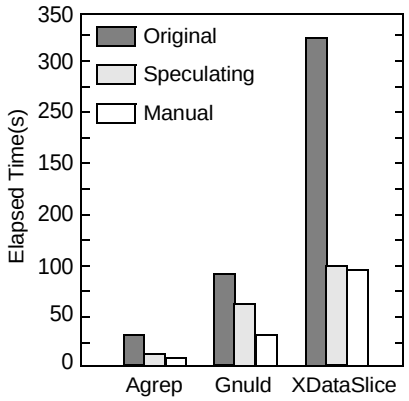
Automatic I/O Hint Generation through Speculative Execution

Chang & Gibson

Proceedings of the 3rd Symposium on Operating Systems Design and Implementation, February 1999.

ABSTRACT

Aggressive prefetching is an effective technique for reducing the execution times of disk-bound applications; that is, applications that manipulate data too large or too infrequently used to be found in file or disk caches. While automatic prefetching approaches based on static analysis or historical access patterns are effective for some workloads, they are not as effective as manually-driven (programmer-inserted) prefetching for applications with irregular or input-dependent access patterns. In this paper, we propose to exploit whatever processor cycles are left idle while an application is stalled on I/O by using these cycles to dynamically analyze the application and predict its future I/O accesses. Our approach is to speculatively pre-execute the application's code in order to discover and issue hints for its future read accesses. Coupled with an aggressive hint-driven prefetching system, this automatic approach could be applied to arbitrary applications, and should be particularly effective for those with irregular and, up to a point, input-dependent access patterns. We have designed and implemented a binary modification tool, called "SpecHint" that transforms Digital UNIX application binaries to perform speculative execution and issue hints. TIP, an informed prefetching and caching manager, takes advantage of these application-generated hints to better use the file cache and I/O resources. We evaluate our design and implementation with three real-world, disk-bound applications from the TIP benchmark suite. While our techniques are currently unsophisticated, they perform surprisingly well. Without any manual modifications, we achieve 29%, 69% and 70% reductions in execution time when the data files are striped over four disks, improving performance by the same amount as manually-hinted prefetching for two of our three applications. We examine the performance of our design in a variety of configurations, explaining the circumstances under which it falls short of that achieved when applications were manually modified to issue hints. Through simulation, we also estimate how the performance of our design will be affected by the widening gap between processor and disk speeds.



Performance improvement. Original corresponds to the original, non-hinting applications; Speculating corresponds to the applications transformed to perform speculative execution for hint generation; and Manual corresponds to the applications manually modified to issue hints. In all cases, the non-hinted read calls issued by the applications invoked the operating system's sequential read-ahead policy.

Scalable Concurrency Control and Recovery for Shared Storage Arrays

Amiri, Gibson & Golding

CMU SCS Technical Report, CMU-CS-99-111, February 1999.

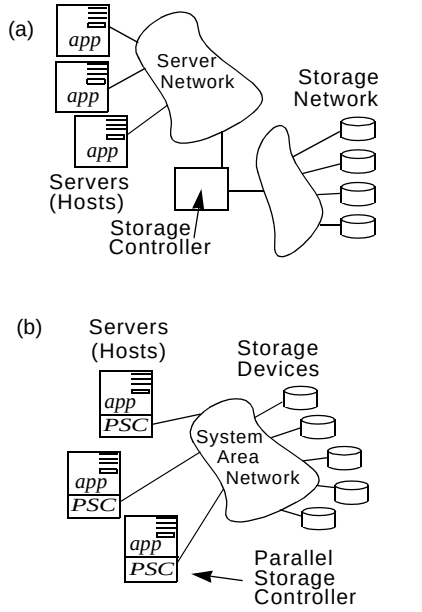
ABSTRACT

Shared storage arrays let thousands of storage devices be shared and directly accessed by end hosts over switched system-area networks, giving databases and filesystems highly scalable, reliable storage. In such systems, however, concurrent host I/Os can span multiple shared devices and access overlapping ranges potentially causing inconsistencies for redundancy codes and data read by end hosts. To enable existing applications to run unmodified and simplify the development of future ones, we desire a shared storage array to provide the illusion of a single controller without the scalability bottleneck and single point of failure of an actual single controller. We show how rapidly increasing storage device intelligence coupled with storage's special characteristics can be successfully exploited to arrive at a high-performance solution to this storage management problem. In particular, we examine four concurrency control schemes and specialize them to shared storage arrays; two centralized: simple server locking, and server locking with leased callbacks and two distributed based on device participation: distributed locking using storage-device-embedded lock servers and timestamp ordering using loosely synchronized clocks. Simulation results show that both centralized locking schemes suffer from scalability limitations.

… continued on pg. 11

… continued from pg. 10

Also, callback locking is particularly suspect if applications do not have much inherent locality and if the storage system introduces false sharing. Distributed concurrency control with device support is attractive as it scales control capacity with storage and performance capacity and offers the opportunity to piggyback lock/ordering messages on operation requests eliminating message latency costs. Simulations show that both storage-optimized device-based protocols exhibit close to ideal scaling achieving 90-95% of the throughput possible under totally unprotected operation. As well, timestamp ordering uses less network resources, is free from deadlocks and has performance advantages under high load. We show how timestamp ordering can be extended with careful operation history recording to ensure efficient failure recovery without inducing I/Os under normal operation. This brings the overhead of concurrency control and recovery to a negligible level thereby realizing the scalability potential of the shared array I/O architecture.



Conventional storage systems: (a) all host requests are serialized at a single storage controller and shared storage array, (b) network-attached storage devices are shared and directly accessed by hosts and serializability must be ensured by a distributed protocol, implicitly transforming low-level host software into a parallel storage controller.

http://www.pdl.cs.cmu.edu/publications/publications.html

Integrity and Performance in Network Attached Storage

Gobioff, Nagle & Gibson

CMU SCS Technical Report, CMU-CS-98-182, December 1998.

ABSTRACT

Computer security is becoming increasingly important in networked computing. We examine the issue of high-performance network security and integrity by focusing on integrating security into a network storage system. Emphasizing the cost-constrained environment of storage, we examine how current software-based cryptography cannot support storage's Gbit/sec transfer rates and introduce a novel message authentication code, based on stored message digests. This lets storage deliver a factor of five improvement in our prototype's integrity protected bandwidth without hardware acceleration for common read operations. For receivers, where precomputation cannot be done, we describe an inline message authentication code to minimize buffering requirements.

NASD Prototype Released

Garth Goodson gets ready to climb the wall at Nemacolin while David Petrou looks on.

… continued from pg. 1

but, in compliance with US government regulations, it does not support encryption and decryption for privacy. This code has been tested on Dell/Intel Pentium workstations, DEC/Compaq Alphas and DEC/Intel StrongArm embedded microcontrollers. Its user-level client functions have also been ported to Solaris, Irix and FreeBSD (client-only).

NASD has been designed by CMU's Parallel Data Lab team of academic researchers working with members of the Parallel Data Consortium (PDC), a group of interested industry partners able to offer insight into the state of the storage industry. The PDL co-founded the National Storage Industry Consortium (NSIC) Network-Attached Storage Device working group (http://www.nsic.org/nasd/), whose participants actively pursue the development, exploration, validation and documentation of network-attached storage technology and standards. PDL's NASD has been a principle inspiration for NSIC's NASD working group. It has also recently become a multiple-working-group technology area in the newly formed Storage Networking Industry Association (SNIA, http://www.snia.org), an organization initiated to develop standards and markets for networked storage.

… continued from pg. 6

THESIS PROPOSAL:

Using Speculative Execution to Automatically Hide I/O Latency, March 22, 1999.

Fay Chang, S.C.S

Previous research has demonstrated that we can greatly improve the performance of disk-bound applications by exploiting I/O parallelism through aggressive prefetching. However, our ability to make good prefetching decisions is limited by our ability to accurately predict the future data needs of applications. While, in some cases, programmers may be willing to manually modify applications to obtain accurate and timely prefetching information, automatic approaches to I/O prefetching are needed to make the benefits of aggressive prefetching readily accessible. Unfortunately, although existing automatic approaches can be very effective, they apply in limited situations. Prefetching according to a common access pattern, like sequential access, can only help applications whose accesses conform to such patterns. Dynamic history-based prefetching approaches can only help when there is recognizable repetition in accesses, and their ability to recognize repetition is limited by their resource requirements. Finally, prefetching approaches based on static analysis are constrained by difficult interprocedural analysis problems.

This thesis project explores a novel approach to automatically determining an application's future data needs that could potentially apply to a wide range of applications. Our approach has three basic components: first, applications attempt to discover their own future data needs by speculatively pre-executing their code; second, the cost of this additional execution is hidden by exploiting otherwise wasted processing cycles during I/O stalls; and third, applications are automatically modified to perform speculative execution using only simple, generic static analyses and transformations. This approach has several advantages over previous approaches, but it is hampered by the lack of complete state information during speculative execution. In addition, speculative execution requires machine resources that may be highly contended in, for example, a multiprogrammed environment. This research explores to what extent, and under what conditions, our approach can determine the future data needs of applications in a manner that will allow applications to automatically obtain significant benefits through I/O prefetching.

Year of Network Attached Storage

… continued from pg. 3

archives that never become outdated because the archives are self contained (include storage, processing elements, and access software), and fully-integrated storage/processing nodes that fully enable transparent function migration. The center has formed a 12 year research agenda, with work already beginning on two fronts. The Electrical and Computer Engineering MEMS Lab is busy building the first generation of read/writable MEMS storage devices. In tandem, the PDL is researching CHIP's system-level design trade-offs (e.g., bandwidth, capacity, data layout, software management), so that CHIPS can best support a wide range of current (e.g., database, file systems, speech recognition) and future applications.

The PDL continues to build a strong, broad research agenda in storage and systems architecture. One of the great strengths of our research projects are their synergy. The lab's initial RAID and prefetching work laid the foundation for understanding how to architect future storage systems such as NASD. NASD research continues with our networking and file system research, and has lead to our Active Disk and ABACUS projects. Now, with CHIPS, we evolve our notion of storage and integration with computing to create the ubiquitous storage infrastructure of the future. However, the most important aspect of the Parallel Data Lab are its students, staff, and faculty. PDL continues to attract the best and brightest people - and it is their ideas, enthusiasm, and very hard work that continue to make the PDL the premier research lab in storage and systems architecture.

Past retreat fun at the climbing wall and on the social hike.

… continued from pg. 7

puppy dog tails" edition), code-named *Connor*. Analysts were impressed that CTO Karen Clay was able to bring this product to market in just nine months. (Although, like many high-tech products, the delivery time did slip by one week). "Karen deserves all of the credit for making Baby 1.0 a big success," says CEO Todd Mowry, who, despite his technical background, played a largely administrative role in this project. Baby 1.0 successfully booted its first program, "Hello World" (pronounced "Waahh!!!") at 11:44pm, May 12th.

May 1, 1999

Christos Faloutsos Involved in Successful NSF Proposal

Besides acting as the ACM SIGMOD '99 Program Chair and joining the Executive Committee of ACM SIG KDD (special interest group on Knowledge Discovery in Data and Data Mining), Christos became the recipient of an NSF grant in May. Also involved in the "Informedia-II: Integrated Video Information Extraction and Synthesis for Adaptive Presentation and Summarization from Distributed Libraries" project are Howard Wactlar (PI), Michael Christel, Alex Hauptmann and Takeo Kanade. Congratulations.

April 23, 1999

First PDS Spring Open House A Success

The first annual Parallel Data Systems Spring Open House was a great success. The Open House offered our Consortium Member companies the chance to meet with the PDL members at CMU. Though some talks were given, the focus of this event was on presenting the PDL's latest research through computer demonstrations, posters and discussion. Comments from industry attendees indicated that they especially appreciated the opportunity the Open House gave them for interaction with the CMU attendees, over and above that available at our fall Retreat.

March 1999

Seth Goldstein Receives NSF Career Award*

Seth Goldstein is the recipient of a National Science Foundation Early Career Development Program CAREER Award, for his project "Architecture and Compilers for Tiger Machines." CAREER awards, until recently known as Presidential Young Investigator awards, provide support to outstanding junior faculty members in their pursuit of "research and education of the highest quality...integral components in stimulating the discovery and learning process."

February 1999

Fay Chang Wins Best Paper at OSDI*

Fay Chang's paper, "Automatic I/O Hint Generation through Speculative Execution," was awarded best student paper (tied) at the 3rd Symposium on Operating Systems Design and Implementation (OSDI), sponsored by USENIX, IEEE TCOS and ACM SIGOPS, in New Orleans (February 23 - 25, 1999). Fay was the conference's first speaker, and did a super job even though her voice threatened to fail at any moment right up to and during her talk. She received compliments from a wide variety of leaders in the field.

February 1999

Garth Gibson Receives Newell Award*

Garth received his gold medal at the CS Whole Department Meeting in February, in recognition of outstanding work and dedication to quality research. Established in 1992, the Newell Award pays tribute to "real science," pursued with the integrity, enthusiasm, and drive exemplified by Allen Newell's scientific career.

February 1999

Todd Mowry Selected a Sloan Research Fellow*

Todd Mowry has been selected as a Sloan Research Fellow. Awarded annually by the Alfred P. Sloan Foundation, this fellowship acknowledges outstanding young scientists and economists who are engaged in research at the frontiers of computer science, physics, mathematics, chemistry, economics and neuroscience. The program originated in 1955 and selects the awardees from hundreds of exemplary candidates, all in the early stages of their professional careers. The small group of final recipients are selected on the basis of "their exceptional promise to contribute to the advancement of knowledge."

January 1999

Karen Lindenfelser Receives College of Engineering Staff Award**

Karen Lindenfelser, faculty secretary in the Department of Electrical and Computer Engineering (ECE), was a top honoree at the College of Engineering's fifth annual Staff Recognition Awards, Jan. 7 in the Singleton Room, Roberts Hall. Selected by a committee of her peers from nominations across the college, Karen was recognized for exception-

… continued on pg. 14

A well deserved rest for Khalil?

Automated Disk Drive Characterization

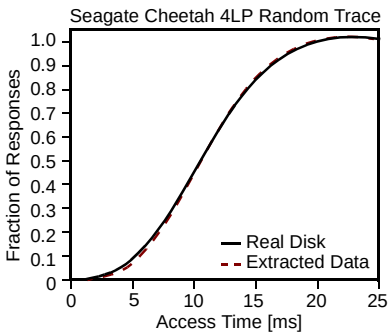
Jiri Schindler & Gregory Ganger

Sophisticated disk schedulers and detailed disk simulators require extensive and accurate characterizations of disk drive performance. Such characterizations include data regarding mechanical delays, onboard caching and prefetching algorithms, command and protocol overheads and logical-to-physical block mappings. Unfortunately, accurate characterizations have traditionally been difficult to acquire, forcing researchers to rely on a collection of ad-hoc techniques. As a result, detailed characterizations exist (in the public domain) for very few disk drives.

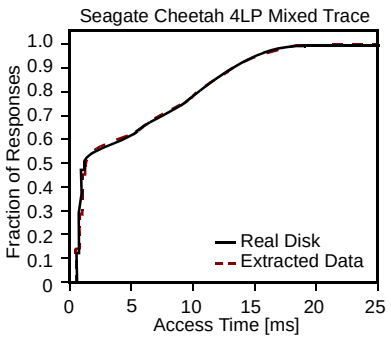
Over the past year, we have developed DIXtrac to address this problem. Without human intervention, DIXtrac can automatically extract from SCSI disk drives accurate values for over 100 performance-critical parameters. DIXtrac runs as a user-level application on LINUX, using the /dev/sg interface to pass SCSI commands directly to the device driver. It uses a collection of pre-programmed test vectors to measure timings for mechanical and command processing overheads and uses expert-system-like algorithms to identify layout and caching policies. Collectively, these parameters provide sufficient characterization to allow extremely accurate simulation of disk performance.

DIXtrac has been used to characterize six different disk models from three manufacturers. Performing a characterization consists of simply pointing DIXtrac at the disk of interest. The characterization process generally requires less than 3 minutes to complete (5 minutes in one case). Using the extracted characteristics to parameterize the DiskSim simulator, we observe very close matches between simulated and measured disk performance. For example, the adjacent figure shows measured and simulated access time distributions for a validation workload of 10000 requests.

Very soon, we expect to release a new version of the DiskSim simulator together with a collection of extracted disk characteristics. We hope to add extracted data for many disks to our on-line database of characteristics. We also plan to release the DIXtrac tool for public use.



Demerit Figure 0.16
Random Trace
10000 requests (2/3 READs, 1/3 WRITEs)
Random LBNs distributed across entire disk
1 KB to 8 KB size



Demerit Figure 0.17
Mixed Trace
5000 requests (2/3 READs, 1/3 WRITEs)
20% sequential requests
30% local requests
1 KB to 8 KB size

Synthetic traces run on a real disk compared to runs on a disk simulator with extracted parameters

AWARDS & OTHER PDL NEWS

… continued from pg. 13

al job performance, extraordinary dedication, an unfailing positive attitude and for being a team player who serves as a role model for colleagues.

November 1998

Garth Gibson awarded Reynold B. Johnson Information Storage Award*

Garth Gibson, David Patterson (UC Berkeley), and Randy Katz (UC Berkeley) have been awarded the IEEE Reynold B. Johnson Information Storage Award for the development of "Redundant Arrays of Inexpensive Disks (RAID)." The award, established in 1991 to acknowledge outstanding contributions in the field of information storage with emphasis in the area of computer storage, is awarded to an individual or a team of not more than three, and consists of a bronze medal, certificate and $5,000 cash prize. The award honors Reynold B. Johnson, recognized as a pioneer of magnetic disk technology and founding manager of the IBM San Jose Research and Engineering Laboratory in 1952. Garth, Dave and Randy join an illustrious and distinguished list of other honored researchers, including most recently Jean-Pierre Lazzari (1998/SILMAG), Alan Shugart (1997/Seagate Technology), Nobutake Imamura (1996/Tosoh Corporation), and James Lemke (1995/Recording Physics, Inc.).

Several of these articles have been drawn from campus news sources: SCS Today (*) and CMU News (**).

FACULTY

Mor Harchol-Balter joined the PDL this fall and is a new faculty member of the School of Computer Science. Please see page 4 for a brief biography.

Tara Madhyastha, a graduate of the University of Illinois at Urbana-Champaign, completed her Postdoctoral Fellowship with the PDL this past June. After an arduous interviewing schedule, she accepted a position as Assistant Professor in the Computer Science and Engineering Department at the University of California at Santa Cruz. There, Tara will teach and continue her research on characterizing and optimizing collective I/O access patterns.

STAFF

Nat Lanza joined the PDL staff as a full time Staff Programmer this spring. He has been working on the NASD project and continues to take courses towards his Bachelor's degree in CS.

David Rochberg joined the PDL staff as a programmer in June, when he decided to take a hiatus from his Ph.D. research.

Mike Bigrigg has moved to CMU's Institute for Complex Engineered Systems (ICES), so we still get to see him occasionally.

Bill Courtright, a former grad student of the PDL and most recently the PDL's Executive Director, has moved to PANASAS to manage Garth.

Sean Levy left the PDL in August to start his own company after deciding that there was a market for his innovative thoughts on the way the web works.

Marc Unangst left the PDL this July after several years with us to join Garth and Bill at PANASAS.

Jim Zelenka, after over 5 years with the PDL, much of which was spent overseeing the NASD project, has also decided to make PANASAS his new place of employment.

GRAD STUDENTS

David Friedman graduated with a BS in Electrical and Computer Engineering from CMU and is now working on his Master's degree under the direction of Dave Nagle. He is currently working on a project in intelligent network interface cards.

Ed Hogan received a B.S. in Computer Science and a Bachelor of Arts in Art Studio at the University of Maryland, College Park. At CMU, he is enrolled in the Information Networking Institution, an interdisciplinary program formed by SCS and the Graduate School of Industrial Administration. He is working towards a Master's degree under the direction of Garth Gibson, researching the problems of semantic mismatch that arise when providing various clients access to files over a multi-protocol network file system.

Andrew Klosterman completed his Bachelor of Science in Electrical Engineering at the University of Dayton (Ohio). Research for his Master's degree in ECE is focused on network based security issues, including intrusion detection, multimodal authentication and user behavior audits. Greg Ganger is his advisor.

Mike Schienholtz graduated from ECE with a Bachelor of Science and is now researching Secure Networked Storage under the supervision of Greg Ganger as part of the requirements for a Master's degree.

John Strunk received his Bachelor's degree from the Department of Computer Engineering at Georgia Tech. He plans on entering the Ph.D. program after completing his Master's degree. He is currently working in the Department of ECE with Greg Ganger on computer system security.

Neal Tibrewala graduated from the Department of ECE in 1998 and is working towards a Master's degree and applying to the Ph.D. program. Greg Ganger is directing his research in fast servers and alternate SAN networking technologies.

Howard Gobioff defended his dissertation in July and left soon thereafter for the San Francisco area where he has become a Software Engineer at Google, a search engine start-up.

John Kliegman finished his M.S. and started with Quackware as a Software Engineer. The company moved to Silicon Valley from Pittsburgh in September.

Erik Riedel defended his Ph.D. research on October 12 and moved west to continue his research in the Storage Systems Program at Hewlett-Packard Laboratories

UNDERGRADUATES

Matt Monroe joined us as an undergrad programmer at the beginning of May this year. He is in his senior year in the Department of CS.

Ross Cohen graduated with a BS in Mathematics this spring and has gone west to work for Quantum Corporation.

Meg Gaj is still at CMU and in her senior year. She decided to leave the PDL to become an assistant to the Department of CS Webmaster.

Adam Phelps finished his degree in Computer Science this spring and stayed on as an undergraduate programmer with the PDL through the summer. This fall, he went on to graduate studies at Stanford University.

Ben Schmidt graduated from CMU last spring, got married and moved to Mountain View, CA and Nomadic Technologies to work on mobile networking and robotics.

NSIC-NASD Meeting Agendas: 1998 - 1999

Informally chaired by CMU's Garth Gibson, the working group on Network Attached Storage Devices in the National Storage Industry Consortium (NSIC) sponsors a quarterly public workshop. Following are the agendas from the past year. Online versions of most of these presentations are available on the NSIC web site at:
http://www.nsic.org/nasd/meetings.html

NASD Meeting: May 18, 1999
File Systems

- Garth Gibson, Carnegie Mellon, NASD Overview plus CMU's NASD Security and Scalable Bandwidth
- Khalil Amiri, Carnegie Mellon, Scalable Concurrency Control and Recovery for Shared Storage Arrays
- Gordon Harris, CrosStor, CIFS over SAN
- John Hufferd, IBM, Shared File Systems: SANs meet LANs
- Mark Carlson, Sun, Service Level Agreements for File Systems
- Remzi Arpaci-Dusseau, University of California, Berkeley, Building Systems with Intelligent Components: Panacea or Pandemonium?
- Jeff Chase, Duke University, Network-Attached Memory in NASD File Services
- Blake Lewis, Network Appliance, Smart Filers and Dumb Disks
- David Golds, Microsoft, NTFS: Raising the Bar for File Systems

NASD Meeting: August 17, 1999
Network Storage for Databases: Asset, Apathy, or Albatross?

- Dave Anderson, Seagate, NASD Overview
- Kimberly Keeton, UC Berkeley, Performance Analysis of Decision Support Database Workloads on Intelligent Disks
- Mustafa Uysal, University of Maryland, Evaluation of Active Disks for Large Decision Support Databases
- Erik Riedel, Carnegie Mellon, Active Disk Architecture for Databases
- Joe Hellerstein, UC Berkeley, Braking the Bandwagon: A Sober Look at Intelligent Disk Research
- Bill O'Connell, IBM, DB2 Directions with NASD and Active Disk
- Dina Bitton and Boris Gelman, Recursion Dynamics, Inc., Exploiting Shared Disk Intelligence
- Bill Bridge, Oracle, Intelligent Disks for Databases, Proposals for Exploiting Shared Disk Intelligence
- James Hamilton, Microsoft, A Database View of Intelligent Disks

SNIA Meeting: October 19, 1999
SNIA Object Based Storage Working Group Kick-off Meeting

- Informal presentations by David Anderson, working group acting chair, Chuck Milligan, Jack Gelb and Garth Gibson.

ANSI Meeting: Nov. 5-9, 1999

- Presentation of OBSD to ANSI X3T10 committee.
- ftp://ftp.t10.org/t10/document.99/99-315r0.pdf

Final NASD Meeting: Nov. 30, 1999

- Presentation of Object-Based Storage Device (OBSD)
- Proposal to ANSI X3T10
- OBSD proposal document available: http://www.nsic.org/nasd/final.pdf

Register to attend, free of charge, by sending email to Barbara Brittain (barbara@nsic.org) and indicate if you would like a 5 minute speaking slot for responding to the OBSD document.

SNIA DBS Meeting: December 1, 1999 (tentative)
First technical meeting; Santa Clara, CA.