To The Faculty Search Committee:

In response to your advertisement, I am applying for a tenure-track position at the rank of assistant professor in your department. I received my Ph.D. in Computer Science from the Georgia Institute of Technology in August 2005, under the direction of Professor Karsten Schwan. I am currently a Postdoctoral Research Fellow at the University of New Mexico, working with Professors Barney Maccabe and Patrick Bridges. My research interests are in experimental systems, data and metadata management, and security, in application domains including high-performance and ubiquitous/pervasive computing.

As a postdoctoral researcher, I have worked closely with faculty and students in the Scalable Systems Laboratory at UNM, as well as with researchers at Sandia National Laboratories. In particular, I am currently investigating frameworks for supporting scalable storage in petascale scientific applications, as well as techniques for building custom OS kernels for the next generation of massively parallel supercomputers. My collaborations during my appointment at UNM have produced peer-reviewed publications and new multi-year research funding awarded by the National Science Foundation.

My doctoral dissertation introduced a novel mechanism which enables efficient data protection and manipulation at the middleware level, supporting the construction of secure and scalable applications in heterogeneous, highly dynamic computing environments. I have significant research experience in areas ranging from efficient low-level communication in heterogeneous systems to wide-area network services; from applications targeting resource-constrained computing environments to high-performance scientific codes; and also in application and network security.

As regards my teaching experience, I was the sole instructor for undergraduate Operating Systems classes both at Georgia Tech (Summer 2002) and at UNM (Spring 2007). I mentor several graduate students in the Scalable Systems Lab at UNM, helping to determine group research strategy and providing individual direction and feedback on specific research issues. My teaching interests include classes in operating systems, distributed systems, computer privacy and security, and related topics.

My academic activities and collaborations to date have been productive and rewarding, and I believe they will provide for me the foundation of a strong, independent research program. Please find enclosed my curriculum vitae, including a list of references and statements of my research and teaching interests. If I can provide any other information, please do not hesitate to contact me. Thank you for your time and consideration; I look forward to hearing from you.

Sincerely,

Patrick M. Widener

encl: Curriculum Vitae
Research Statement
Teaching Statement

Patrick M. Widener

Curriculum Vitae

Department of Computer Science
University of New Mexico
MSC 01-1130
1 University of New Mexico
Albuquerque, NM 87131

(505) 277-9211 (Office)
(505) 553-4564 (Home)
pmw@cs.unm.edu
http://www.cs.unm.edu/∼pmw/

RESEARCH INTERESTS

Experimental systems; operating systems; generation and use of metadata; information privacy and security; application domains including high-performance, pervasive, and agile computing.

PROFESSIONAL PREPARATION

Doctor of Philosophy, Computer Science, 2005, Georgia Institute of Technology, Atlanta, GA.
Dissertation title: *Dynamic Differential Data Protection for High-Performance and Pervasive Applications and Services*
Professor Karsten Schwan, advisor.

Master of Computer Science, 1992, University of Virginia, Charlottesville, VA.

Bachelor of Science (with distinction in Computer Science), 1990, James Madison University, Harrisonburg, VA.

APPOINTMENTS

Postdoctoral Research Fellow — September 2005 - present
Department of Computer Science, University of New Mexico, Albuquerque, NM

Graduate Research Assistant — May 1999 - August 2005
College of Computing, Georgia Institute of Technology, Atlanta, GA

Intel Foundation Graduate Research Fellow — September 2002 - August 2003
College of Computing, Georgia Institute of Technology, Atlanta, GA

Instructor — Summer 2002
College of Computing, Georgia Institute of Technology, Atlanta, GA

Graduate Teaching Assistant — September 1998 - May 1999
College of Computing, Georgia Institute of Technology, Atlanta, GA

PUBLICATIONS

Journal Articles

1. Rolf Riesen, Ron Brightwell, Patrick Bridges, Trammell Hudson, Arthur Maccabe, Patrick Widener and Kurt Ferreira. 'Designing and Implementing Lightweight Kernels for Capability Computing'. In *Concurrency and Computation: Practice and Experience*. Accepted for publication.
2. Patrick Widener, Greg Eisenhauer, Karsten Schwan and Fabian E. Bustamante. 'Open Metadata Formats: Efficient XML-Based Communication for High-Performance Computing'. In *Cluster Computing: The Journal of Networks, Software Tools, and Applications* (5), 2002, pp.315-324. Invited submission.

Book Chapters

1. Karsten Schwan, Brian F. Cooper, Greg Eisenhauer, Ada Gavrilovska, Matt Wolf, Hasan Abbasi, Sandip Agarwala, Zhongtang Cai, Vibhore Kumar, Jay Lofstead, Mohamed Mansour, Balasubramanian Seshasayee and Patrick Widener. "AutoFlow: Autonomic Information Flows for Critical Information Systems" in *Autonomic Computing: Concepts, Infrastructure, and Applications*, Manish Parashar and Salim Hariri (eds.), CRC Press, 2006.

Refereed Conference/Workshop Publications

1. James Horey, Jean-Charles Tournier, Patrick Widener, Arthur Maccabe and Ann Kilzer. "A Filesystem Interface for Sensor Networks". In *Proc. Fifth Workshop on Embedded Networked Sensors (HotEmNets)*, Charlottesville, Virginia: ACM, June, 2008. To appear.
2. Jiantao Kong, Ivan Ganev, Karsten Schwan and Patrick Widener. "CameraCast: Flexible Access to Remote Video Sensors". In *Proc. Fourteenth Annual Multimedia Computing and Networking Conference (MMCN'07)*, San Jose, California, January, 2007.
3. Ron A. Oldfield, Patrick Widener, Arthur B. Maccabe, Lee Ward and Todd Kordenbrock. "Efficient Data Movement for Lightweight I/O". In *Proc. 2006 Workshop on high-performance I/O techniques and deployment of Very-Large Scale I/O Systems (HiPerI/O 2006)*, Barcelona, Spain, September, 2006.
4. Ron A. Oldfield, Arthur B. Maccabe, Sarala Arunagiri, Todd Kordenbrock, Rolf Riesen, Lee Ward and Patrick Widener. "Lightweight I/O for Scientific Applications". In *Proc. 2006 IEEE Conference on Cluster Computing*, Barcelona, Spain, September, 2006.
5. Jiantao Kong, Karsten Schwan and Patrick Widener. "Protected Data Paths: Delivering Sensitive Data via Untrusted Proxies". In *Proc. 2006 International Conference on Privacy, Security and Trust (PST 2006)*, Ontario, Canada, October, 2006.
6. Vibhore Kumar, Zhongtang Cai, Brian F. Cooper, Greg Eisenhauer, Karsten Schwan, Mohamed Mansour, Balasubramanian Sheshasayee and Patrick Widener. "IFLOW: Resource-aware Overlays for Composing and Managing Distributed Information Flows". In *Proc. European Conference on Computer Systems (EuroSys)*, Leuven, Belgium, April, 2006.
7. Zhongtang Cai, Vibhore Kumar, Karsten Schwan, Brian F. Cooper, Greg Eisenhauer, Mohamed Mansour, Balasubramanian Seshasayee and Patrick Widener. "Implementing Diverse Messaging Models with Self-Managing Properties using IFLOW". In *Proc. 3rd IEEE International Conference on Autonomic Computing*, Dublin, Ireland, June, 2006.
8. Patrick Widener. "Reverb: Middleware Support for Distributed Application Forensics". In *Proc. IEEE Workshop on Challenges for Large Distributed Environments*, Research Triangle Park, North Carolina, July, 2005.
9. Balasubramanian Seshasayee, Karsten Schwan and Patrick Widener. "SOAP-binQ: High-Performance SOAP with Continuous Quality Management". In *Proc. 24th International Conference on Distributed Computing Systems (ICDCS 2004)*, Tokyo, Japan, March, 2004.
10. Yair Wiseman, Karsten Schwan and Patrick Widener. "Efficient End-to-End Data Exchange Using Configurable Compression". In *Proc. 24th International Conference on Distributed Computing Systems (ICDCS 2004)*, Tokyo, Japan, March, 2004.
11. Patrick Widener, Karsten Schwan and Fabian Bustamante. "Differential Data Protection in Dynamic Distributed Applications". In *Proc. Annual Computer Security Applications Conference*, Las Vegas, Nevada, December, 2003.
12. Fabian Bustamante, Patrick Widener and Karsten Schwan. "Scalable Directory Services Using Proactivity". In *Proc. Supercomputing 2002*, Baltimore, Maryland, November, 2002.

13. Patrick Widener, Greg Eisenhauer and Karsten Schwan. "Open Metadata Formats: Efficient XML-Based Communication for High-Performance Computing". In *Proc. Tenth International IEEE Symposium on High Performance Distributed Computing (HPDC-10)*, San Francisco, California, August, 2001.
14. Fabian Bustamante, Greg Eisenhauer, Patrick Widener, Karsten Schwan and Calton Pu. "Active Streams: An approach to adaptive distributed systems". In *Proc. 8th Workshop on Hot Topics in Operating Systems (HotOS-VIII)*, Elmau/Oberbayern, Germany, May, 2001.
15. Beth Plale, Patrick Widener and Karsten Schwan. "Taking the Step from Meta-Information to Communication Middleware in Computational Data Streams". In *Proc. IEEE Heterogeneous Computing Workshop (at IPDPS 2001)*, San Francisco, California, April, 2001.
16. Fabian Bustamante, Greg Eisenhauer, Karsten Schwan and Patrick Widener. "Efficient Wire Formats for High-Performance Computing". In *Proc. Supercomputing 2000*, Dallas, Texas, November, 2000.
17. Beth Plale, Greg Eisenhauer, Lynn K. Daley, Patrick Widener and Karsten Schwan. "Fast Heterogeneous Binary Data Interchange for Event-based Monitoring". In *International Conference on Parallel and Distributed Computing and Systems (PDCS-2000)*, Las Vegas, Nevada, November, 2000.
18. Charles W. Reynolds and Patrick Widener. "A Study of the Support of the Programming Language C++ for Generic Abstract Data Types". In *Proc. Fourth National Conference on Undergraduate Research*, Schenectady, New York, April, 1990.

Refereed Posters

1. Patrick M. Widener, Matthew Barrick, Jack Pullikottil, Patrick G. Bridges and Arthur B. Maccabe. "Metabots: A Framework for Out-of-Band Processing in Large-Scale Data Flows". In *Proc. 2007 International Conference on Grid Computing (Grid 2007)*, Austin, Texas, September, 2007.
2. James Horey, Jean-Charles Tournier, Patrick Widener and Arthur Maccabe. "Koseki: A Sensor Network Filesystem". In *Proc. 2007 Annual International Conference on Mobile Systems (MobiSys)*, San Juan, Puerto Rico, June, 2007.
3. Fabian Bustamante, Patrick Widener and Karsten Schwan. "A Case for Proactive Directory Services". In *Proc. 11th IEEE Symposium on High Performance Distributed Computing (HPDC-11)*, Edinburgh, Scotland, July, 2002.
4. Fabian Bustamante, Patrick Widener and Karsten Schwan. "The Case for Proactive Directory Services". In *Proc. Supercomputing 2001*, Denver, Colorado, November, 2001.
5. Fabian Bustamante, Greg Eisenhauer, Karsten Schwan and Patrick Widener. "Active Streams and the effects of stream specialization". In *Proc. Tenth IEEE International Symposium on High Performance Distributed Computing (HPDC-10)*, San Francisco, California, August, 2001.
6. Patrick Widener, Karsten Schwan and Greg Eisenhauer. "Open Metadata Formats: Efficient XML-Based Communication for Heterogeneous Distributed Systems". In *Proc. 21st IEEE International Conference on Distributed Computing Systems*, Phoenix (Mesa), Arizona, April, 2001.

Dissertation

1. Patrick Widener. *Dynamic Differential Data Protection for High-Performance and Pervasive Applications*. Ph.D. Dissertation, College of Computing, Georgia Institute of Technology, August, 2005.

Technical Reports and Non-Refereed Publications

1. Yair Wiseman, Karsten Schwan and Patrick Widener. 'Efficient end to end data exchange using configurable compression'. In *ACM SIGOPS Operating Systems Review* 39 (3), July, 2005.

2. James Horey, Jean-Charles Tournier, Patrick Widener, Arthur B. Maccabe and Ann Kilzer. *A Filesystem Interface for Sensor Networks*. Technical Report TR-CS-2008-01, University of New Mexico, March, 2008.
3. Patrick M. Widener, Matthew Wolf, Hasan Abbasi, Matthew Barrick, Jay Lofstead, Jack Pullikottil, Greg Eisenhauer, Ada Gavrilovska, Scott Klasky, Ron Oldfield, Patrick G. Bridges, Arthur B. Maccabe and Karsten Schwan. *Structured Streams: Data Services for Petascale Science Environments*. Technical Report TR-CS-2007-17, University of New Mexico, November, 2007.
4. Jean-Charles Tournier, Patrick G. Bridges, Arthur B. Maccabe, Patrick M. Widener, Zaid Abudayyeh, Ron Brightwell, Rolf Riesen and Trammel Hudson. 'Towards a Framework for Dedicated Operating System Development in High-End Computing Systems'. In *ACM SIGOPS Operating Systems Review* 40 (2), April, 2006, pp.16-21. Special Issue on Operating and Runtime Systems for High-End Computing Systems.
5. Karsten Schwan, Brian F. Cooper, Greg Eisenhauer, Ada Gavrilovska, Matt Wolf, Hasan Abbasi, Sandip Agarwala, Zhongtang Cai, Vibhore Kumar, Jay Lofstead, Mohamed Mansour, Balasubramanian Seshasayee and Patrick Widener. *Autonomic Information Flows*. Technical Report GIT-CERCS-05-22, Center for Experimental Research in Computer Systems, Georgia Institute of Technology, November, 2005.
6. Patrick Widener and Karsten Schwan. *Dynamic Differential Data Protection*. Technical Report GIT-CERCS-04-36, Center for Experimental Research in Computer Systems, Georgia Institute of Technology, 2004.
7. Patrick Widener, Fabian Bustamante and Karsten Schwan. *PDS - A Proactive Directory Service*. Technical Report GIT-CC-00-32, College of Computing, Georgia Institute of Technology, December, 2000.
8. Patrick Widener, Karsten Schwan and Greg Eisenhauer. *Open Metadata Formats: Efficient XML-Based Communication for Heterogeneous Distributed Systems*. Technical Report GIT-CC-00-21, College of Computing, Georgia Institute of Technology, December, 2000.
9. Patrick Widener, G. Scott Briercheck, Saad C. Himmich, Sally A. McKee, Ramesh V. Peri and William A. Wulf. *The WM Protection Mechanism*. Technical Report TR-92-28, Department of Computer Science, University of Virginia, August, 1992.
10. Ron A. Oldfield, Arthur B. Maccabe, Sarala Arunagiri, Todd Kordenbrock, Rolf Riesen, Lee Ward and Patrick Widener. *Lightweight I/O for scientific applications*. Technical Report SAND2006-3057, Sandia National Laboratories, May, 2006.

RESEARCH FUNDING

Co-PI, *Petascale Storage for High End Computing*, National Science Foundation, $400,000, 2006-2009.

RESEARCH PROJECTS

Petascale Storage for High-End Computing

with A. Maccabe, K. Schwan, P. Bridges, R. Oldfield, G. Eisenhauer, M. Wolf

Providing high-performance I/O for data-intensive high-end computing applications requires working with I/O systems at an unproductively low level of abstraction. This project seeks to provide higher-level I/O abstractions which make possible complex I/O tasks. The central abstraction of this approach is the *structured stream*. Structured streams provide a model for embedding application-specific functionality between application components. This functionality is applied to data as it moves through *I/O graphs*, which perform routing and in-band modification. The metadata necessary to connect these components is constructed out-of-band by autonomous *metabots*, moving the performance impact of metadata maintenance out of the "fast path" of high-end computing applications.

Lightweight I/O for High-Performance Computing
with A. Maccabe P. Bridges, R. Oldfield

Today's high-end massively parallel processing machines have thousands to tens of thousands of processors, with next-generation systems planned to have in excess of one hundred thousand processors. For systems of such scale, efficient I/O is a significant challenge that cannot be solved using traditional approaches. In particular, general purpose parallel file systems that limit applications to standard interfaces and access policies do not scale and are a performance bottleneck for many scientific applications. This project investigates the use of a "lightweight" approach to I/O that requires the application or I/O-library developer to extend a core set of critical I/O functionality with the minimum set of features and services required by its target applications. We argue that this approach allows the development of I/O libraries that are both scalable and secure.

Configurable Operating Systems for High-Performance Computing
with A. Maccabe, P. Bridges, R. Brightwell, R. Riesen

The use of general-purpose operating systems for high-performance computing applications (defined in terms of application and/or hardware scale) imposes unacceptable overhead on application performance. Dedicated operating systems address performance issues, but are difficult to build, maintain, and adapt. The ConfigOS project is developing a component-based framework which supports the development and adaptation of *a la carte* operating systems for high-performance computing. Such operating systems are customized to application requirements but remain easily configurable as those requirements (or other factors, such as hardware capabilities) change.

Dynamic Differential Data Protection for High-Performance and Pervasive Applications
with K. Schwan, G. Eisenhauer

A key concern among developers of extensible systems is the ability to provide adaptation approaches without sacrificing the level of security achievable with more static (and less adaptable) solutions. For example, where will adaptations execute? With what environment will they run? What level of access will they have to the existing system? Dynamic Differential Data Protection (D3P) has addressed these issues through the creation and evaluation of protection mechanisms for middleware infrastructures. D3P provides control over the data typing space for such middleware as well as abstractions provided by the middleware itself. D3P also provides a general and flexible extension/customization model for distributed applications based on publish/subscribe middleware. We are now exploring how D3P concepts can be applied to advanced software architectures for high-performance computing.

High-Performance Publish/Subscribe Communication
with K. Schwan, G. Eisenhauer

The EVPath communication infrastructure, along with PBIO, its companion data representation library, is the foundation for many research efforts at Georgia Tech. I continue to collaborate with researchers there on topics ranging from wire-formats for heterogeneous communication, to peer interaction paradigms (pull/push styles of communication), to required services and driving applications.

Proactive Directory Services
with F.E. Bustamante, K. Schwan

We argue that traditional passive client interfaces to directory services are not sufficient for the application environments enabled by grid and pervasive computing, where data are updated at high frequency. In particular, an *exclusively* passive interface not only hinders service scalability but also indirectly restricts the behavior of potential applications. Consequently, we have proposed a customizable active mode through which clients can subscribe to be notified of changes to data of interest. We have designed and implemented the Proactive Directory Service to test our ideas. PDS clients can dynamically tune the levels of detail and granularity of these notifications through filter functions instantiated at the server or at the object's owner, and by remotely tuning the functionality of those filters. We are currently building a next generation of these tools as part of the Petascale Storage project.

Filesystem-like Control Interfaces for Sensor Networks
with J. Horey, J.-C. Tournier, A. Maccabe

Currently, users of sensor network applications must adopt custom methods and interfaces to perform common management tasks. This makes monitoring, tasking, diagnosing and debugging sensor networks and their applications cumbersome; for example, users often cannot transfer skills learned for one application onto another. Our work provides standardized end-to-end (from user to sensor nodes) communication and control over a sensor network. A POSIX-style filesystem interface enables users to view and update data, organize groups of sensors, and retask sensor nodes. Sensor nodes appear as directories containing sensor and data files. Users are then able to use common command-line utilities to interact with the sensor network. We are currently testing the fidelity of our approach by applying management tools such as file system visualizers to our work.

Open Metadata Formats for High-Performance Communication 1999 - 2002
with K. Schwan, G. Eisenhauer

This project explored support for decoupling metadata and applications in high-performance communication environments. We designed and implemented the XML Metadata Integration Toolkit (XMIT), enabling application developers to exploit the generality and flexibility of XML for data description, while preserving the option of using a high-performance binary wire format for data exchange. Our solution proved superior to XDR- or "end-to-end" XML-based data transfer approaches, both in terms of data encoding size and and transfer bandwidth provided to applications. Research products from this project are now being used and improved as part of the Petascale Storage project.

PROFESSIONAL SERVICE

Program Committee Member, ACM International Conference on Computing Frontiers (2006).

Paper Referee, ACM Symposium on Operating Systems Principles (2004), IEEE International Symposium on Cluster Computing (2006), IEEE Symposium on High Performance Distributed Computing (2003-2005), IEEE International Parallel and Distributed Processing Symposium (2005), Journal of Grid Computing (2005), Annual Computer Security Applications Conference (2005), International Conference on Architecture of Computing Systems (2005).

Member of the Graduate Student Council, 1999-2003, College of Computing, Georgia Institute of Technology.

Member, Association for Computing Machinery and IEEE Computer Society

TEACHING EXPERIENCE

Instructor, Department of Computer Science, University of New Mexico Spring 2007
Prepared syllabus, reading list, class lectures, projects and exams for CS 481, "Digital Computer Operating Systems".

Instructor, College of Computing, Georgia Institute of Technology Summer 2002
Prepared syllabus, reading list, class lectures, projects and exams for CS 4210, "Advanced Operating Systems".

Guest Lecturer, College of Computing, Georgia Institute of Technology 2001 - 2005
Multiple lectures on varied material for CS 6210, "Advanced Operating Systems" graduate course.

Teaching Assistant, College of Computing, Georgia Institute of Technology 1998 - 1999

Instructor, Continuing Education Department, College of Computing, Georgia Institute of Technology 2001

OTHER PROFESSIONAL EXPERIENCE

Software Developer, Universal Systems, Inc., Chantilly VA	Mar 1998 - Aug 1998
Technical Lead / Specialist, Bell Atlantic Network Services, Inc., Silver Spring, MD	Sep 1996 - Mar 1998
Programmer/Analyst, American Management Systems, Fairfax VA	Sep 1992 - Aug 1996

REFERENCES

Arthur (Barney) Maccabe, Ph.D.
Chief Information Officer (Interim); Professor, Department of Computer Science
University of New Mexico
MSC 01-1130, Albuquerque, NM 87131
maccabe@cs.unm.edu, 505-277-8125

Karsten Schwan, Ph.D.
Professor, College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
schwan@cc.gatech.edu, 404-894-2589

Patrick Bridges, Ph.D.
Assistant Professor, Department of Computer Science
University of New Mexico
MSC 01-1130, Albuquerque, NM 87131
bridges@cs.unm.edu, 505-277-3032

Greg Eisenhauer, Ph.D.
Research Scientist, College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
eisen@cc.gatech.edu, 404-894-3227

My research interests lie in experimental systems. My preferred method of conducting research is building proof-of-concept systems, scaling them up through iterative design-implement-test cycles, and examining them to characterize their strengths and weaknesses. I am interested in applying this approach to research problems in operating systems, parallel and distributed computing, and security; my preferred application domains include pervasive/ubiquitous computing and high-performance computing. A unifying theme in my research is the management of data and metadata in such systems. More precisely, I am interested in exploring how to best describe, manage, and protect data efficiently and effectively at different levels in the application/operating system/hardware stack.

Metadata description and representation is becoming increasingly important. In the high-performance computing domain, high-end computational experiments pose several challenges. The immense volume of data (DOE researchers routinely aim at petascale data environments) must be monitored, moved, and manipulated; novel approaches to the construction and dissemination of metadata can relieve real-time I/O bandwidth constraints while supporting existing codes. Heterogeneity of application components (codes, data, machines) is another area where intelligent, integrated metadata management can provide immediate benefits. Finally, collaboration between researchers requires common vocabularies and mechanisms for resource description, exclusion/mediation, and interoperability. Pervasive computing applications composed of many distributed sensing and computing devices share related concerns; how will individuals assert control over the data they comprise (personal profiles, for example) and the data they generate (real- and virtual-world interactions such as passing through the viewing angle of a camera)?

Viewing the issues from the other direction, data privacy is an issue in both pervasive and high-performance environments: scientists may wish to exert control over sensitive data they produce while still cooperating with remote researchers, and sensor data such as images may need to be "scrubbed" in some application-specific manner before being made publicly available. Likewise data provenance; as both pervasive and high-performance applications begin to interact by using the same data, the quality and history of that data becomes critical. My research investigates abstractions and mechanisms intended to give application developers traction in addressing these concerns.

My earlier research has produced published results on portable-yet-efficient representation schemes for binary data formats, novel directory service abstractions for high-performance applications, and a mechanism providing differential data protection for applications in domains including pervasive and high-performance computing. These efforts are still active and have served as the basis for new research vectors for myself and my collaborators. Currently, I am investigating software architectures for providing efficient storage in petascale scientific applications, and techniques for building custom OS kernels for high-performance applications.

Petascale Storage — Large-scale HPC applications face significant I/O challenges. This is particularly apparent for emerging classes of data-intensive HPC applications where I/O may require extensive data staging, reorganization, or conversion. Seismic exploration data processing, for example, requires multiple stages of data filtering and transformation. Each of these stages requires the loading and interpretation of hundreds of terabytes of data. These actions require not only access to massively parallel computational facilities, but also data movement at near-physical disk speeds in order to keep pace with output data production and continual checkpointing of application state. Finally, the design and successful deployment of large data-intensive applications is highly dependent on the existence of metadata (of the right kinds and in the right places); however, current I/O solutions perform the generation of such metadata in-line with application processing, reducing effective I/O bandwidth and stealing scarce processor cycles.

This project seeks to provide high-level I/O abstractions which make possible complex I/O tasks. The central abstraction of this approach is the *I/O Graph*. I/O Graphs provide a model for embedding application-specific functionality between application components. The metadata necessary to connect these components is constructed out-of-band by autonomous *metabots*, moving the performance impact of metadata maintenance out of the "fast path" of high-end computing applications. I am a co-PI on a National Science Foundation grant supporting this research.

A foundation for this project is the Lightweight File System (LWFS), a collaboration between the University of New Mexico and Sandia National Laboratories. LWFS investigates a "lightweight" approach to I/O, in which a core set of critical I/O functionality (access control and low-level storage) is all that is provided. Any additional storage semantics or functionality must be provided by application developers as libraries. The lightweight approach allows applications to manage their own I/O while choosing the storage semantics (and associated tradeoffs) they need. We argue that this approach, by making the choices of I/O-related overheads explicit, allows the development of layered I/O subsystems which are both scalable and secure. Our ongoing work on LWFS includes the design and development of lightweight transactional semantics and mechanisms, as well as application compatibility libraries.

Configurable OS Kernels — The use of general-purpose operating systems for high-performance computing applications imposes unacceptable overhead on application performance. Dedicated operating systems address performance issues, but are difficult to build, maintain, and adapt. The ConfigOS project is developing a component-based framework which supports the development and adaptation of *a la carte* operating systems for high-performance computing. Such operating systems are customized to application requirements but remain easily configurable as those requirements (or other factors, such as hardware capabilities) change. We are defining a tree of components and associated deployment/composition infrastructure which can be used for this purpose, specializing areas such as memory management, scheduling and multiprogramming.

Dynamic Differential Data Protection — My dissertation research explored data protection challenges found in a continually-changing, dynamic world of open and connected applications, services and systems. In this world, businesses must share information with partners while protecting proprietary data. Scientists must collaborate by exchanging data sets and coordinating access to shared research equipment without jeopardizing their personally publishable work. Data streams from remote sensors must be managed in order to satisfy a diverse set of consumers, from end-user applications such as surveillance to computational tasks like face recognition.

To address this class of application problems, I have proposed the concept of *dynamic differential data protection* (D3P). D3P is a set of abstractions providing a data access model which is based on familiar object- and capability-based protection models. This data access model incorporates application-defined security policies and metadata in order to provide *differential* protection, where the granularity of access control changes as needed by applications. D3P incorporates adaptive behavior by providing a model for dynamically-deployed, application-specific code to operate on data "in-flight". For example, a middleware system that provides typed data interchange facilities implicitly supports a type system. Metadata associated with such a type system can provide the basis for flexible and decoupled definition and evolution of data. Also possible are applications that differentiate services based on the structure and/or content of typed data. Such differentiation can be employed to enforce application-specific security policies or to enable service provision on resource-limited application platforms. My dissertation explores possible candidates for such services, details sample implementations, and evaluates them in the contexts of high-performance and pervasive computing.

Future Directions — I am particularly interested in issues surrounding data provenance, and I would like to explore more fully how provenance information can be introduced and used in my chosen application domains. To do this effectively will require close collaboration with the domain researchers who are creating and using the data in question, a process already underway in the Petascale I/O project. To do this efficiently will require investigation of novel solutions in the construction of metadata, many of which will be realized in the metabot framework. I must look upward, into service specification languages and code generation tools and learn how to better represent application needs in metadata. I must also look downward into hypervisor, operating system, and hardware layers, to integrate with existing data typing, code mobility, and protection mechanisms as seamlessly as possible. I will apply the lessons learned in high-performance application metadata design to pervasive/ubiquitous applications, where data provenance issues are also paramount. I also would like to broaden my research in these areas to applications in the healthcare domain, where questions of data protection and application interaction will only become more important as our nation's population ages.

I actively seek opportunities for collaboration with computer scientists in other sub-disciplines, I have pursued collaborative research throughout my academic career and have found it both rewarding and productive. As examples, my interests in metadata specification and manipulation have ties to the database and software engineering communities, and I have expanded my data protection research into application forensics (addressing topics such as intrusion detection, execution-path tracing, and configuration change auditing).

My experience as a professional technologist and software developer has produced in me a strong bias toward things that work, rather that those that are simply aesthetically appealing. This attitude has two different consequences for a teacher. First is that understanding a technology comes most directly from hands-on experience, learning not only what it does but why it does what it does. The important corollary here is that experimentation is not only desirable — it is necessary if true understanding is to be achieved. Second is that things that work but are not elegant, or those that do not work, present important opportunities for iterative design and evaluation. This can inform the coverage of historical material as well as hands-on projects. Perhaps most important here is the opportunity to involve others in the iterative process through group projects, term-long projects that build on themselves, and class interaction through such activities as code reviews.

My experience in teaching Operating Systems to undergraduates at both the University of New Mexico and Georgia Tech has greatly influenced my thinking. In both instances, I was responsible for preparing the class syllabus and reading list; delivering all lectures; devising all homeworks, programming projects and exams; and supervising teaching assistants or graders. My students were junior- and senior-level undergraduates, as well as graduate students from other disciplines. I find it most rewarding to view that class material as an evolution of concepts, and so organized the class content to emphasize the connections between those concepts. For example, my lectures presented the development of Remote Procedure Calls as a logical consequence of the adoption of both programming abstraction and networked computing. I then demonstrated how these basic ideas persist today in the current realization of RPCs in technologies such as SOAP and .NET.

I believe that one of the most effective ways of demonstrating the relevance of classroom topics is to show how they translate into improved tools for use by computer scientists (or, increasingly, by non-computer scientists). Stressing the historical roots of "new" technologies demystifies them and encourages students to begin to think critically about them. This is the first step toward experimenting with technology, which is in turn the first step toward learning it.

Communication and interaction are essential at all points. At least as important as the production of some classroom artifact (such as a programming project) is the ability to communicate about that artifact: what does it do? why? does it fulfill its requirements? how? It is the responsibility of the instructor to provide examples and feedback to guide students in developing these communication skills — they are not found in the textbooks.

One of the most rewarding things about learning is that non-trivial effort is required. If the effort needed to master a skill or understand a concept is acknowledged up front, the sense of accomplishment and self-confidence realized by the student after going through the process is that much greater. This is not to say that teachers cannot make learning fun, that the two are mutually exclusive. However, I do believe that a recognition that work is involved pays off in the end.

It is much easier to teach a motivated student. I believe it is easiest and most beneficial to motivate students by helping them understand the *relevance* of their classwork. Many undergraduates approach classroom topics with a single question: "How will this help me get a job?" This kind of attitude can easily result in an adversarial relationship between instructor and student, especially if the instructor chooses to rely upon the curricular necessity (i.e. "you HAVE to take this class") of the class to overcome any such reservations on the part of students. In my experience, an approach that emphasizes that there is work involved, but also emphasizes the relevance of that work, provides a framework where students are willing to invest time and effort in their own education. This in turn provides the instructor with several and varied opportunities for teaching *actively*: frequently interacting with students, encouraging critical thinking and analysis by students, and making the class a participatory experience rather than a passive one.

I have also had numerous opportunities to act as a mentor, both in industry as the leader of a team of developers and in academia with graduate students. I feel very strongly that it is impossible to be a good academic researcher without also being a good mentor. The scope of work necessary to perform quality research demands the assistance of quality graduate students. Retaining those students, ensuring that they make progress in their own careers, and helping them deal with the work/life issues endemic to graduate school requires a strong mentoring relationship. Teaching also provides an opportunity for mentoring students — while presenting research papers as a lecturer, I have made a point of discussing what makes a research paper good or not and how quality research is conducted and communicated.

My research expertise is in operating systems, distributed computing, computer security, and high-performance and ubiquitous computing applications. I am interested in teaching courses on these topics at both the graduate and undergraduate levels. I am also capable of teaching undergraduate classes in several other areas (such as networking, programming languages, and software engineering). I would also welcome the opportunity to teach seminar-style courses that explore the state-of-the-art in current research.

March, 2008