ITR/SY: A Distributed Programming Infrastructure for Integrating Smart Sensors

NSF Program CCR-0121638

PIs: Umakishore Ramachandran
Kenneth Mackenzie
Steve DeWeerth
Irfan Essa
Thad Starner

College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
Phone: (404) 894-5136
FAX: (404) 385-6506
e-mail: rama@cc.gatech.edu
WWW URL: http://www.cc.gatech.edu/~rama

Annual Report via Fastlane May 31, 2005

1 Activities and Findings

1.1 Research and Education

The proposed research is integrating sensing hardware, embedded processing and distributed system support to build a seamless programming infrastructure for ubiquitous presence applications. Fundamental invention and integration of techniques spanning programming idioms and runtime systems for distributed sensors, and building blocks for embedded processing are expected as the primary intellectual contributions of the proposed research. Interfacing these technologies to emerging applications on the one end and novel off-the-shelf sensors at the other end are secondary goals of the proposed research.

In our exploration of novel techniques to support integration of smart sensors into a distributed programming infrastructure, we are conducting focused studies into 1) efficient resource management for streaming applications in wireless sensor networks, cluster computing, and grid computing environments; 2) universal middleware for integrating multiple protocol families; 3) efficient biometric authentication over wireless networks; 4) efficient and robust code dissemination in multihop wireless networks; 5) leveraging web service technologies to support stream based distributed computing applications; and 6) using software (network) caching rather than hardware caching for energy efficient embedded sensor processing. We are also developing specific driving pervasive computing applications to assist in design and evaluation of our distributed infrastructure components. Further, we are developing a sensor lab as a testbed for integrating sensor technologies into pervasive computing applications. This subsection details the research accomplishments this past year (June 2004-May 2005).

1.1.1 Distributed Systems Technologies

Simulation-Based Study of Wireless Sensor Network Middleware Future wireless sensor networks (WSN) will transport high-bandwidth, low-latency streaming data, and will host sophisticated processing, such as image fusion and object tracking, in-network on sensor network nodes. Node computation capabilities will continue to increase; however, it is likely that energy will continue to remain a constrained resource.

We extend our previous DFuse [9] work to further evaluate performance of our infrastructure for these sensor network applications. DFuse provides capabilities for higher-level in-network processing while minimizing energy drain on the network, and consists of a data fusion API and a distributed algorithm for energy-aware role assignment. The fusion API enables an application to be specified as a coarse-grained dataflow graph, and eases application development and deployment. The role assignment algorithm maps the graph onto the network, and optimally adapts the mapping at run-time using role migration.

Our previous work considered wireless communication to be the only significant cause of energy loss in WSN. However, hosting complex processing on WSN nodes incurs additional processing energy and latency costs that impact network lifetime and application performance. There is a need for a WSN planning framework to explore energy saving and application performance trade-offs for models of future sensor networks that account for processing costs in addition to communication costs.

In this work [21, 22], we present a simulation framework to analyze the interplay between resource requirements for compute- and communication-intensive in-network processing for streaming applications. We simulate a surveillance application workload with middleware capabilities for data fusion, adaptive policy-driven migration of data fusion computation across network nodes, and prefetching of streaming data inputs for fusion processing. Our study sheds light on application figures of merit such as latency, throughput, and lifetime with respect to migration policy and node CPU and radio characteristics.

We show that 1) increasing radio bandwidth while using prefetching may not improve latency or throughput for compute-intensive workloads and may actually decrease network productivity; 2) migration can significantly extend WSN lifetime for topologies and applications two orders of magnitude larger than previous studies; 3) imprudent choice of migration policy may degrade performance for specific workloads; and 4) adaptive computation migration enables lifetime extension with increase in the number of network nodes, using a migration policy scalability study for a small application.

We are currently performing simulator extensions and experiments to 1) further analyze the scalability of the DFuse role assignment algorithm for large applications in large networks, 2) incorporate more realistic MAC and routing layer models, and 3) to study dynamic application workloads and mechanisms to further improve WSN efficiency under these dynamic workloads.

Adaptive Resource Utilization Mechanisms for Streaming Cluster Applications A large emerging class of interactive multimedia streaming applications that are highly parallel can be represented as a coarse-grain, pipelined, data-flow graph. One common characteristic of these applications is their use of current data: A task would obtain the latest data from preceding stages, skipping over older data items if necessary to perform its computation. When parallelized, such applications waste resources because they process and keep data in memory that is eventually dropped from the application pipeline. To overcome this problem, we have designed and implemented an Adaptive Resource Utilization (ARU) mechanism that uses feedback to dynamically adjusts the resources each task running thread utilizes so as to minimize wasted resource use by the entire application. A color-based people tracker application is used to explore the performance benefits of the proposed mechanism. In our recent work [10], we show that ARU reduces the application's memory footprint by two-thirds compared to our previously published results, while also improving latency and throughput of the application.

Streamline: A Scheduling Heuristic for Streaming Applications on the Grid Streaming applications such as video-based surveillance, habitat monitoring, and emergency response are good candidates for executing on high-performance computing (HPC) resources, due to their high computation and communication needs. Such an application can be represented as a coarse-grain dataflow graph, each node corresponding to a stage of the pipeline of transformations that may be applied to the data as it continuously streams through. Mapping such applications to HPC resources has to be sensitive to the computation and communication needs of each stage of the pipeline to ensure QoS criteria such as latency and throughput. Due to the dynamic nature of such applications, they are ideal candidates for using ambient HPC resources made available via the grid. Since grid has evolved out of traditional high-performance computing, the tools available, especially for scheduling, tend to be more appropriate for batch-oriented applications. We have developed a scheduler, called *Streamline* [2], that takes into account dynamic nature of the grid and runs periodically to adapt scheduling decisions using (a) application requirements (per-stage computation and communication needs), (b) application constraints (such as co-location of stages on the same node), and (c) current resource (processing and bandwidth) availability. The placement generated by Streamline heuristic can be used by other services (such as task migration) in a grid environment in order to dynamically adapt the performance of a streaming application.

We have designed our scheduling heuristic over the existing grid framework, using Globus Toolkit. In our experimental study, we compare Streamline with the Optimal placement (for small dataflow graphs) and approximation algorithms using Simulated Annealing (for both small and large dataflow graphs). Most of the existing grid schedulers such as Condor, Legion and Nimrod-G focus on allocating resources for batch-oriented applications. We analyze how such existing schedulers in grid can be enhanced to support streaming applications using Condor as example. Condor is a well studied resource allocator for grid and it uses DAGMan for task graph based applications. DAGMan is designed for batch jobs with control-flow dependencies and ensures that jobs are submitted in proper order, whereas different stages of a streaming application work concurrently on a snapshot of data. Thus, we have extended Condor to meet the particular streaming requirements, resulting in a baseline scheduler called *E-Condor*.

We compared the performance of Streamline with Optimal, Simulated Annealing, and E-Condor for “kernels” of streaming applications. The results show that our heuristic performs close to Optimal and Simulated Annealing, and is better than E-Condor by nearly an order of magnitude when there is non-uniform CPU resource availability, and by a factor of four when there is non-uniform communication resource availability. We have considered two variants of Simulated Annealing algorithm with different execution times and observe that neighbor-selection and annealing schedule in a Simulated Annealing algorithm have a relatively higher impact on the performance of generated schedule for communication-intensive kernels than for computation intensive kernels. We have also conducted scalability studies and demonstrate the scalability of our heuristic for handling large-scale streaming applications. The results show that Streamline is more effective than E-Condor in handling large dataflow graphs, and performs close to Simulated Annealing algorithms, with smaller scheduling time.

While Streamline does the placement for such streaming applications well, it is clear that the application dynamics may result in the computation and communication characteristics of the application changing over time. Perhaps even the dataflow graph of the application could change over time with the addition and deletion of new stages to the pipeline. If the application characteristics that are profiled and used in the placement are considered the “typical” (for e.g. mean) values for the respective stages, then the mapping given by Streamline would result in an acceptable level of performance despite this dynamism. Nevertheless, it is important to consider the impact of the application dynamism and the consequent adaptation of the scheduling heuristic. Such an adaptive scheduling heuristic is part of our future work. We are also looking at using the placement generated by Streamline heuristic in other grid services to dynamically adapt the performance of an application.

MediaBroker MediaBroker is a distributed framework designed to support pervasive computing applications. Key contributions of MediaBroker are efficient and scalable data transport, data stream registration and discovery, an extensible system for data type description, and type-aware data transport that is capable of dynamically transforming data en route from source to sinks. Specifically, the architecture consists of a transport engine and peripheral clients and addresses issues in scalability, data sharing, data transformation and platform heterogeneity.

Continuing from our work reported in the 2003-2004 annual report, we have coupled MediaBroker with EventWeb, a continuous query, internet-scale driving application we are also developing that is described below. Details of the MediaBroker architecture, implementation, and this concrete application example are published in a recent journal extension [20] of our initial MediaBroker study [13]. Experimental study shows reasonable performance for selected streaming media-intensive applications. For example, relative to baseline TCP performance, MediaBroker incurs under 11% latency overhead and achieves roughly 80% of the TCP throughput when streaming items larger than 100 KB across our infrastructure. The EventWeb application demonstrates the utility and graceful scaling of MediaBroker for supporting pervasive computing applications.

MediaBroker continues to gain interest from undergraduate researchers, and has served as an excellent research vehicle for exploring the design and performance analysis of pervasive computing middleware in the context of real-world applications. Our current MediaBroker research efforts focus on integrating our efficient stream processing scheduling technologies into a seamless API spanning WSN and HPC environments.

uMiddle and Novel User Interfaces Dr. Jin Nakazawa, a visiting research scholar from Keio University, is associated with Professor Ramachandran's research group. He is developing a universal middleware called *uMiddle* [14] that allows devices from different protocol families (UPnP, Bluetooth, etc.) to be used without any special configuration in a pervasive computing application. On top of this middleware, Dr. Nakazawa has developed a novel user interface that combines physical artifacts (such as maps) and off-the-shelf digital devices such as a mouse to explore multimedia data across space and time.

Cyberforaging Research Scientist Phil Hutto is leading a group of undergraduates working in the sensor lab on a *cyberforaging* project. The basic idea is to use the ambient computing infrastructure (such as high performance clusters) from mobile devices for compute intensive tasks. This activity is just starting this summer, and we will have more to report on it in our next annual report.

Remote Authentication over Wireless Networks Remote authentication over a long range wireless network using large signature keys such as biometric samples (fingerprint, retinal scans etc.) is soon going to become an integral feature of various kinds of transactions. In the domain of mobile and ad hoc networking, this will become even more relevant due to the intrinsic dynamism in the applications. Because of the large size of the authentication keys, and continual need for authentication, considerable power and bandwidth are consumed by such a process. Authentication being only a background process supporting other transactions, should not take away too much of resources, especially bandwidth and power that are quite critical for small mobile devices.

To enable such applications, we developed LAWN, a light-weight authentication protocol for wireless networks that trades computation for communication and can be tuned for any desired security guarantee. For an authentication token of length n, LAWN prepares a small sketch of length $O(\log n)$ (adding very low computational overhead), and transmits the sketch over the network. Under a reasonable energy consumption model, we show that this technique results in 70% to 80% saving in power for long-range wireless applications.

We had the LAWN protocol developed before May 2004. Since then, we have carried out multiple tests to validate the viability of the protocol. It was tested on iPAQ test bed for performance in terms of bandwidth consumption and computational overheads. We also power-profiled the energy consumption of this protocol through a power profiler called JouleTrack (developed at MIT). We have published our preliminary results [18] and more recent results will appear this July at IEEE NCA 2005 [17].

Code Dissemination in Wireless Ad Hoc Networks Code dissemination in wireless ad hoc network is an important aspect of network deployment. Once deployed, the network nodes may still need software updates to keep up with the changing application demand, thus making wireless broadcast an important aspect of any wireless network deployment. We developed FBcast, a new broadcast protocol based on the principles of modern erasure codes. FBcast provides high reliability and data confidentiality: often considered critical for disseminating codes. We extend FBcast with the idea of repeaters to improve reliable coverage for multi-hop and large deployments. The scope of FBcast includes general data broadcasting in unreliable WiFi, wireless mesh networks and sensor networks. Simulation results on TOSSIM (a simulator for TinyOS running on Intel motes) show that FBcast offers higher reliability with lower number of retransmissions than traditional broadcasts. We have published a preliminary FBcast technical report [7] and initial results [8]. We have communicated details of the protocol with extended results, currently under review [6].

State Management in Web Services This is a continuation of the work reported in the 2003-04 annual report to provide certain applications wishing to use the web service paradigm with rapid, robust state maintenance. Based on the loosely-coupled D-Stampede.NET infrastructure built last year, we design and implement the closely-coupled cluster-based Stampede.NET on Microsoft .NET platform. The former infrastructure is good for low-bandwidth, high-latency Internet connection while the latter one is good for a high-bandwidth, low-latency cluster of machines. We also re-design the C library to object-oriented structure in C# and develop an enhancement of original garbage collection algorithm in Stampede. The preliminary performance results show the effectiveness of the new design and the enhancement. We are preparing results for conference communication currently. Two undergraduates, Jason Whitehurst and J. D. Courtoy, have recently joined this focused research effort.

Software Caching for Embedded Sensor Processing This is continuation of the work we reported in the 2001-02, 2002-03 and 2003-04 annual reports to reconcile programmability with cost for embedded processing associated with sensors in a distributed sensing infrastructure.

The reconstruction of the instruction cache based SoftCache system now supports basic-block level division of arbitrary ELF binary programs. The client-server system is now designed to run inside the U-Boot framework on the client, on top of an Intel Sitsang 400 XScale PDA platform. The server remains an x86-hosted dynamic binary translator that parses and rewrites ARM binaries for the client needs.

Additional work in reconstruction of control- and data-flow graphs from the arbitrary ELF binaries without source code or debugging symbols has also been completed. This now enables the isolation of data addresses inside the text stream, as well as providing support for data cache support in the SoftCache framework. The key insight required to allow a feasible implementation of data caching requires the UTI/MTI behavior classification [4]. By exploiting this novel classification and by adding a limited instruction-manipulated tag array, the SoftCache is a viable and complete solution for ubiquitous embedded network systems. We have submitted additional results that are currently under review [15].

1.1.2 Applications

EventWeb While the volume and diversity of multimedia permeating the world around us increases, our chances of making sense of the available information do the opposite. This environment poses a number of challenges which include achieving scalability while accessing all the available media, live and archived, inferring its context, and delivering media to all interested parties with its context attached. We envision [12] a solution to this set of challenges in a novel system architecture. As a starting point, however, we select a previously described framework, EventWeb, suitable for annotating raw multimedia data with context meaningful to end users. We then map it onto a distributed architecture capable of correlating, analyzing, and transporting the volumes of data characteristic of the problem space. This paper first presents the requirements for our architecture, then discusses this architecture in detail, and outlines our preliminary implementation efforts. More recently, we coupled the EventWeb application with our MediaBroker stream discovery and transformation engine to demonstrate utility of MediaBroker to support large scale pervasive computing applications [20].

Family Intercom We use the Family Intercom as another application for evaluating the integration of our systems technologies for smart sensors with driving applications. The Family Intercom provides communication assistance within a given location such as a home. Primarily, it is a control system for a collection of microphone and speaker combinations that can track the movement of people inside their location and establish dynamic communication links accordingly such that a conversation can be maintained as people freely move within their home. In addition to the academic significance, the system has potential real-world applications such as assisting the elderly within their home. Our focus is both on qualitative (integration experience) and quantitative (connection management latencies) performance analysis of the integrated application. MediaBroker's ease of use and of integration is currently being tested by integrating it with off the shelf components such as Universal Plug and Play (UPnP) and the Lightweight Directory Access Protocol (LDAP) systems. This study has been worked on by several undergraduates, including REU-funded Vladimir Urazov and Jose Caban. Experience from this study is feeding back into the next generation design of MediaBroker.

1.1.3 Sensor Technologies

We reported on the development of a sensor lab in our annual report last year. We are continuing to develop this lab and use it as part of our experimental testbed. Starting in summer of 2004, Research Scientist Phillip Hutto has been leading a group of undergraduates in the design and development of a Java-based Sensor Toolkit (*JSTk*) [5]. The system was designed to provide a comprehensive middleware framework for tight integration and coordination of sensor and actuator resources.

The JSTk infrastructure provides high-level, network-aware, user-space "device drivers" that control and mediate their associated devices. Devices differ widely in their characteristics and capacities so a common core API (interface) will provide an intersection of capabilities with additional device-specific interfaces for unique characteristics. Thus, the infrastructure resembles three OS (Linux) abstractions: the device abstraction, the virtual filesystem interface, and the networking interface. It is layered, like the networking interface, with basic "low-level" capabilities used to implement higher, more powerful abstractions (like persistence, streaming, eventing, etc.).

1.2 Training and Development

We continue to attract bright and interested graduates and undergraduates to research projects in our group. Undergraduate participation in research within the College is facilitated by the excellent UROC program (`www.cc.gatech.edu/program/uroc`), coordinated by Professor Amy Bruckman. A variety of institute-wide programs are also available (`www.undergraduateresearch.gatech.edu`) including a special fund sponsored by the president of Georgia Tech (PURA) and several NSF-sponsored projects. We were pleased to support 17 undergraduates on our ITR-related projects during Summer/Fall 2004 and Spring/Summer 2005. They were: James Kim (SensorLab, MediaBroker and EventWeb), J. D. Courtoy (Stampede.NET), Jason Whitehurst (Stampede.NET), Vladimir Urazov (MediaBroker and Family Intercom), Jose Caban (MediaBroker and Family Intercom), Simon Chen (SensorLab), Ryan Graciano (SensorLab), Rex Sheridan (SensorLab), Robert Thomas (SensorLab), Thomas Shanks (SensorLab, uMiddle, Cyberforaging), Kevin Adkisson (Stampede.NET demo), Seth Horrigan (Cyberforaging), Paolo Mentonelli (uMiddle), Ricky Pattillo (Cyberforaging), Bill Phillips (uMiddle), Nate Rivard (MediaBroker) and Eric Sample (MediaBroker). For details of the PURA program, along with a list of recipients, see the website.

Many of the ongoing ITR-related projects are partially staffed by students working in the context of the Systems Hackfest. This is a group of undergraduates who participate in various research projects for pay, course credit, or just for fun. Hackfest is supervised by Research Scientist Phil Hutto and runs throughout the year. Summer sessions are most productive and have recently involved 6-10 students. Students meet briefly in a weekly session to report progress and plan milestones for the coming week. The group meeting allows cross-fertilization of project ideas and helps to educate the students. In addition, it provides an opportunity for group brainstorming on design and debugging issues. Weekly project meetings are focused on specific research tasks and often involve relevant faculty, grad students and staff.

During the last year undergraduates have participated in the following projects: SensorLab, MediaBroker, Stampede.NET, Family Intercom, EventWeb, uMiddle, and Cyberforaging. We are also pleased by the number of undergraduates in our group who continue on to graduate study both here at Georgia Tech and at other top schools.

We believe the Hackfest is an excellent opportunity for initiating undergraduates into the form and substance of academic research. In addition, the size and maturity of the inter-related research efforts provides a fertile matrix for varied interactions and training. Each group – undergraduates, Masters students, PhD students, research scientists and senior faculty – have regular opportunities for cross-group interactions. For example, undergraduates can look to senior faculty for vision and research goals, to research scientists for design advice, to graduate students for technical assistance and literature questions, and to each other for day to day camaraderie.

1.3 Outreach

Professor Ramachandran along with Santosh Pande, were instrumental in organizing the second Georgia Tech Symposium on Computing Systems and Networking Technologies, held in February 2005 on the Georgia Tech campus. This two-day event exposed students, particularly graduates, to the processes and forces that influence innovation in the commercial world. Attendees gained insight into how cutting-edge technologies transform from ideas into products, and had the opportunity to meet with technology researchers and innovators. Other Symposium highlights included a panel that examined career paths after graduate school ranging from start-ups to academia. The symposium also provided an opportunity for graduate students to display their work to the invited industry participants via a poster session.

During the summer of 2004 we have established comprehensive websites for all of our related projects and activities including personnel and publications. These websites offer extended abstract-style overviews of each project and discuss work in progress. We believe this effort will contribute to our research group's visibility. We continue to keep the content, especially the publication lists, current on these websites.

Through the auspices of the Center for Experimental Research in Computer Systems (CERCS) we continue to invite and host key individuals from academia and commercial research labs engaged in complementary research. Recent visitors include: Prof. Krithi Ramamritham (IIT Bombay), Kevin Kloker and Phil May (Motorola Labs), Brian Savory (Executive Director, SLR) and Ron Hutchins (CTO, Georgia Tech) in Summer 2004; Hermann Haertig (Technische Universitaet Dresden), Dr. Ajay Mohindra (IBM T. J. Watson), Dr. Roy Ju (Intel Corp.), Kevin Chang (University of Illinois, Urbana-Champaign), Saurabh Bagchi (School of Electrical and Computer Eng., Purdue University), Dr. Gopal Pingali (IBM T. J. Watson), Dr. Joshua Goodman and Dr. Yi Min Wang (Microsoft Research), Dr. Larry Huston (Intel Research Lab, Pittsburgh) and Dr. Manish Parashar (Rutgers University) in Fall 2004; Dr. Nageswara S. V. Rao (Oak Ridge National Laboratory), Prof. Gregory R. Andrews (U. of Arizona, former Division Director, Computer and Network Systems, NSF), Warren Matthews (OIT, Georgia Tech), KC Claffy (CAIDA), Prof. Haesun Park (NSF and Univ. of Minnesota), Prof. Rajeev Barua (UMD, College Park), Dr. Gautam Shroff (VP Technology Program, Tata Consultancy Services), Prof. Elisa Bertino (CERIAS and Purdue University), Prof. Sol Shatz (UI Chicago), Dr. Jeff Z. Pan (School of Computer Science, U. Manchester) and Prof. David August (Department of Computer Science, Princeton University) in Spring 2005; and Jason Nieh (Columbia University); Anthony Young-Garner (Software Engineer, IBM); Mark Weitzel (Architect, IBM Software Group, Autonomic Computing) and Chen Ding (University of Rochester) in Summer 2005.

The annual CERCS NSF/IUCRC Workshop on Experimental Research in Computer Systems was held in October 2004 and gave an opportunity for our group to interact with a distinguished list of advisory board members such as John Chilenski (Boeing), Phil Bernstein (Microsoft), Dean Compton (Delta Technology), Felipe Cabrera (Microsoft), Richard Friedrich (HP Labs), Thomas Zacharia (ORNL), Alan Ganek (IBM), Jaynarayan Lala (Raytheon), Bryant Bigbee (Intel), Daniel Reed (UNC at Chapel Hill), Raj Yavatkar (Intel), Dennis Gannon (Indiana University), Terence Critchlow (Lawrence Livermore National Lab), and Mathai Joseph (TATA Research Development and Design Center).

1.3.1 Technology Transfer

The Stampede programming systems developed under the auspices of this ITR project is being used as the basis for a dynamic cluster scheduling framework for use in technical applications at the Federal Reserve Bank - Atlanta. Two PhD students, Namgeun Jeong and Hasnain Mandviwala, are working at FRB-Atlanta as year-round interns helping the economists at FRB to run their economic forecasting applications on high-performance clusters using this dynamic scheduling framework.

The Stampede programming model also served as an inspiration for the T-Streams programming model that Dr. Kath Knobe developed at HP Cambridge Research Lab. T-Streams is being used for supporting scientific applications on high performance clusters.

1.3.2 Visitors

The ITR-funded project has attracted visitors who spend extended periods of time working with Professor Ramachandran's research group.

- Professor HeonChang Yu, Korea University, Seoul, S. Korea was a sabbatical visitor from Jan 2004–Jan 2005. He worked on meta-scheduling issues related to grid computing during his stay at Georgia Tech.
- Dr. Jin Nakazawa, Keio University, Japan, is currently visiting the group since Oct 2004. He is interested in pervasive computing and is contributing to developing universal middleware and novel UI combining physical artifacts and digital devices.
- Dr. Kath Knobe, HP Cambridge Research Lab, spends shorts bursts of time with the research group. She made two such visits in Fall 2004 and Spring 2005. She is slated to spend a month this summer working with the research group.

2 Publications and Products

2.1 Publications

See the references at the end of this document for publications that appeared in the period that covers this annual report.

2.2 Web Site

Please visit the project web site at www.cc.gatech.edu/~rama/ubiq-presence/

3 Contributions

The activities we are currently undertaking have resulted in a significant number of publications and software artifacts. These are listed in the references at the end of this report.

3.1 Human Resources

Roughly 17 graduate students, 19 undergraduate students, 2 research scientists, and 2 visiting faculty members have been associated with this project during the period covered by this annual report.

3.2 Student Placement

We continue to place student members of our research group in interesting project-related internships, graduate programs and industry jobs.

Undergraduate alumni of this project are now grad students at the following places:

- Ilya Bagrak is currently a graduate student at University of California, Berkeley. He received an honorable mention by the CRA for outstanding undergraduate and received an NSF and DOE graduate fellowship in 2004.
- Ansley Post is a graduate student at Rice University.
- Kirill Mechitov is a graduate student at UIUC.

Several students trained by this projects are sought after for pursuing internships in industries:

- PhD student Xiang Song is spending the Summer and Fall of 2005 working with Dr. Raj Kumar at HP Labs in Palo Alto on grid infrastructure for "appliance computing." This is a continuation of the work he did for that group in Summer/Fall 2004.
- PhD student Rajnish Kumar is interning at NEC labs.
- PhD student Dave Lillethun is interning at Motorola labs.
- PhD students Namgeun Jeong and Hasnain Mandviwala are working as year-round interns (since January 2004) at the Federal Reserve Bank in Atlanta with their cluster computing group. Some of the technologies developed through partial support from the ITR grant such as the Stampede system is being used as a dynamic cluster scheduling framework for running compute intensive applications developed by economists.

Recent alumni of this project include:

- Undergraduate Zachary Crowell is finishing his degree and joining Microsoft (this was enabled by an internship he did in the Summer of 2004);
- Yavor Angelov is pursuing his PhD under Professor Ramachandran while employed at Microsoft.
- Former Master's student Derick Pack has joined a Naval Research laboratory in Charleston, South Carolina.
- Dr. Sameer Adhikari [1] graduated in December 2004 and joined Intel, Portland as a system engineer.
- Dr. Arnab Paul [16] graduated in May 2005 and has joined Dartmouth University as a post-doctoral fellow under Dr. David Kotz.

3.3 Research and Education

The research artifacts from the project are finding their way into graduate courses and we have significant undergraduate participation in project-related research. We have funded a number of undergraduates through the REU supplement attached to this ITR grant, sponsored a number of independent undergraduate research projects for course credit (CS 4903), and have sponsored three capstone senior design projects (CS 3901) that each result in a poster presentation at the annual Undergraduate Research Symposium. One senior design project related to MediaBroker Federation (James Kim) was completed in Fall 2004, and another SensorLab project (Robert Thomas) was completed in Spring 2005.

SensorLab software was used for a project in a graduate course on Distributed Systems (CS 559) taught by Phillip Hutto at Emory University in Fall 2004.

The ITR project has reinforced the connectedness of hardware and software and the need to train students in system architecture quite early in their undergraduate preparation. With this in mind, Professor Ramachandran embarked on writing a set of course notes [19] for use in the sophomore level course on systems and networks. The course notes were well received by the students.

4 Special Requirements

The total budget for this 5-year proposal is $1.35M. However, due to fiscal constraints NSF chose to front-load the total budget by awarding $750,000 in the first year. The second, third, and fourth year increments were $200,000 each. The understanding with the program manager (Dr. Helen Gill) is that our spending plan for the award will be more balanced over the 5-year period despite the front-loaded nature of the allocation.

References

[1] Sameer Adhikari. *Programming Idioms for Pervasive Computing*. PhD thesis, College of Computing, Georgia Institute of Technology, December 2004.

[2] Bikash Agarwalla, Nova Ahmed, David Hilley, and Umakishore Ramachandran. Streamline: A scheduling heuristic for streaming applications on the grid. Technical Report GIT-CERCS-05-14, Georgia Institute of Technology, Atlanta, GA, May 2005.

[3] Yavor Angelov, Umakishore Ramachandran, Kenneth Mackenzie, James Matthew Rehg, and Irfan Essa. Experiences with optimizing two stream-based applications for cluster execution. *Journal of Parallel and Distributed Systems*, 2005. To Appear.

[4] Joshua B. Fryman, Hsien-Hsin S. Lee, Luke A. Snyder, Chad M. Huneycutt, and Gabriel H. Loh. Intelligent cache management by exploiting dynamic UTI/MTI behavior. Technical Report GIT-CERCS-05-08, Georgia Institute of Technology, CERCS, 2005. Also submitted to IEEE MICRO 2005.

[5] Phillip Hutto, Simon Chen, Ryan Graciano, James Kim, Thomas Shanks, Rex Sheridan, Robert Thomas, Matthew Wolenetz, and Umakishore Ramachandran. Sensorlab: A Java sensor toolkit (jstk). *Third IEEE International Conference on Pervasive Computing and Communications (PerCom'05)*, March 2005. Under submission.

[6] Rajnish Kumar, Arnab Paul, and Umakishore Ramachandran. On improving broadcast reliability of sensor networks using erasure codes. Under submission to IEEE International Conference on Network Protocols (ICNP 05).

[7] Rajnish Kumar, Arnab Paul, and Umakishore Ramachandran. Fountain broadcast for wireless sensor networks. Technical Report GIT-CERCS-05-02, Georgia Institute of Technology, CERCS, Atlanta, GA, February 2005.

[8] Rajnish Kumar, Arnab Paul, and Umakishore Ramachandran. Fountain broadcast for wireless sensor networks. In *Proceedings of the Second International Workshop on Networked Sensing Systems (INSS 2005)*, San Diego, CA, June 2005. To appear.

[9] Rajnish Kumar, Matthew Wolenetz, Brian Cooper, Bikash Agarwalla, JunSuk Shin, Phillip Hutto, Arnab Paul, and Umakishore Ramachandran. Dynamic data fusion for future sensor networks. *IEEE Transactions on Sensor Networks*, 2004. Expanded version of the DFuse paper that appeared in the First ACM Sensys Conference. Under submission.

[10] Hasnain Mandviwala, Nissim Harel, Umakishore Ramachandran, and Kath Knobe. Adaptive resource utilization via feedback control for streaming applications. In *Proceedings of the 19th International Parallel and Distributed Processing Symposium*, Denver, CO, April 2005.

[11] Hasnain A. Mandviwala, Nissim Harel, Kathleen Knobe, and Umakishore Ramachandran. Distributed garbage collection for timestamped data. *IEEE Transactions on Parallel and Distributed Systems*, 2005. Expanded version of three papers that appeared in earlier conferences, *PODC 2000, ICPP 2002* and *LCPC 2002*, under submission.

[12] Martin Modahl, Ilya Bagrak, , Matthew Wolenetz, Ramesh Jain, and Umakishore Ramachandran. An architecture for eventweb. In *Proceedings of the 10th IEEE International Workshop on Future Trends in Distributed Computing Systems (FTDCS'04)*, pages 95–101, May 2004.

[13] Martin Modahl, Matthew Wolenetz, Phillip Hutto, and Umakishore Ramachandran. Mediabroker: An architecture for pervasive computing. In *Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, pages 253–262, March 2004.

[14] Jin Nakazawa, Hideyuki Tokuda, W. Keith Edwards, and Umakishore Ramachandran. G2ui: A geographical user interface for multimedia exploration using diverse off-the-shelf devices. 2005. Under submission to ACM Multimedia.

[15] Ripal Nathuji, Joshua B. Fryman, and Karsten Schwan. Dynamic voltage and frequency scaling in real-time pervasive computing environments: A systems-level perspective. 2005. Under submission to RTSS 2005.

[16] Arnab Paul. *Application of Error Correcting Codes to Distributed Pervasive Computing*. PhD thesis, College of Computing, Georgia Institute of Technology, December 2004.

[17] Arnab Paul, Rajnish Kumar, and Umakishore Ramachandran. LAWN: A lightweight authentication protocol for wireless networks. In *Proceedings of the 4th IEEE International Conference on Network Computing Applications (NCA 05)*, July 2005. To appear.

[18] Arnab Paul and Umakishore Ramachandran. Computation-communication trade-off for power and bandwidth saving for remote authentication over wireless networks. Technical Report GIT-CERCS-04-01, CERCS, Georgia Tech, January 2004.

[19] Umakishore Ramachandran and Bill Leahy. Course notes for CS 2200 - introduction to systems and networks, 2005. Draft of textbook for CS 2200, URL: http://swiki.cc.gatech.edu:8080/cs2200/1.

[20] Umakishore Ramachandran, Martin Modahl, Ilya Bagrak, Matthew Wolenetz, David Lillethun, Bin Liu, James Kim, Phillip Hutto, and Ramesh Jain. MediaBroker: a pervasive computing infrastructure for adaptive transformation and sharing of stream data. *Elsevier Journal on Pervasive and Mobile Computing, special issue*, (2), 2005. To appear.

[21] Matthew Wolenetz, Rajnish Kumar, Junsuk Shin, and Umakishore Ramachandran. Middleware guidelines for future sensor networks. In *Proceedings of the First Workshop on Broadband Advanced Sensor Networks*, 2004. To appear.

[22] Matthew Wolenetz, Rajnish Kumar, Junsuk Shin, and Umakishore Ramachandran. A simulation-based study of wireless sensor network middleware. *International Journal on Network Management, special issue on sensor networks*, 2005. To appear.