Planning Grant I/UCRC: Net-Centric Software and Systems Center

Introduction

This planning grant proposal describes our plans to create a Purdue University site to the already established Texas-based Net-Centric Industry-University Collaborative Center (I/UCRC), which is being funded by the National Science Foundation (NSF). This planning grant proposal aims at: a) making Purdue a site of the already established NSF Net-Centric I/UCRC, b) increasing the number of industry members to levels that meet or exceed the requirements of NSF, and c) adapting the policies, fees, bylaws, and agreements with existing NSF Net-Centric I/UCRC sites.

The funds from this planning grant will be used primarily to organize an industry-university meeting and open house that will help market the new Purdue Net-Centric I/UCRC site to industry in order to bring the Purdue membership at levels that meet or exceed those required by this NSF program. Research problems that the proposed I/UCRC site will tackle revolve around research in Net-Centric environments with focus on Cloud, end-to-end secure software design, construction of reliable net-centric distributed storage repositories, secure and adaptable mobile-cloud centric computing, data-sharing across concurrent computations, reducing energy consumption of mobile-centric systems, as well as information flow tracking data privacy protection in net-centric systems.

Net-centric systems are important in industry as well as in homeland security, Air Force and Army applications. The center will enable the creation of additional capabilities in net-centric computing. The proposed site is uniquely positioned to promote research, education, and large-scale industry activities. Our faculty members are doing leading research topics related to net-centric in mobile computing, distributed systems, parallel programming, and network security among many other fields of Computer Science. They have funding from multiple sources including the National Science Foundation (NSF), the Air Force Research Lab (AFRL), Army Research Lab (ARL), Northrop Grumman, Cisco, Motorola, Microsoft Corporation, NEC Laboratories, Department of Energy (DOE), Samsung, etc.

Planning Grant Objective

The objective of this planning grant is to organize an industry–university meeting both with the industrial partners who have shown interest in joining the center and with several new prospects who are willing to explore the partnership benefits. The meeting will showcase the work of our faculty and graduate students in the area of cloud centric systems, secure and adaptable mobile-cloud computing, end-to-end security in service-oriented architecture, reliable distributed storage for data-intensive applications, data-centric programming models, and tackling energy consumption issues of mobile systems by treating them as *network-centric* devices.

Marketing Plans

We have discussed a common research and education agenda of the Purdue site with the directors of the existing center sites. The next step is to get feedback from the prospective members from industry. This will be done by sending them a copy of this proposal, brochure describing the existing center at University of Texas and ongoing research activities. Perspective industrial members will be invited to various meetings of the center that we will attend in the next year. We will seek advice from current members of the consortium and the chair of the industrial advisory board.

Potential Industry Members and Recruitment Plan

The directors of development in CS and ECE will provide publicity of the center's plan to industry and corporate partners. The meeting will be held either a day before or after the

corporate partners meeting that is held twice a year in Purdue. The next two meeting are planned for April and September 2013. It coincides with the job fair that attracts over 200 companies to Purdue. The list of current companies involved with the CS department is available at: http://www.cs.purdue.edu/external_relations/corporate/partners/. Almost every week representatives of companies come to the CS department to recruit students (http://www.cs.purdue.edu/external_relations/corporate/opportunity_update/), and one-on-one contact with them will be made directly during their visits. Alumni and former students will be contacted for encouraging their research groups in net-centric systems to learn about the center and join as a member.

We have already contacted over thirty corporate partners and friends of computer science and ECE departments. The development officers in the colleges of Science as well as Engineering and associate deans of research have been asked for contacts and help. The vice-president for research at Purdue has agreed to provide the necessary support to help recruit new members to support the Purdue site.
The second step is to present the center plans to the thirty corporate partners and friends in the next meeting in April, 2013 and seek their help. We will discuss this with various industry representatives who regularly come to Purdue for recruiting and during the annual spring job fare. We will contact several computer science related companies in the Purdue research park in West Lafayette.
We have support letters (attached) from eight potential members who have shown interest in joining the center and endorsed our proposal to become a Net-Centric NSF I/UCRC center. The I/UCRC Industrial membership agreement letter and support letter form current director (Prof. Krishna Kavi at UNT) are attached.

The plan is to recruit companies based on other NSF I/UCRC centers experience. A plan will be designed to organize an industry day for recruiting additional companies. The plan consists of: a) identifying contact persons in the area, b) visits to company sites with presentations by consortium director and by consortium faculty, c) invitation of company representatives to facilities for student posters and laboratory tours, d) invitations to the open house, e) teleconferences and additional email exchanges, f) exchange of list of potential projects to be executed by the consortium faculty and students, e) IP and agreement discussions between university and industry legal departments.

Meeting of Potential Industry partners at Purdue
The next major step will be to hold a meeting of all interested industrial partners in Purdue in September, 2013 when all corporate partners come to Purdue. The timing is flexible depending on the funding/budget cycle of industry partners. This meeting will expand our agenda and include elements from our proposed partner universities in Texas, Missouri, and Arizona, which will be invited to present their work at this meeting. The meeting will establish a process for our current members to adopt the policies, bylaws and membership agreement of the Net-Centric I/UCRC. The Director of the proposed site was involved in meetings with industry at the University of North Texas during the formation of this center, and has worked with industrial partners for many Purdue-based centers such as CERIAS and SERC. The Director of the proposed NSF I/UCRC site has preciously founded the IEEE International Symposium on Reliability in Distributed Systems in 1981 (31st will be held in Irvine, CA in Oct., 2012) and served as the chair of steering committee. This will provide a forum for wider publicity for the proposed NSF center. He also has ongoing collaboration with the site director of NSF I/UCRC Center on Net-Centric System Software at Missouri University of Science and Technology.

The major meeting to finalize the funding from industrial partners will be organized as follows:

Meeting arrangements:

- **Proposed location:** Lawson Computer Science building Commons area and room 1142. The facility can host 150 attendees and 50+ student posters. It has ample parking space, a food court and hotels (including one in Purdue Memorial Club) are within walking distance. The research laboratories of the faculty are in the same building.
- **Meeting format and organization:** The meeting will have morning and afternoon sessions. The morning session of the meeting will have presentations by the director, presentations by some potential industry members, and presentations by faculty highlighting their Net-centric research areas. The afternoon session will feature about 10 graduate student posters presenting research results and some related demos.
- **Responsibilities of staff and presenters:** The director will present the plan to join Net-Centric as a site, a summary of the bylaws and elements of the consortium agreement. The director will also introduce the Purdue faculty members who have agreed to join Net-Centric I/UCRC. The current industry sponsors of different projects will present their potential benefits from joining the center and profiles of their companies while faculty will introduce their areas and their student posters.
- **Draft agenda**: A draft agenda is given below:

Morning Session

8:30 am Registration
9:00 am Introduction (Director)
9:15 am Opening Remarks by the Vice President for Research at Purdue University and College of Science Associate Dean for Research
9:30 am Remarks by NSF and Net-Centric representatives
10:00 am Status of the Potential Members
10:30 am Profiles of Potential Member Companies (Industry representatives will be invited)
Coffee Break
11:00 am -12:30 am Presentations by the PI on the proposed projects
Lunch

Afternoon Session

1:30-2:30 pm Other Faculty Presentations
2:40-4:00 pm Student Posters
4:00-6:00 pm Social Get Together

Marketing Materials

The center created will enhance marketing materials including:

- Database of target companies;
- Contact names and industry group interests / matching with consortium faculty areas;
- A comprehensive web site with links to other member sites and industry partners
- Power point presentations outlining benefits of the consortium and technical area descriptions
- List of current projects and resources available and a center brochure

Project Description

The mission of the proposed center is to perform basic and applied research and train students in net-centric systems, cloud-based software design and implementation and issues dealing with security, trust, privacy, information management platforms, as well as energy issues in mobile computing. Applications addressed will include information and software systems, defense and homeland security, real-time data processing, cloud-based computation and information processing systems, mobile applications and opportunistic networks among others.

Markets in resource management, QoS, data management, security and trust in mobile computing, and cloud computing have been growing exponentially during the last five years and will accelerate with the wider adaptation of lightweight mobile computation devices such as smartphones.
The goals of the proposed NSF I/UCRC site include:
(a) To advance the research in Net-centric computing, cloud computing, mobile computing;
(b) To produce a highly trained workforce that will support emerging industry research and development activities;
(c) To promote scholarship and research dissemination by presenting and publishing papers in prestigious conferences and journals;
(d) To establish scientific leadership and organize workshops, special sessions, and special issues in new areas;
(e) To establish a patent portfolio that will be shared with partner industry;
(f) To create short courses and write monographs for continuing education of local and national industry engineers and practitioners;
(g) To attract high quality graduate students with diverse backgrounds in the Ph.D. program and produce theses with cutting edge results;
(h) To launch outreach activities to attract high quality students in engineering.

Our multidisciplinary faculty participation brings capabilities in mobile-centric system design, programming languages, parallel programming, compilers, software engineering, energy measurements and evaluations in net-centric computing test-beds, security and privacy in net-centric environments, and large software development projects. Our faculty activities are already supported by industry partnerships, federally funded programs that include, fundamental research funded through AFRL, Naval Research Laboratory, DARPA, Northrop Grumman, Cisco, Motorola, etc.

Adaptation of Operations of the Site According to Net-Centric I/UCRC Policies:
The Purdue site will adapt its operations according to the Net Centric I/UCRC as outlined below:
• **Director:** The Purdue site Director will work closely with the Net-Centric Director Dr. Krishna Kavi, and the IAB on the research focus and related administrative matters. The Purdue site Director will be responsible for bringing the project proposals to the Net-Centric IAB for voting.
• **Agreement between the Purdue site and supporting companies:** An agreement has been drafted by Net-Centric and is attached as part of the supporting documents of this planning grant proposal. This agreement is approved by the Purdue Office of Sponsored Programs.
• **Corporate Relations:** The Center will work with a full time corporate relations/marketing staff at Purdue University to ensure technology transfer, attract new members to the center and to ensure good relations with the supporting companies.
• **By Laws:** The operations and voting process will be adopted after approval by the IAB.
• **Research Funding:** The fee for a company to join the Center will be $30,000 per year (as already established by the Net-Centric Center). For smaller companies, smaller amounts will be considered. Research in the Center will be carried out by graduate/undergraduate students and postdoctoral researchers, under the supervision of the participating faculty.
• **Center Evaluator:** The industry/university interactions will be independently observed and formally evaluated by an Independent Center Evaluator (to be named).
• **Center Benefits:** The supporting companies will get all the usual benefits already outlined as part of the Net-Centric Center. These include (1) opportunity to direct the research for industry's benefit, (2) access to student recruiting, (3) leveraging of research supported by other corporate members and others' projects already carried out by the participating faculty, (4) direct access to faculty members, and (5) understanding of capabilities within the participating labs and other universities.

• **Meetings:** As already setup by the Net-Centric I/UCRC, Purdue will be participating in the Net-Centric Center meetings. New projects will be proposed for review by the IAB at the summer meeting. Purdue will offer to hold such meetings periodically at its conference facilities and will be responsible for the organization of the meeting.
• **Proposal Selection:** Purdue will follow the I/UCRC proposal selection process.

Purdue Recruitment of Students from Underrepresented Groups
For recruiting talented minority students, Purdue has created a welcoming environment and campus culture by establishing Student Diversity & Academic Support Programs in College of Science and Engineering as well as in CS and ECE departments to assist and support them during the period of their study. As part of Purdue's strategic action plan, the colleges have established a team of faculty and staff to develop a plan to recruit and retain minorities and females into programs in science and engineering. The NSF AGEP program is being led by the dean of graduate school. The representative of each college for minority programs in each college will help in coordinating with the partner HBCUs in recruiting interested and qualified candidates, including women.

Purdue has a special unit called Services for Students with Disabilities in the office of students' affairs, which caters to all the needs of such students. The office has agreed to provide their services in case of any need. The CS department has a video conferencing facility donated by Cisco for direct delivery. In addition, Purdue is fully accessible to disabled students including the library, classrooms and desks.

The PI has been active in recruiting and graduating students from underrepresented groups. He will work closely with minority representatives that run and attend minority conferences. The PI will also provide minority recruitment committee demonstration modules designed specifically by some of the minority students in their orientation events to stimulate interest of students from underrepresented groups in computing. The PI and faculty will coordinate with minority offices in science and engineering colleges that have databases and content that target native Indian, Hispanic, and African American students'

Transformative Research Aspects
Research conducted under this I/UCRC will revolutionize net-centric computing; resulting in paradigm shifts in mobile, cloud and distributed computing, with enhanced security, reliability and performance. The products of the research activities performed at the center will be largely adoptable by industry, leading to changes in the ways companies and organizations process, analyze, and design net-centric and cloud software, services and systems. The research to be performed in the field of mobile-centric computing will create models providing significant performance and energy gains over existing mobile systems. Enhanced security of mobile computing achieved with these projects will result in products that will be highly adoptable in critical missions and defense as well as various other fields using mobile computing. Enhanced real-time response achieved in mobile-cloud centric computing will transform computing preferences of millions of users. Projects focusing on enhanced parallelism in data-centric applications will attract many top-notch software companies and transform the way they process big data through further advances in the field. Reliable and secure storage and dissemination of big data to be achieved with the center activities will result in greater adoption of cloud computing resources by many organizations, hence leading to savings in computing resources.

Transformative aspects of this I/UCRC will deal with a plan that includes a research internship program where graduate students involved in this I/UCRC spend time at the core research laboratories associated with the application areas such as at AFRL. This program will also be complemented with an internship program in collaboration with our Industry Consortium

members. As part of their paying membership agreement our industry partners are very interested to host students for summer internships.

Relation of Current Research to Current and Past Projects
The faculty group of this I/UCRC have had several projects funded by ARL, AFRL, NSF and DOE including reliable pervasive systems, dynamic program reasoning, compiler optimization, distributed storage systems, adaptable distributed systems, SOA security, secure mobile systems etc. The experiences gained from the faculty team with the prior work will be fundamental in making I/UCRC projects a success. We will engage a large number of students in EPICS (Engineering projects in Community Service).

Enhancement of Courses for NSF I/UCRC Research Students
Faculty members are working on designing and enhancing courses related to Net-Centric computing. For example, the PI has taught a course on cloud computing for the blind and hearing-impaired in Spring 2010. Some of the industry projects will involve capstone projects assigned to undergraduate students. Teams of I/UCRC graduate and undergraduate students will work on these projects. The objectives of the course projects will not only include exposure to particular research topics but also exposure to practices and logistics of research, such as:
Working on team-oriented research tasks with collaborators and supervisors**:** Interaction with industry partners, Promotion of ethics of research: Includes formal lectures to educate students and promote practices that are consistent with the notion of research ethics as stated in NSF and IEEE/ACM policies, Research literature survey practices using web and library resources**:** Students will be exposed to successful literature survey models, Reporting formats of research results with standard IEEE/ACM specifications: Helps students learn how to write papers with page limitations, Research presentations to peers, supervisors, and potential sponsors: Students will be exposed to technical writing and presentations.

The following are the courses which I/UCRC associated students will be recommended:

Courses	Description	Proposed Project
CS 525	Parallel Programming (G)	Project 8
CS 390/CS 505	Distributed Systems (UG/G)	Projects 2, 3, 4, 6, 7, 8
CS 426	Computer Security (UG)	Projects 1, 3
CS 626	Advanced Information Assurance (G)	Projects 1, 3, 5, 6
CS 307	Software Engineering I (UG)	Project 2
CS 526	Information Security (G)	Projects 1, 3, 5
CS 422/CS 536	Computer Networks (UG/G)	Projects 2, 3, 4, 6
ECE 364	Software Engineering Tools Laboratory	Project 1
ECE 461	Software Engineering	Projects 1, 2, 6
ECE 468	Introduction to Compilers	Project 1
ECE 568	Embedded Systems	Project 2
ECE 573	Optimizing Compilers	Project 1

Table 1: List of Courses related to the proposed projects

Sample Research Projects in Net-Centric Computing Areas

Project 1: Attribution in Cyberspace through Calling Context Encoding

Attribution in cyberspace is the capability of attributing any activities in cyberspace to their intents. It is critical to forensic analysis and cyber-attack deterrence in net-centric environments as any malicious behavior can potentially be traced back to the entities that are responsible. Despite its importance, accurate attribution is extremely hard, as it demands an efficient representation of intents; it entails propagating potentially large amount of such intent information across the entire IT infrastructure.

The closest existing technique is system wide information flow tracking that taints information coming to an IT system as safe or unsafe depending on the policy, e.g., packets received from an untrusted source are tainted as unsafe. Such taint information is then propagated throughout the system according to the activities within the system to indicate if these activities are safe/unsafe. Information flow tracking falls short in cyberspace attribution because of the following reasons. First, it can only flow a few bits of information, and increasing the bandwidth incurs substantial runtime overhead. The state of the art instruction level dynamic information flow technique causes a few times slow down to system execution. Second, it only encodes simple safety information and does not provide a solution to encode intents or origins. While ids can be generated to represent entities, such ids are not adequate to precisely represent intents or origins. For instance, assume an application in the cyber space is compromised by being injected a piece of malicious code. The injected code may not always be active and hence the application may behave completely differently over time, giving rise to the necessity of discerning its different intents.

This proposal aims to develop a modeling scheme that can precisely encode the intent of an activity in the cyber space. It encodes a system intent into one integer value, which can be flowed through the whole infrastructure using existing dynamic information tracking systems to achieve the final goal of attribution. The encoding scheme is general and is applicable to most systems in an IT infrastructure. It is transparent, without demanding developers to change their systems. It is precise with the capability of disambiguating the different intents of a running system. It is cost-effective, incurring trivial runtime overhead.

In particular, we leverage our recent progress on efficient system calling context encoding. Calling contexts provide substantial information about intents. For instance, in our aforementioned compromised software example, even though both the benign code and the malicious injected code call the same low level socket communication methods to send messages, the contexts of these messages sends clearly represent their different intents. Hence, we propose to use calling contexts to encode intents. Recently, our research shows that we are able to encode the calling contexts of almost all C applications with a 32 bit integer. In other words, at any point of program execution, we maintain an integer that precisely encodes the current context. This integer can be tagged with a message to represent the original intent of the message, which can be further flowed through the whole IT infrastructure. Our experiments show that the runtime overhead of encoding is about 2% on average.

In the project period, we plan to extend the work to applications developed in other languages. Note that our encoding algorithm is general, although our current encoding implementation only supports C. We will build an infrastructure wide attribution prototype by integrating the encoding technique with information flow systems. Our group has developed a few information flow systems in the past. We have projects funded by NSF on building information flow systems.

Team: Xiangyu Zhang (http://www.cs.purdue.edu/homes/xyzhang/)

Proposed Work and Milestones:

• *Month 1-9.* In our study, we observe for some programs that use recursion intensively, such as the SPECint program 197.parser, a few integers may be needed to encode the calling contexts

in order to ensure precise decoding. We observe that these extra numbers are needed not because we are running out of the encoding space, but rather due to cycles in the contexts. Hence, we will explore the solution of unrolling recursions in static call graphs so that our current encoding algorithm can uniquely encode recursive paths. We will also study the solution of hashing a few numbers to one number. Our current prototype is implemented on an infrastructure called CIL that supports C programs. It is robust enough to handle programs up to 800K LOC. However, in order to achieve the generality goal, we plan to port the technique to GCC such that all programs supported by GCC will be supported by our technique, including those written in C/C++, Fortran, and so on. We have expertise on the GCC platform and we are very familiar with the GCC intermediate representation GIMPLE.

• *Month 10-15*. Integrate the proposed technique with our GCC based information flow system to build a proof-of-concept system. Evaluate our system on a few networking applications. The experiment will be designed as follows: we will hijack one of the nodes in the network and issue malicious requests to the network. Then, we will observe how the entailed malicious actions on other nodes can be attributed to the malicious intent of the contaminated node. We expect our technique can locate the precise code injection place.

Project 2: Analyzing and Reducing Energy Consumption of Mobile-Centric Systems

For mobile-centric systems, limited energy from batteries is a major constraint. Today's mobile systems are connected through wireless networks (Wi-Fi, 3G, Bluetooth, etc.) to a wide range of heterogeneous systems: laptops, desktops, servers, and so on. Many studies have shown that migrating heavy computation to grid-powered desktops or servers may save mobile systems' energy [1-4] and prolong mobile systems' battery life, even though communication through networks also consumes energy. Thus, when we develop solutions for analyzing and reducing energy of mobile systems, we have to consider the effects of network connections. This project intends to answer the following questions:

- How much energy is consumed by a mobile system for using various types of services? How do the interactions between mobile systems and servers affect energy consumption? More specifically, if a service is changed, how would that affect the mobile systems' energy? The changes may be invisible to mobile users. For example, for cloud storage service, the service may compress files to use network bandwidths more efficiently. This adds loads to the processors of mobile systems, while reducing loads to the network interfaces. We will study whether compression always saves energy. If the answer is negative, we will investigate under what conditions compression saves energy. The changes may be visible to users, for example, through different user interfaces or additional options. There is no systematic way evaluating the energy consumption of these services.
- Mobile systems are becoming much faster. Dual core smartphones are common. Quad-core tablets with 1GB memory are available. Meanwhile, network latency is remaining nearly unchanged. This trend suggests that computation should migrate back to mobile system. Meanwhile, another trend is moving in the other direction. The needs for complex computation on mobile systems are growing, for example processing images, playing video, presenting high-quality graphics, and gaming. Even though solutions exist today for deciding where to perform computation, these solutions are piecemeal and uneasy to generalize.

As can be seen from the previous description, we study how to save energy of mobile systems by treating them as *network-centric* devices, not stand-alone devices. The main challenges are modeling the interactions of these connected systems and the effects on energy consumption.

Each system consumes energy due to four types of activities: (1) external activities from users or environment, (2) responses from another system through networks, (3) self initiated activities because of events (such as a timer) inside a system, and (4) continuation from previous activities. We use discrete-time models to express how one system affects another system's power consumption at the next time interval. Our approach is scalable because each system is connected only a few other system (most likely only to a network switch or an access point).

Based on the energy models, we develop energy-aware software that can self evaluate the energy of different features. This information is then propagated through network to dynamically adjust the services for better energy efficiency. We have already developed a programming framework in which software can adapt and migrate to improve energy efficiency. We have developed several case studies for image and video processing [5-8]. These programs estimate the energy consumption on mobile systems (smartphones and tablets) based on the amount of operations and contents. With this estimation, computation is performed on servers or mobile systems. Through the collaboration with the faculty and industry partners in this IUCRC, we will generalize our current work for a wider range of software for more services.

Team: Yung-Hsiang Lu (https://engineering.purdue.edu/HELPS/Faculty/yunglu.html) and

Bharat Bhargava (http://www.cs.purdue.edu/homes/bb)

Deliverables: Programming framework, reference implementations for the case studies, research papers

Milestones: Year 1: Estimate energy consumption of services. Year 2: Develop a general programming framework. Year 3: Construct demonstration cases.

Project 3: Secure and Adaptable Mobile-Cloud Centric Computing

Current mobile applications mostly involve an inflexible split of computation between the mobile and cloud platforms. This inflexibility prevents applications from adapting to conditions such as high network latency, which could result in poor performance when cloud resources are preferred over computation on the device. One reason to avoid flexibility in computation resources is the security risks associated with using the cloud: The lack of control on resources and multi-tenancy of different users' applications on the same physical machine make cloud platforms vulnerable to attacks. In this project, we propose to design and develop a secure mobile-cloud computing framework that dynamically partitions a mobile application for distributed execution on mobile and cloud platforms. By focusing on achieving trusted communication with and trusted execution on the cloud platforms, this body of work is expected to result in a general-purpose secure performance optimization framework for mobile-centric computing. More specifically the following problems will be addressed:

- Adaptable application partitioning: A good application-partitioning scheme is crucial for optimum performance in terms of real-time response, low communication costs and low energy consumption on mobile systems. The partitioning scheme has to take the constraint into consideration that methods accessing specific features of the mobile system (such as sensors including GPS, camera, accelerometer etc) must be pinned to that device. This work will involve the development of a high-performance, constraint-based application partitioning scheme for mobile applications, building on existing work such as [9] in Java application partitioning.
- Protecting the integrity of offloaded data/computation: The multi-tenant structure of the cloud makes user code and data vulnerable to attacks by other malicious guests. A malicious host/guest with access to the virtual machine of a user can compromise the integrity of the data or computation to achieve a particular task with foreseen benefits. The research

conducted for the task of protecting the integrity of offloaded data/computation in this work will involve development of a dynamic integrity checking algorithm and self-destruction of data and computation upon detection of integrity-violating activities on the data that were offloaded to the cloud. The application partitioning algorithm will be advanced such that security-critical parts of application code are always executed locally, instead of in the cloud.

- Protecting privacy of sensitive data/computation in mobile-cloud computing: In the case of a malicious cloud host or malicious applications running on the same machine as the mobile application partitions, there is the possibility of proprietary code leakage. The current industry practice is either not to use the cloud at all due to possible leakage or run programs in clear text, which makes them vulnerable. Our proposed computation-offloading framework will alleviate this problem by partitioning the program and running the partitions on different machine instances, which allows for hiding the complete details of the program. In addition to that, using a program encryption scheme during execution will minimize the exposure of program code to malicious parties.

Our proposed framework for secure & adaptable mobile-cloud computing is based on the collaborative execution of program code by mobile agents sent to different machine instances in a Cloud. We plan to develop of a tamper-resistant program execution algorithm in the Cloud through "program self-encryption", which reveals (decrypts) program segments incrementally based on the previously executed segments, resulting in program failure upon tamper with program code. Through the collaboration with the faculty and industry partners in this IUCRC, we will design and develop a secure mobile-cloud computation framework integrating the latest industry standards and addressing the most common security and privacy needs in mobile computing.

Team: Yung-Hsiang Lu and Bharat Bhargava

Deliverables: Programming framework, reference implementations for the case studies, research papers

Milestones: Year 1: Develop adaptable mobile application partitioning framework. Year 2: Develop secure program execution algorithm and framework. Year 3: Construct demonstration cases and conduct performance studies.

Project 4: Opportunistic Data Analysis in Mobile-Centric Systems

One major challenge for all Internet users today is the rapid growth of heterogeneous data: email, attachment, twitter, multimedia, etc. The data are unstructured because they do not easily fit into a rational database. Moreover, they grow rapidly reflecting the social lives of users. This is one example of the "Big Data" problems. The data are already analyzed by business for planning and personalized promotions. However, data owners, the users, are rarely able to control the flow of data, protect their privacy, share data with their friends, not to mention take advantage of the knowledge embedded in the data for improving their own lives. However, as mobile-centric systems become the primary computing platforms for millions of users, their mobile systems do not have the performance, storage, and energy for running intensive computation to analyze data. As an example, a user may analyze their photographs and build a personalized image database. In this project, we propose an *opportunistic* approach by analyzing data when resources are available. The analysis programs run only when mobile systems are plugged in or when faster computers (such as desktops) or high-speed networks (Wifi vs. 3G) are available so that the analysis does not drain battery too fast. This is different from the typical mode of operations in mobile applications: they tend to be interactive and user-triggered. In contrast, in our approach, execution is triggered by availability of resources. The analysis improves gradually as more data are analyzed. This approach is also progressive: new

data are frequently added and analyzed. The results of this opportunistic analysis is that users obtain increasingly more accurate results using their own data without degradation of performance or shortened battery life. Our preliminary work [10-12] estimates the amount of energy consumed for analyzing images and representing each image by a set of features. Through is I/UCRC, we will work with industry partners to identify the types of personal data to analyze the information that may be extracted from the analysis.

Team: Yung-Hsiang Lu and Bharat Bhargava

Deliverables: Case studies, reference implementation of the analysis programs, research papers

Milestones: Year 1: Identify useful personal information. Year 2: Develop analysis programs. Year 3: Conduct usability study.

Project 5: Enhancing Data Sharing and Dissemination Security in Net-Centric Environments

Common approaches for protecting disseminated data against privacy violations in net-centric environments use a client application at the host receiving the disseminated data to enforce the privacy policies associated with the data. When data leaves the trusted domain or owner's sphere of control, it may be eavesdropped or stolen during transfer and may not reach its intended destination. The common approaches trust the client application to enforce the privacy policies associated with the data. These approaches are unable to protect data in such scenarios. An alternative approach for protecting disseminated data is to bundle together the data and metadata (policies) into a Managed Information Object (MIO), and provide a protection mechanism that enforces the policies. An MIO consists of data/content in its payload and metadata that describes the content and specifies various policies to protect that content [15]. Attaching policies to the data provides capabilities to MIO to control the behavior of data in a trusted domain. To have control over the enforcement of policies even when the MIO leaves the sphere of control or the trusted domain, we extend MIO to provide it with complete control over enforcement of its policies in any domain whether trusted or unknown. This gives MIO the capability to protect and control any interaction with its data. One approach to extend MIO is to encapsulate the data and metadata with another layer that is responsible for enforcement of any policies over its data and control data dissemination. This extended MIO is called Active Bundle. An Active Bundle (AB) is a data protection mechanism, which can be used to protect data at various stages throughout its life cycle. The active bundle scheme [13, 14] is based on the AB software construct. An Active Bundle bundles together sensitive data, metadata, and a Virtual Machine (VM) specific to the bundle.

In this project, we discuss the MIO approach for data dissemination and show that bundling data with metadata provides better dissemination control, enhanced security, and reduces the risk of privacy violations for disseminated data in net-centric environments. We intend to answer the following questions:

- Determine the capabilities gained by attaching metadata (policies) to the data, e.g. enhanced security properties and controlled data dissemination that reduce the risk of data leakage and privacy violation; determine ways to adjust and update metadata policies dynamically, based on a context, host, history of interactions, trust level etc.; determine suitable parameters for trust evaluation of hosts, e.g. history of interactions, context parameters etc.

- Determine the decision on a host request and decide the content of information to be disseminated, e.g. if in an emergency context in an unsecured location with possibility of attack, decide what data to disseminate immediately.

• Design, prototype and evaluate the approach in different models like with and without trusted third party, multiparty computation, publish-subscribe model, peer-to-peer model, service-based model (SOA) etc. Evaluate performance of the approach in different deployment setups such as data sharing and dissemination scenarios like cloud computing, mobile cloud, SOA etc. and investigate how changing the distribution of the system components among hosts affects the performance of the prototype.

Team: Bharat Bhargava

Deliverables: Prototype framework, reference implementations for the case studies, research papers

Milestones: Year 1: Analyze the security properties and protection capabilities of MIO. Year 2: Prototype and extend the MIO approach. Year 3: Construct demonstration cases.

Project 6: End-to-End Security Auditing in Service Oriented Architecture (SOA)

Modern cloud-based services are distributed and loosely-coupled in nature. We cannot trust them and assume they are behaving as expected. Moreover, complexity of the SOA-based systems leads to a large attack surface. To provide high security assurance in cloud computing, it is not enough to model the services as blackbox and then inspect only the corresponding inputs and outputs of them. Particularly, the blackbox abstraction fails in scenarios that infrastructures and services are not under the direct control of computing consumers (as in cloud computing). Moreover, any pre-certification of services will fail in such scenarios. This is because of the fact that services may get compromised by attackers or modified by cloud service providers due to economic incentives. Therefore, firstly, it is essential for a monitoring framework to inspect the services dynamically during their execution. Secondly, the monitoring framework must not be easy to bypass by malicious services. This requirement gives priority to the monitoring mechanisms, which operate outside of the service execution domain. Thirdly, the monitoring scheme must not incur a high performance overhead to make it viable for large-scale deployment. Finally, a good monitoring mechanism should be as transparent as possible to the service providers and users. This condition is necessary for real world success and adoption by industry.

We have proposed and developed a new monitoring system based on taint analysis which discovers the services that may be either malicious or get compromised over time [16]. The proposed auditing system is an essential part of secure cloud based systems and needs to be efficient. In this project, we will extend our taint analysis framework to monitor the internals of SOA components as well as interactions among them in the cloud environment. We will provide a service-level information flow control (IFC) module. This IFC scheme will enable us to observe how data is manipulated within each service and how this data is transferred among services.

In our proposed approach, the taint analysis framework monitors the activity of services (at runtime) and inspects the data exchanges (information flow) between them to detect certain events. To achieve runtime execution monitoring, we need dynamic program instrumentation. In this mechanism, extra instructions are automatically added to service implementation on the fly during their execution, without the users' awareness. The execution of the instrumentation tracks the information flow in the execution. Intuitively, one can consider such instrumentation as hooks to the execution so that the taint analysis component can gain control when certain events occur. In this project, we leverage aspect oriented programming (AOP) to achieve a special information flow control for auditing. This technology empowers our solution to work outside of the application servers and execution domains (i.e. at JVM level). This property complies with the second requirement.

Most of the currently proposed taint analysis schemes are heavyweight, which perform static and dynamic binary execution profiling on commodity software [20, 21]. These solutions need a major change in the current operating systems. But, as mentioned earlier, we employ AOP for information flow control, which has a reasonable performance [16]. Therefore the third condition is met by our solution. Our taint analysis framework (based on AOP) is transparent to the services and users. Therefore, service providers are not required to change their programs or deployment (or need minimal changes). This feature satisfies the last condition.

This system will be used to collect and report a wider range of audit trails related to manipulation of data by services. It will be used as an agent to communicate other source of audit trails. Data collection will include (but not limited to):

- Information path within a service (How service inputs are propagated/manipulated within a service)
- Information path between services (How will services interact? Are sensitive data propagated to unauthorized services?)
- Performance data (to benchmark the overhead of taint analysis service)

Moreover, we will take advantage of the currently available web service security standards (WS-*)[17-19] in our design.

We are planning to deploy and test the taint analysis framework in a cloud test-bed. This experiment will demonstrate how we can inspect the behavior of closed-source services by runtime analysis of their execution. This will enable the monitoring and reporting of illegal service invocations and suspicious data manipulation. We will adapt our framework for general SOA practice through the collaboration with the faculty and industry partners in this IUCRC.

Team: Bharat Bhargava and Xiangyu Zhang
Deliverables: Methodology for evaluating the security of SOA systems, prototype implementation of the proposed security architecture, research papers
Milestones: Year 1: Design a new auditing and policy enforcement architecture for SOA. Year 2: Develop prototype of taint analysis based information flow control. Year 3: Large-scale cloud based deployment and improvement.

Project 7: Reliable Distributed Storage for Data-Intensive Applications Using Graph Formalisms

While economies of scale have made it feasible to envision petabyte storage repositories, ensuring reliable and efficient access to stored data remains a significant challenge. Tradeoffs between performance, storage efficiency, and availability often lead to ad hoc solutions, or infrastructure specialized for specific domains. This project takes a more principled approach towards the construction of reliable net-centric distributed storage repositories. The key insight is in formulating availability metrics in terms of distributed graph operations. The graphs manipulated by the storage system correspond to the generating matrices of erasure codes [22, 25]. We argue that the structure of these graphs can be exploited to guide data block placement and access, and can be used to minimize communication and computation overheads. We propose to demonstrate the feasibility of our ideas by building a cloud-based distributed storage repository in which data blocks used for primary storage can have the same degree of availability, as archival data. To achieve this goal, we will leverage techniques from the programming language and systems communities to enforce atomicity of block updates, mask latency via controlled speculation, and devise caching policies based on application characteristics and demands.

The proposed research plan tackles challenging questions in (distributed) algorithms and coding theory, has a substantial system-building component, requires a deep understanding of application requirements, and will necessarily involve construction of specialized application-level APIs. Specifically, the project will investigate (i) development of algorithms, analyses, implementation, and benchmarking for reliability in distributed storage systems; (ii) formulation of adaptive storage requirements and their implementation; and (iii) characterization of application requirements, application-specific optimizations, analyses, and implementation. Each of these poses deep theoretical, systems-oriented, and application-interfacing challenges, which form the proposal's technical core. These challenges are addressed through a mix of graph-based techniques, distributed implementation of numerical methods (matrix updates, matrix factorizations), average case analyses, and formalization of storage semantics.

The proposed research poses significant challenges. While extensive prior research [23, 24] has focused on important measures such as rate, redundancy, computation, and memory requirements, distributed environments have other constraints as well. These include (i) communication and synchronization overheads -- specifically, it may not be feasible to assemble all requisite blocks of coded data to decode a required block; (ii) scalability -- the data sizes may be such that no single host may store all data associated with a decode operations; (iii) performance -- optimizations such as caching interfere with erasure coding; (iv) application characteristics -- leveraging application characteristics (relative read write mix) often leads to tremendous performance improvements; and (v) customizability -- the ability to modify availability requirements of all or part of the dataset dynamically. These form the focus of the research and development efforts proposed in this project.

Our approach to these problems relies on our insights from distributed sparse linear algebra (coding and decoding operations are mat-vecs and direct linear solves over a finite field, respectively), semantics of concurrent access in wide-area overlay-based distributed systems, scheduling techniques for load balancing and communication, and caching/replication strategies of storage blocks for performance. Relying on a graph-theoretic view of the computation, we pose various operations (coding, decoding, recoding) as distributed computations with replicas providing needed degrees of freedom for optimization. Semantics of access and availability of resources constrain the space of solutions. The goal is to guarantee user-specified levels of reliability, while minimizing resource utilization and access time.

Team: Ananth Grama (http://www.cs.purdue.edu/homes/ayg/) and

Suresh Jagannathan (http://www.cs.purdue.edu/homes/suresh/)

Milestones: Year 1: Design and Implementation of System Primitives. Year 2: Consistency mechanisms and caching techniques. Year 3. Implementation of Distributed Coding/Decoding

Project 8: Data-Centric Programming Beyond Data Parallelism

Data-centric programming models like MapReduce [26, 27] and Dryad have received considerable attention over the past few years. The success of these models can be attributed to their simplicity, support for fault-tolerance, and scalable performance. The computation is typically split into multiple compute elements (tasks); each task is (potentially) executed on a different machine for scalability and performance. Fault-tolerance is achieved through deterministic replay of failed tasks. In the absence of side effects, re-executed tasks deterministically produce the same outputs, thus providing well-defined semantics.

MapReduce, however, does not support data sharing across concurrent computations, limiting its applicability to embarrassingly data-parallel applications. This severe limitation is mainly because deterministic-replay based fault-tolerance mechanisms fail in the presence of side

effects (e.g., writes to persistent storage or communication over the network) or non-deterministic operations (e.g., using a random number generator). Consider a map function writing to the underlying distributed file system. If this instance is replayed (in case of a fault), the re-execution is oblivious of the previous write and hence rewrites the data. Both of these writes are, however, visible to external processes leading to non-deterministic behavior. For this very reason, the MapReduce framework disallows side effects.

The application scope of MapReduce and related models can be greatly improved by allowing communication/data sharing across concurrent computations. Data sharing through shared address space (e.g., a shared disk-resident key-value store) enables speculation and task-parallelism. Speculative-parallelism (or amorphous data-parallelism) is where parallel execution of tasks can lead to potentially conflicting concurrent computations. Though non-conflicting computations can be executed in parallel, these conflicts can only be detected at runtime. Exploiting this form of parallelism requires communication across computations to detect and resolve potential conflicts. Boruvka's minimal spanning tree, Dijkstra's single-source shortest-path, etc. are examples of applications exhibiting speculative-parallelism.

This project will extend the applicability of MapReduce and related models by allowing data sharing across concurrent computations. Towards this goal, we propose (1) the use of distributed key-value stores as the underlying storage for MapReduce, and (2) semantics for transactional execution of computations for supporting side effects over this shared address space [29-30]. In addition to allowing data sharing for wider applicability, the use of distributed key-value stores can also improve the performance of regular data-parallel applications by reducing the synchronization overheads. A subset of MapReduce applications (e.g., WordCount), particularly those in which the reduce phase performs simple aggregation, can be implemented as Map-only jobs when run over key-value stores. However, it is non-trivial to run MapReduce with side effects on shared address-spaces. Bridging the disparate fault-tolerance mechanisms adopted by the storage and computation layers presents significant technical challenges relating to definition of semantics, efficient implementations, and application integration.

This project will make the following specific contributions: (1) We will explore the use of distributed key-value stores as the underlying storage for data-centric programming models like MapReduce for performance. (2) We will explore the need for data sharing among concurrent computations for wider applicability of these models. (3) We will devise semantics for transactional execution of computations (map/reduce functions) over distributed key value stores, using primitives adapted from Software Transactional Memory (STM) literature. By restricting side effects only to the key-value store, we derive effective mechanisms for avoiding the consistency problems associated with deterministic replay. In our model, upon completion, results of one computation (writes to the global key-value store) become atomically visible to other computations, and to other concurrent jobs. (4) We intend to build a prototype-implementation of our semantics in the context of MapReduce over Bigtable [28]. We will explore design decisions in terms of concurrency (optimistic/pessimistic), fault-tolerance, and performance optimizations. (4) We will support our claims of performance and enhanced application scope in the context of diverse speculative-parallel applications; we address both application classes --- (1) those that cannot be expressed in the current MapReduce framework, and (2) those with limited data-parallelism in the current framework.

Team: Ananth Grama and Suresh Jagannathan

Milestones: Year 1: Design and Semantics. Year 2: Implementation of distributed key-value stores and optimizations. Year 3. Benchmark analysis and applications.