Executive Summary: Collaborative Information Systems

Bharat Bhargava (bb@cs.purdue.edu)
Department of Computer Sciences, Purdue University

1 Summary of Proposal

Information sharing is a critical element in a large number of social, governmental and military activities. It involves the distribution, discovery and aggregation of information generated or requested by multiple parties participating in an activity either explicitly or implicitly. We propose a framework for decentralized and collaborative information sharing, where different parties (peers) are able to share information with each other, without having to rely on any centralized repositories. Collaborative information sharing is expected to be more secure, efficient, fault-resilient, and trustworthy. The autonomy of each party will be maintained, and the load at each party will be balanced. Meanwhile, there will be no single points of failure or vulnerability in the system. An appropriate degree of information redundancy and diffusion will make our framework highly prepared and survivable when facing malicious attacks or other natural/social disasters.

Center for Collaborative Information System will integrate research in databases, distributed systems, networks, peer-to-peer cooperation, security, user interfaces and visualization, and social computing. The research will specifically target situations in homeland security such as terrorist attack prevention, emergency reactions, and disaster recovery.

Fundamental Research Ideas and Approach Our framework is initially inspired by the peer-to-peer communication paradigm. In the peer-to-peer model, the communicating parties, now called peers, share data and computation services among each other. In collaborative information sharing, the collaboration takes place in three ways: (1) between information providers and requesters, (2) among information providers, and (3) among information requesters, all in a de-centralized fashion. This poses new challenges in maintaining system autonomy, trustworthiness, cost-effectiveness, and survivability. Unfortunately, current research in peer-to-peer communication lacks an in-depth investigation into the fundamentals of collaboration; and little insights are available to address the new challenges in collaborative information sharing.

We therefore move beyond peer-to-peer communication, and propose a number of novel ideas for collaborative information sharing. These ideas will be reflected in the algorithms and protocols in the system. The key novel ideas include: (1) *proactive information diffusion*, (2) *collective and opportunistic information aggregation*, (3) *reliable information storage in unreliable environments*, (4) *economics-inspired information sharing incentive*, (5) *integrity and trust maintenance*, (6) *decentralized information mining and event co-relation*, (7) *isolation of malicious participants*, (8) *collaboration-oriented user interfaces and visualization*, (9) *collaboration in mobile ad hoc environments*, and (10) *social and organizational networking* In fact, some of the ideas have yielded initial but promising results, and are currently being evaluated in a wide-area and multi-institution testbed on a 24/7 basis.

Potential Applications The proposed investigation will contribute towards an in-depth understanding of collaborative information sharing among a variety of institutions such as schools, hospitals, government agencies, airlines and airports, and industrial plants. Some of its potential applications include:

- *Medical data* Multiple hospitals and health-care providers collaborate in research or patient care. Their collaboration requires sharing data while preserving anonymity of patients. An example is an ongoing research at Indiana University Medical School on the sharing of medical records for more effective diagnosis and treatment.

- *E-commerce, e-government, and remote education* Electronic commerce among buyers and sellers in supply chains requires collaborative information sharing of databases involving price, inventory and shipping data. When multiple suppliers and customers are involved, the databases need to interoperate in a cooperative manner yet maintaining autonomy and competitiveness.

- *Social, military, commercial, and governmental cooperation* The *first* example is bio-security and disaster management. During a biological or chemical attack, multiple agencies of the government, hospitals, pharmacies and Red Cross volunteers must all coordinate information about evacuation sites, medical care, supplies of safe food and water, and maintain calm and order. This requires a collaboration for all partners in disaster recovery. The *second* example is coordination of operations of the Armed Forces and the ally forces in dynamic military environments, which requires

autonomy-aware sharing of intelligence data and resources. The *third* example is the integration of information resources of multiple agencies (including CIA, FBI, INS, IRS) involved in developing fundamental principles and public policies for homeland security. However, such integration should not create any single point of potential intrusion and privacy violation.

2 Research Collaboration

Expertise of the PI's and the Personnel This research will be based on collaboration of researchers from different departments at Purdue and other universities. It will integrate the expertise of Professors Bharat Bhargava (distributed database management systems and trust models), Catherine Rosenberg (P2P and networking), Dongyan Xu (distributed systems and P2P), David Yau (Quality of Service), Philipp Afeche (Economics at Northwestern), Suresh Jagannathan (Storage systems), Gopal Pandurangan (P2P algorithms), Charlie Hu (P2P infrastructures), Karthik Ramani (supply-chain and manufacturing), Clements McDonald (Distinguished professor of medicine at IU, medical record and bio security), and Dr. Leszek Lilien (distributed database systems and security). Faculty in Envision center will also be involved in user interfaces and the visualization of exchanged data.

International Collaboration We plan to collaborate with Professor Evaggelia Pitoura at the University of Ioannina, Greece. She is an expert in broadcast models for data distribution. We will work with Professor Elisa Bertino at the University of Milan, who has worked for many years on formal security models. We will collaborate with Professor Elisa Bertino on developing formal models of security and trust in dynamic peer-to-peer collaborative environments.

Brief Plan of Work We plan to design, implement, and evaluate the framework for collaborative information sharing. Our technical goal is to achieve high autonomy, trustworthiness, scalability, usability, and survivability. Furthermore, our research aims at making scientific, engineering, as well as social impact, by integrating expertise in multiple disciplines such as computer systems, telecommunications, security, economics, social computing, and medical informatics.

We plan to extend current testbeds in (1) type of information: the shared information will span a wide spectrum such as medical records, bioinformatic sequences, sensor data streams, and event notifications. (2) diversity of participants: the testbed will be open to our research partners - hospitals, research institutes, companies, and even individuals. (3) collaboration pattern: collaborations will involve complex patterns such as patient care workflow, parallel data mining/event co-relation, and multi-round information trading and negotiation. Based on this testbed, our proposed technical ideas will be extensively tested and evaluated. Furthermore, real-world information sharing systems can be easily constructed by re-using the components in the testbed. Especially, we plan to implement multiple sample applications in the domain of homeland security and disaster recovery, as described in Section 1.

The major design of the proposed framework will be completed in years 1-3. Then, in years 4-5 we will significantly expand the testbed and system prototype, and distribute our results and software. Publications and presentations are already in progress. Furthermore, we plan to distribute knowledge to law enforcement agencies through seminars and conferences at Purdue as well as nation-wide. Both our technical findings and testbed infrastructure will be made available to researchers, practitioners, undergraduate and graduate students, and the general public.

3 Management Plan

We will employ a flexible administrative structure which involves all the interested entities, i.e., Purdue faculty from Computer Science and Electrical and Computer Engineering who are also involved in the Center of Wireless Systems and Applications (CWSA), EPICS, and CERIAS, Envision Center as well as staff from ITaP, faculty and physicians from IU Medical School, and faculty from Northwestern Kellog School of Management. The proposed center will be governed by a Center Management Committee (CMC) consisting of one member from each participating organization. The CMC will meet quarterly to discuss the overall direction of the Center, to present scientific and technical results of general interest, and to examine collaboration on strategic funding opportunities. Prof. Bhargava will serve as the Center's director and will also be responsible for day-to-day management. The responsibilities of the CMC are:

- To provide direction for the Center and to manage Center activities.
- To develop a Center Annual Program Plan (CAPP) and to track research progress against the plan.
- To address institutional membership changes as appropriate.

- To address leadership changes as needed in the technology transition areas and research areas

In addition, the CMC will establish a Center Advisory Board (CAB) to monitor the overall progress in achieving the goals of the program. The advisory board membership is drawn from prominent individuals in academia and industry in the USA and abroad. The responsibilities of the CAB are to advise the CMC regarding research directions, technology transition, and intellectual property. We will budget funds to allow the advisory board to travel to West Lafayette once per year to attend a board meeting. The advisory board will also be available at other times during the year via conference call.

The proposed center will be organized into several research areas and technology transition areas. Teams will be identified for each of these focus areas. There will be considerable overlap in the membership of the focus area teams to reflect both the natural overlap in the technical issues and the desire to foster interaction across areas (particularly between research and technology transition). The responsibilities of the research area teams and the technical transition area teams are:

- Support the CMC in developing and refining the Center Annual Program Plan (CAPP). Develop a Research Area Annual Program Plan (R-APP) for the particular research area which targets key research problems in that area and sets milestones for meeting the overall research area objectives set by the CMC.
- Ensure timely reporting to the CMC of significant technical developments and achievements. Track project progress versus plan timelines.
- Prepare recommendations for consideration of CMC for redirection of work within existing resources.
- Leverage research external to the Center.

4 Education Plan

We will develop new experimental and simulation infrastructures for teaching and public access. We plan to deploy an information portal for the sharing and exchange of research ideas pertaining to the topic of *information sharing* itself. On the education front, our research results will be included in new undergraduate and graduate courses covering the areas of cooperative systems, peer-to-peer systems, trust models, secure human-computer interaction, economic and social computing, and user interfaces and visualization.

More specifically, we regularly offer an undergraduate course in *database systems* and this research will bring security issues as a major enhancement of the course. We regularly offer a graduate course in *distributed database systems* and have offered a seminar on *peer-to-peer systems*. We plan to update our distributed database and multimedia database courses with research results from peer-to-peer and security, as a by-product of conducting this research. This is a unique effort to bring economic issues to computer science research. This provides students an environment where in addition to science and engineering, business and behavior aspects are discussed. This was particularly recognized as a necessity in the NSF PI meeting. The projects in EPICS program will be enhanced by utilizing the tools and software developed in this center.

Contributions to Undergraduate Education, Curriculum Development, and Support for Minorities This research will train five new women Ph.D. students: Ms. Yuhui Zhong, Ms. Sarika Agarwal, Ms. Yunhua Lu, Ms. Tamara Morris, and Ms. April Savoy. Tamara and April have joined us from traditionally minority black institutions. Sarika came from UC-Berkeley. Professors Xu, Yau, and Rosenberg are working actively with Ph.D. as well as undergraduate students in this area.

The PI teaches the Computer Science freshman honors seminar and has involved twelve undergraduate students through NSF REU grants. Eight of these students are Hispanics or black. The PI is a member of the science minority affairs committee, the Horizon Program (to mentor bright minority students), and the undergraduate research day committee. He will continue to involve minorities in this research program through REU grants.

5 Conclusion

Currently, the faculty involved have several million dollars of research support. Over thirteen faculty in CS, ECE, and ME at Purdue and the Kellogg Business School at Northwestern and Medical School at Indiana University are involved. This center will build upon the strength of ITaP, the Envision Center, the Center on Wireless Systems and Applications, EPICS, and Regenstrief Health Center at IU Medical School.

More information is available on the web site http://www.cs.purdue.edu/homes/bb/p2p/index.html