

Interim Report - Workshop I
20 November 2011

China US Software Workshop
Report from First Workshop
26 – 28 September 2011
Peking University, Beijing

Summary

Software is critical for enabling advances in all branches of science and engineering and is a key component of the emerging global cyberinfrastructure. In parallel, science and engineering are becoming more collaborative, requiring multi-investigator teams with a diversity of expertise in order to efficiently attack increasingly complex fundamental questions and processes, on a global scale; including societal issues such as health, energy, climate, disaster mitigation, and the environment. Solving problems on a global scale requires both that scientists work together and that their software and related cyberinfrastructure interoperate and become sustainable infrastructure for innovation through science and engineering

This report summarizes discussions from the first of two workshops to explore and catalyze long-term collaborations among researchers in China and the United States in three areas of software: trustworthy software, extreme-scale software, and software for emerging architectures. These workshops are supported by the US National Science Foundation (NSF) and the National Natural Science Foundation of China (NSFC). A second workshop will be held in March 2012 to expand upon topics of interest from the first and move another step closer to developing collaborations among participants. The goal of these activities is to accelerate the formation of mutually beneficial collaborations of China/US bilateral research teams, and to accelerate the formation of substantive software research programs that are supported in the longer term through NSF and NSFC peer-reviewed programs. More generally, it is hoped that these interactions will open the door to broader and deeper China-US collaborations in some critical areas of software and break down significant barriers between researchers in the world's two largest economies.

The technical discussions at the first workshop focused on three areas of software development: trustworthy software, extreme-scale software, and software for emerging architectures. In addition, the workshop attendees articulated the motivation for and benefits of collaborations between researchers in China and the US. Finally, the participants discussed mechanisms to support future collaborations.

This report summarizes these preliminary discussions on the technology areas, benefits of collaboration, and possible mechanisms.

Structure of the First Workshop

This workshop series has the goal of enabling researchers to begin the process of collaborative research across international boundaries. With few exceptions, the researchers invited had never met their international counterparts. The group had the initial challenge of briefing each other on their specific current research activities and areas of possible future collaboration. With 30 (15 US, 15 China) leading researcher, time did not allow for each person to give an in-depth overview of their research accomplishments. Instead we used a four-pronged approach to

Interim Report - Workshop I
20 November 2011

communicate as much technical information as possible in the shortest period of time. Day 1 of the workshop consisted of the following:

1. Participants each spent 2 minutes to introduce themselves and their broad areas of interest (approximately 75 minutes)
2. Prior to the meeting, each software area had identified leads (1 US, 1 China), who developed a 45-minute presentation representing the research of their country's participants. The US and China presentations were given back-to-back for each area (1.5 hours total per area)
3. After each pair of overview presentations, the entire workshop participated in an extended panel/group discussion of each area to clarify and begin the process of discovering mutual interest (30-45 minutes per area)

While Day 1 was quite a long day, we were able to, as a group, rapidly begin the process of discovery.

Day 2 began with a:

4. Poster session in which each participant had prepared an in-depth technical discussion of their research (75 Minutes). The poster session provided a venue for each individual to describe his or her research in depth in a one-on-one setting. This session finished mid-morning on Day 2.
5. The workshop was divided into smaller groups to further clarify areas of mutual interest, identify other researchers that could have a potential interest (for invitation to Workshop 2 and for development of collaborations beyond the workshop), begin the definition of practical mechanisms that enable bilateral collaborative research, and make follow-on plans to continue the exploration of opportunities and define "homework" to be completed before Workshop 2. Each working group¹ were given the same set of broad questions to address.

The rest of this report describes the substantive output of Workshop 1. Tables and charts are primarily the output of the working groups' focused discussions.

Technical Areas and Possible Topics for Collaboration

The workshop was structured around three broad software research themes. It was understood that these topics have overlaps, and this was re-enforced at the workshop by the desire of groups to interact. For example, trustworthy software for emerging architectures was just one such identified overlap. It was clear that keeping the group together for technical briefing in a plenary session on Day 1 was quite beneficial.

We list some of the topics of common interest to the participants. In each case, there needed to be interest expressed by both a US and a Chinese researcher to make it to this list. Participants will continue discussions between the two workshops to refine the questions of interest. These overlaps will help focus the technical aspects of the next workshop.

¹Because the areas overlapped in some cases, there was some mixing of group participants between morning and afternoon sessions of these breakout groups.

Interim Report - Workshop I
20 November 2011

In what follows, there is a listing of topics and some comments on them. In the breakout groups, research areas were identified, along with a brief description of the benefits of collaboration.

Trustworthy Software

Part of discussion reflected the wide range of topics in this software area, ranging from dependability and reliability of the software to the security of the software from external interventions.

Research Area	Benefits of Collaboration
Static Analysis of Android Native Code [Note: this is a large topic, potentially broken into multiple topics.]	Android is popular and growing, and protection needs are important.
Toward Reliable Large-Scale Software/Information Systems for Social Media	Social media is rich with large-scale software systems. Techniques for analysis, debugging/testing could help deal with reliable, large-scale software system for social media. Culture-dependent nature makes cross-cultural collaboration important. Property-mining/requirements engineering.
Data Privacy for Healthcare Systems	Requirements engineering, non-functional requirements
Reliable and Secure Applications in the Cloud	A formally verified hypervisor could incorporate fault tolerance of benign and malicious faults. Moving beyond formal methods for enhanced scalability. End-to-end guarantees.
Trust and Dependability for Concurrent and Multicore systems	Increasing prevalence of concurrency and of multicore systems.

Extreme-Scale Software

Of the three areas, extreme-scale computing quickly identified problems of interest. In part, this was expected since there were latent synergies among all participants more so in this area than the others. This resulted in very engaged technical discussions, and significant progress towards identifying directions for future workshops. Key findings include the need to target in these future workshops opportunities not only for increased dialog among the researchers engaged in extreme-scale software but across this group and the community of researchers engaged in moving forward specific domains of science and engineering through computation and data, namely the “application scientists”.

Interim Report - Workshop I
20 November 2011

Research Topic	Benefit of Collaboration
1. Performance Modeling and Optimization of Interprocessor Communication Interfaces	There is need for optimized of non-blocking and blocking collectives, in particular all-to-all. These can benefit a very broad range of systems frameworks and applications for both China and US collaborators)
2. Optimized Sparse Computations	Sparse matrix and graph operations including solvers, kernels such as SpMV, partitioning and load balancing, Benefits the large community of physics-based modeling and simulation applications.
3. Data, Viz, and Analytics	Large data analysis, machine leaning, graph and network mining, developing and refining models from data, verification and validation, data and information visualization. These are critical for design through simulation and the creation of knowledge from data across a wide range of science and engineering domains.
4. Energy-Aware Performance Optimization	There is a need to explore the tradeoffs between performance, energy, resiliency and accuracy of solution to co-manage multiple measures at multiple scales including at facilities scale, system cross-layer (application through h/w) scales, within application, etc. This is primarily from the viewpoint of optimizing facilities and systems efficiencies for enhanced science and engineering outcomes that are achieved at lower costs of operations.
5. Parallel and Distributed Applications Development Environments	Code capability enhancements, task-based enhancements, optimizations for coupled wide-area and local computing; distributed data sharing and software lifecycle managements, These are critical for sustainable software infrastructure for collaborative science and engineering.
6. Resilience	Managing system level through algorithm level resilience. This is critical for achieving correct results and managing performance degradations across most applications from increasing rates of both permanent and transient hardware errors (that are caused by technology trends and very large system scales).

Software for Emerging Architectures

This discussion reflected both existing software architecture issues as well as emerging architectures, with a great emphasis, a priori, on the word “emerging.” Expectably, these emerging systems also have requirements related to extreme-scale and trustworthy computing, but the focus is on a holistic view of system-level issues and their management. The following research topics arose from the discussions:

Research Area	Benefits of Collaboration
<p>Cloud computing software and abstractions for high-performance computing (HPC). Among others, subtopics include the following:</p> <ul style="list-style-type: none"> • Parallel programming models and heterogeneous scheduling (GPUs) • Data management including File IO in parallel systems/HPC • Virtualization • Source-to-source translation 	<p>Powerful cloud programming environments and computational primitives, such as MapReduce, need to be redesigned in order effectively use clouds for HPC applications; collaboration enable access to diverse applications and infrastructures, thus exposing requirements and test cases needed for designs that are generally applicable and useful.</p>
<p>Middleware for management of cloud-oriented infrastructure:</p> <ul style="list-style-type: none"> • Energy concerns • Middleware/runtime and programming environments • Systems that use cloud components • Virtualization layer for HPC 	<p>Management of cloud infrastructure and systems built with components from distinct cloud must take into consideration multiple operational costs and service level agreements. The inherent, likely transnational, distributed nature of these systems can be best accounted for when using software approaches and infrastructure hosted in different countries.</p>
<p>Fault-tolerance in large-scale systems. Among others, subtopics include the following:</p> <ul style="list-style-type: none"> • Automatic fault identification • Risk assessment in software • Self-healing 	<p>Failures are common events in large-scale systems. The software used in large-scale systems must operate correctly regardless of faults. Some of the largest HPC and social networking systems in the world are in China and the USA.</p>
<p>Autonomic software architectures for dependable cloud computing and social networking. Some of the topics overlap with those discussed in the trustworthy software group.</p>	<p>Complex emerging systems are too hard and expensive too manage unless (1) feedback is used to automatically adapt them to new requirements, environment changes and components provisioned by third-parties, and (2) self-organization and discovery enables them to scale. Automatic adaptation requirements are exacerbated when considering transnational systems. This is an area where researchers from the US and China have complementary expertise.</p>

Motivation for / Benefits of Collaboration

“The primary driver of most collaboration is the scientists themselves. In developing their research and finding answers, scientists are seeking to work with the best people, institutions and equipment which complement their research, wherever they may be.” [KNN]²

The discussions at the workshop identified several motivations to and benefit of collaborations, many of which reflect the findings in the report cited above, but several of which are unique to the collaboration between China and US researchers in the areas of software covered by this workshop.

Problems and Ideas: Collaboration among researchers in China and US allows two communities to focus on a class of problems beyond the research of either group, yet that affect both groups. Put another way, China-US bilateral collaboration brings more intellectual power to common problems important to both countries than each country working unilaterally.

While several applications areas were mentioned at the workshop, only a few applications areas were discussed at any length (albeit briefly) at this workshop. However, there was a strong sense that we should focus on large application problems affecting both countries. Furthermore, these problems should generally fall into the broad category of ones that researchers from one country could not solve on our own. The following list is illustrative rather than exhaustive and should be expanded in the future:

- Energy consumption of large systems and the need for energy-efficient software design.
- Climate change and adaptation,
- Hazard and risk management

Complementary Expertise and Resources:

A benefit of China-US bilateral collaboration is that there are natural complementary assets for collaborations to leverage. The sharing and leveraging of these benefits reduces the need for duplication, leading to the benefit of all involved, allowing researchers to address larger problems and making the research more cost-effective. Initial discussions at Workshop 1 identified the following opportunities to share resource and expertise: :

- Infrastructure:
 - Access to unique infrastructure can provide researchers with research opportunities not possible in their own country
 - Test software on systems at a scale not possible otherwise (e.g. access to largest IPv6 network in China)
 - Prototype software on unique infrastructures (e.g. new CPU, new software architectures)
 - Compare different approaches to new phenomena (e.g. social media)
 - Gain access to and share applications, code bases, systems and data sets to develop and test new ideas

² Knowledge, networks and nations (KNN): Global scientific collaborations in the 21st century. The Royal Society. ISBN: 978-0-85403-890-9, March 2011

Interim Report - Workshop I
20 November 2011

- Providing a diverse set of design points

It is worth noting that access to unique resources benefits both parties in a collaboration, since the tests run on the platform will help inform future designs, make those systems more usable, create new research ideas, and lead to publications.

- Expertise:
 - Access to leading researchers stimulates and challenges other researchers to think more deeply about scientific questions. Also, different individuals bring different perspectives – which when combined allows collaborations to address problems not tractable to either group individually.
 - In areas of common priorities (e.g. extreme scale) collaborations can tackle problems of joint interest (e.g. trade-off between energy and performance);
 - In areas of different priorities (e.g. cloud computing), collaborations can indicate both trends and can recruit new expertise to the areas.
 - In areas of different approaches (e.g. social media), there are opportunities to compare approaches.
 - Access to software engineers, which may be more available in China

Benefits to Students, Researchers, Institutions

- Students: Increasingly, we are working in an interconnected world. International experiences allow students to be exposed to different approaches to research and to education that is a reflection of different cultural approaches to research. Engaging students as early as possible in international collaborations will help prepare them for their careers, whether in academic, industrial, or government positions.
- Faculty researchers: Collaboration allows faculty to increase the pool of talent they can work with. This can lead to the ability to tackle more challenging problems, in turn leading to better papers, increased career development including more success with funding and promotions, attracting brighter students, and having a broader impact.
- Institutions: Institutions benefit by an increased international reputation, having more successful faculty (publications, funding), and a stronger program for students (better students).

Research Risk and Opportunity: Collaborations bring both risk and opportunities.

- Opportunities include those listed above (benefit) as well as understanding research trends in the different countries.
- Risks are inherent in collaborations if they are not well conceived, or if they do not provide mutual benefit and demand mutual contributions. These risks can be mitigated by ensuring that these issues are thought through at all stages of the collaboration.
- International collaboration provides a unique means to balance research risk with broader impacts of learning about trends as well as exposing students to international interactions. In particular, the collaboration may tolerate greater research risk because of the other benefits of the collaboration. In turn, greater risk can lead to potential greater reward.

Impact: Bilateral collaboration can produce results of greater impact than working unilaterally:

- Publications can have increased citation impact (KNN, and other studies³)
- Software produced from joint collaborations could have broader user base and be disseminated more broadly
- Software developed jointly could be applied to problems shared by both societies – if driven by users, the adaptation would presumably be faster than adaptation of software developed unilaterally.

Mechanisms to Support Collaboration

Participants in the workshop discussed mechanisms to support collaborations in the areas discussed. The ideas fell into two temporal categories: the time between the two workshops (i.e. between the end of September 2011 and the beginning of March 2012); and the time after the second workshop. A premise of the workshop is that collaborations don't just "happen", they need real mechanisms for long-term support if they are to be sustainable.

Between the Workshops: Participants agreed to continue discussions and identify more concretely problems of mutual interest. Three particular types of activities were identified:

- Organize a session at the 2011 Supercomputing Conference to continue discussions, in particular in extreme-scale software. This is underway and will occur on the Monday November 14th
- Determine opportunities for both physical and virtual visits among smaller group of potential collaborators and arrange the visits. Specific visits are still in process of being defined. An example visit already planned is a visit by Prof. Zhi Jin to Rutgers and University of Florida.
- Develop technological mechanisms, such as Google+/social network communication, mailing list, or blog for match-making and plans for proposals. An initial test with Diaspora as an engine is being created by Felix Wu, UC Davis.

After the Second Workshop: Ideas in this category focused on how to build broader dialog between the communities, train or exchange students, and fund joint research and development projects. These identify an expansive set of possible collaborative mechanisms, some of which are quite time- and cost-effective. These mechanisms reflect ideas discussed at the first workshop. They also reflect that collaborations take on different forms and take time to develop. In the subsequent section there are several NSF funding mechanisms that could support US participants in a China-US collaboration. Some of these collaborative mechanisms will be developed further at the next workshop.

- Building Broader and Sustained Dialog
 - Develop a series of regular summer schools lasting for about 2 weeks, over a period of at least 3 years, to allow both groups to interact and learn. Summer

³ See the publication metrics in the following study,
http://www.forskningsradet.no/en/Newsarticle/Increased_copublication_with_the_United_States_and_Canada/1253968889036?lang=en

Interim Report - Workshop I
20 November 2011

schools should include graduate students and postdocs. A thematic approach with a venue for publishing lectures as a book might be desirable.

- Pursue exchanges/short visits (students, scholars) to pursue specific questions
 - Conduct virtual meetings w/longer presentations – and more focused content
 - Develop technical (week or longer) workshops
 - Construct and submit a workshop report to IEEE Computer, at the right time, to discuss the opportunities for China-US collaborations
- Train or Exchange Students
 - Develop opportunities to exchange students. Students are often the conduit of ideas and a key component of a collaboration.
 - Develop plan for student co-mentoring / joint supervisors. The longer-term intent could involve closer collaborations between institutions via jointly offered courses or ultimately joint degree programs.
 - Joint Research Funding
 - Work with funding agencies to create opportunities for coordinated submissions/ review/ funding. Community support for joint programs could encourage dialog between agencies for specific programs of mutual benefit.
 - Submit proposals to support joint projects.
 - Leverage supplemental funding (and at NSF the EAGERS)
 - Strategic Directions Umbrella program
 - Work with NSFC and NSF to introduce a “strategic directions” umbrella program. This special umbrella program could host a portfolio of activities including student training and funding for specific research project etc. Funding should be commensurate with the scope and scale of the umbrella program and its strategic value.

Existing Funding Mechanisms Available via NSF

Several funding mechanisms were mentioned by NSF staff who attended the workshop. We list several of these below, with links to the urls. We note that there are at least three categories of funding opportunities:

1. Explicitly international by design: Activity in this category include support for US-based graduate students to visit international partners (e.g. EAPSI), for advanced institutes (e.g., PASI), or for virtual institutes (e.g., SAVI).
2. General programs that can be useful to support US researchers in an international collaboration. Activities in this category include Research Coordination Networks (RCN), Research Experiences for Undergraduates (REU), regular research programs, solicitations (e.g., SI2) and other mechanisms (e.g., EAGER).
3. Program that support US researchers within a framework of collaboration that underpins the activities. An example of this is the Dimensions of Biodiversity.

We list several of the activities that were mentioned in the workshop or in the paragraph above.

- East Asia Pacific Summer Institute (EAPSI): Deadline **9 November 2011**

Interim Report - Workshop I
20 November 2011

- http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5284
- Provide support for graduate students in science and engineering during the course of the summer
- The graduate student is the PI! And must be a US citizen or permanent resident.
- Check site for eligible institutions in China (more institutions were added this year; although China only accepts 40 students per year)
- Advanced Studies Institute (aka PASI): Deadline **15 January 2012**
 - http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5327
 - New solicitation expected before deadline
 - Provides support for intense training experiences for US and in this case Chinese students and postdocs
 - This does allow for awards to countries outside of its named “Pan-American”, but please contact the program officer
- Research Experiences for Undergraduates (REU): both sites as well as individual students
- SI2 Software Infrastructure for Sustained Innovation (new solicitation not released)
 - <http://www.nsf.gov/pubs/2011/nsf11539/nsf11539.htm> (old solicitation - needs to be rewritten)
 - Conceptualization Awards:
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503489&org=NSF&sel_or_g=XCUR&from=fund Deadline **14 December 2011**
 - Two components of the SI2 that might be relevant in the short term for participants of these two workshops are the following:
 - Scientific Software Elements (SSE): SSE awards target small groups that will create and deploy robust software elements for which there is a demonstrated need that will advance one or more significant areas of science and engineering
 - Scientific Software Integration (SSI): SSI awards target larger, interdisciplinary teams organized around the development and application of common software infrastructure aimed at solving common research problems. SSI awards will result in sustainable community software frameworks serving a diverse community
- SAVI - Science Across Virtual Institutes (SAVI)
 - Details announced on 5 October 2011
 - <http://www.nsf.gov/pubs/2011/nsf11087/nsf11087.jsp>
- Dimensions of Biodiversity (<http://www.nsf.gov/pubs/2011/nsf11518/nsf11518.htm>)
 - The relevance of this solicitation is that it gives an example of a framework for collaboration between China and US researchers in a focused area of research.
- Research Coordination Networks
 - http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=11691
 - This mechanism allows for five years worth of interactions to help develop research areas
- Other NSF funding mechanisms for this group to consider:
 - Supplements to existing awards
 - EAGER - EARly-concept Grants for Exploratory Research

Interim Report - Workshop I
20 November 2011

Participants

US Participants

Coordinators

Philip Papadopoulos	UCSD	users.sdsc.edu/~phil/homepage.html
Peter Arzberger	UCSD	www.nbcr.net/parzberger.php

Trustworthy Software

Wright, Rebecca	Rutgers (Lead)	www.cs.rutgers.edu/~rebecca.wright
Alvisi, Lorenzo	U Texas	www.cs.utexas.edu/~lorenzo/
Traynor, Patrick	GA Tech	www.cc.gatech.edu/~traynor/research.html
Wu, Felix	UC Davis	www.cs.ucdavis.edu/~wu/

Extreme Scale Software

Raghavan, Padma	Penn State U (Lead)	www.cse.psu.edu/~raghavan/
Chow, Edmond	GA Tech	www.cc.gatech.edu/~echow/
Gropp, Bill	UIUC	www.cs.uiuc.edu/~wgropp/
Ng, Esmond	LBL	crd.lbl.gov/~EGNg/
Patra, Abani	SUNY Buffalo	www.mae.buffalo.edu/people/full_time/a_patra.php

Software for Emerging Architectures

Fortes, Jose	U FL (Lead)	www.ece.ufl.edu/people/faculty/fortes.html
Dinda, Peter	Northwestern	www.cs.northwestern.edu/~pdinda/
Figueiredo, Renato	U FL	byron.acis.ufl.edu/~renato
Parashar, Manish	Rutgers	nfscc.rutgers.edu/people/parashar/
Qiu, Judy	Indiana U	www.soic.indiana.edu/people/profiles/qiu-judy.shtml

NSF Participants

Gabrielle Allen, Emily Ashworth	OCI NSF Beijing Office	http://www.cct.lsu.edu/~gallen/ http://www.nsf.gov/od/oise/beijing/nsf-beijing-ofc-about-us.jsp
Nancy Sung	OISE	

PRAGMA Liaison

Cindy Zheng	UCSD	http://www.sdsc.edu/~zhengc/zhengc.html
-------------	------	---

China Participants

Trustworthy Software

Jin, Zhi	Peking U. (Lead)	http://www.sei.pku.edu.cn/people/zhijin/
Dong, Wei	NUDT	http://www.nudt.edu.cn
Gu, Ming	Tsinghua U.	http://info.thss.tsinghua.edu.cn/docinfo_out/index.jsp
Zhang, Jian	ICS, CAS,	http://lcs.ios.ac.cn/~zj/

Interim Report - Workshop I
20 November 2011

Zhao, Jianjun Shanghai Jiao Tong University
<http://cse.sjtu.edu.cn/~zhao/>

Extreme-Scale Software

Qian, Depei Beihang U. (Lead)
<http://scse.buaa.edu.cn/teacher/10.html>,
Chen, Wenguang Tsinghua U.
<http://hpc.cs.tsinghua.edu.cn/research/cluster/cwg.html>,
cwg@tsinghua.edu.cn
Chen, Yifeng Peking U. (PI and Local Organization)
sei.pku.edu.cn/~cyf,
Feng, Xiaobing ICT, CAS
[http://sourcedb.cas.cn/sourcedb_ict_cas/](http://sourcedb.cas.cn/sourcedb_ict_cas/en/eictexpert/fas/200909/t20090917_2496613.html)
[en/eictexpert/fas/200909/t20090917_2496613.html](http://sourcedb.cas.cn/sourcedb_ict_cas/en/eictexpert/fas/200909/t20090917_2496613.html),
Zhang, Yunquan ICS, CAS <http://www.rdcps.ac.cn/~zyq>

Software for Emerging Architectures

Guo, Minyi Shanghai Jiao Tong University (Lead)
<http://epcc.sjtu.edu.cn/?q=node/34>,
Lu, Yutong NUDT <http://www.nudt.edu.cn>
Peng, Xin Fudan University <http://www.software.fudan.edu.cn>,
Zhang, Lu Peking U <http://sei.pku.edu.cn/~zhanglu>,
Zhou, Yuming Nanjing U. <http://cs.nju.edu.cn/zhouyuming>

NSFC Participants

Zhaotian Zhang, Dept. of Info. Sci.
Huai Chen Bureau of International Cooperation
Ke Liu Dept. of Info. Sci.
Xiaochun Cao Dept. of Info. Sci.,
Xiuping Liu Bureau of International Cooperation
Qing Wei Bureau of International Cooperation