Statement of Purpose

Andrew Pavlo

A distinguishing characteristic of distributed computing research is that it incorporates facets from many other fields in computer science. While there are challenging problems at the lowest system level, there are also problems that require knowledge of network and application levels. This broad encapsulation, as well as the problem domain itself, arouses my interest the most. My academic and professional career so far has allowed me to directly observe and apply key tenets of distributed systems; yet there are many fundamental principles that I still need to learn. Graduate school will not only allow me to apply these principles to new and exciting problems, but also to develop my own techniques all throughout my career. I know that my experiences, from my independent projects to my current work with a world-class research team, make me a strong candidate for graduate studies at the University of Wisconsin-Madison.

My initial foray in formal academic research concerns interactive graph visualization. With support from Dr. Christopher Homan, my master's thesis advisor, we developed an experimental graph visualization system that displays large, complex graphs as a series of spanning trees [Pav06, PHS06]. Our algorithms generate transitions that minimize crossings between edges and minimize disruptive movement of nodes between similar, non-isomorphic trees of the same graph. I still continue to conduct research with my advisor even though I have moved on to another university [HPS06]. I hope to improve the aesthetic qualities of our new drawing algorithm and to develop a more meaningful animation algorithm before I begin the next stage of my formal education.

Although I enjoy my work in graph visualization, my main research interests are now focused elsewhere. Since September 2005, I have been a member of the Condor research group lead by Dr. Miron Livny at the University of Wisconsin-Madison. My work at Wisconsin spans the three major projects in our group: (1) the Condor distributed batch system, (2)

NSF's Middleware Initiative Build & Test Laboratory, and (3) the Virtual Data Toolkit. Working for the Condor project is often hectic, and research on certain problems may only last for a month before one is switched to another task; our priorities change quickly based on the needs of our various collaborators and users. For example, just on the Condor system itself, in the last year I investigated time synchronization issues between daemons, made contributions to a new execution environment for managed jobs running on submit nodes, and assisted in porting Condor to BSD-based operating systems.

While Condor is the most established project of the three that I work on, one newer research system I help to develop is the NMI Build & Test Laboratory [PCG+06]. The NMI system is a framework for building and testing software in a heterogeneous, multi-user, distributed computing environment. Unlike Condor, which has an established community and a mature code base, the NMI project is still quite new and there are many interesting problems still unsolved. One of my notable contributions was developing job migration mechanisms in the NMI system in collaboration with CERN and the INFN in Europe. This allows build and test submissions to automatically execute outside of local administrative domains based on resource availability. Another important feature that I developed was integrated support for the TeraGrid's dynamic environment management system. I am also excited about the possibilities of my research for the NMI system in the upcoming year: expanding the framework to include more robust configuration provenance of computing resources and investigating new methodologies for testing distributed computing software.

Lastly, I am also a member of the Virtual Data Toolkit operations team, under the guidance of Dr. Alain Roy and Dr. Timothy Cartwright at Wisconsin. This project is responsible for integrating and supporting the software stack deployed on computing resources in the Open Science Grid. Working with Dr. Roy provides an invaluable understanding of the components that make up a large-scale grid infrastructure.

My time with Dr. Livny has taught me the importance of viewing distributed computing research from a social sciences perspective. That is, to understand and solve problems in computer science based on how humans interact with the world, rather than to just blindly develop new technology. Working on Condor has also motivated me and intensified my desire

to pursue an advanced degree in computer science.

Because of my positive experiences with the Computer Sciences department here at Wisconsin, I am excited about the possibility of transitioning from academic staff to a full-time graduate student. This would provide me the opportunity to continue further with my Condor-related research, but also to work with other systems faculty members, such as Drs. Remzi and Andrea Arpaci-Dusseau, Dr. Barton Miller, and Dr. Michael Swift. Furthermore, I plan to remain an active participant in the systems reading group and seminars, and become more involved with the graduate student community.

After earning my doctorate, I will pursue an academic career at a research institution. Thus, it is imperative that I continue to work hard and fulfill my educational goals. I would also like to experience teaching while in graduate school. My time at Wisconsin has proven to me that the Computer Sciences department is the ideal environment for nurturing my research endeavors. I look forward with anticipation to the many challenges that I will experience in the upcoming years.

References

[HPS06] Christopher Homan, Andrew Pavlo, and Jonathan Schull. Smoother transitions between breadth-first-spanning-tree-based drawings. In *Graph Drawing, 14th International Symposium, GD 2006, Karlsruhe, Germany, September 18-20, 2006, Revised Papers*, Lecture Notes in Computer Science. Springer, 2006.

[Pav06] Andrew Pavlo. Interactive, tree-based graph visualization. Master's thesis, Rochester Institute of Technology, 2006.

[PCG+06] Andrew Pavlo, Peter Couvares, Rebekah Gietzel, Anatoly Karp, Ian D. Alderman, Miron Livny, and Charles Bacon. The NMI Build & Test Laboratory: Continuous integration framework for distributed computing software. In *Proceedings of the Large Installations Systems Administration (LISA) Conference*. Usenix Association, 2006.

[PHS06] Andrew Pavlo, Christopher Homan, and Jonathan Schull. A parent-centered radial layout algorithm for interactive graph visualization and animation. Technical Report `http://arxiv.org/abs/cs/0606007`, arXiv.org, 2006.