Virtually-Organized Interdisciplinary Teams using Distributed Cyberinfrastructure: A Performance Model

Sara Kraemer
Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 W. Johnson Street, Madison, WI, 53704
sbkraeme@wisc.edu
+608-265-5624

Christopher Thorn
Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 W. Johnson Street, Madison, WI, 53704
cathorn@wisc.edu
+608-263-2709

ABSTRACT

This paper presents a model of virtual team performance, specifically concentrating on multi- and interdisciplinary research teams' use of high-throughput computing (HTC) for computational research. We define HTC as environments that can deliver large amounts of processing capacity over long periods of time and are designed to be extremely fault-tolerant and require minimum human intervention. There is increasing evidence that HTC—and grid technologies in general—are ubiquitous in and mission-critical to interdisciplinary research that requires access to distributed resources, ranging from large amounts of computing power to targeted expert advice. The effective performance of these virtual organizations depends on not only on an in-depth understanding of the technical characteristics of HTC and the ways in which those characteristics affect virtual team interaction, but also on social and organizational factors that influence virtual team performance in distributed computational environments.

Author Keywords

Virtual teams, sociotechnical systems, high-throughput computing, interdisciplinary.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Computer-supported collaborative work

INTRODUCTION

There is increasing evidence that HTC—and grid technologies in general—are ubiquitous in and mission-critical to interdisciplinary research that requires large amounts of computing power as well as access to expert advice [1]. We define HTC as an environment that can deliver large amounts of processing capacity over long periods of time. In addition to computational power delivered, there is a second, critical measure of system quality: HTC systems are designed to be extremely fault-tolerant and require minimum human intervention [2]. By design, these technologies enable and support distributed teams. These and other characteristics drive the interactions and forms of collaboration that emerge when users from various scientific domains use HTC resources to work on computational problems. Virtual teams in high throughput computing may vary across: time and geography, domains of science, team size, background or culture, type of task, type of research problems (e.g., applied, basic), computational needs, fluidity of membership in the HTC community, and degree of interdisciplinarity either within their scientific domain and/or across research projects.

Some HTC systems, such as the Condor Project software at UW-Madison, have unique characteristics that foster interdisciplinary virtual team collaboration, such as high degrees of resource flexibility, end-user control, open-ended planning, and distributed resource management [3]. HTC generally, and Condor specifically, are drivers of leading-edge science in local research teams and in large-scale, internationally distributed production environments.

VIRTUAL TEAM PERFORMANCE IN HIGH THOUGHPUT COMPUTING

A virtual team is a group that works across time and distance and whose interactions are mediated by technology [4]. Although there are differences in the types of technology and communication enabled in virtual team environments, the core feature of a virtual team is that it is one in which interdependent group members work together on a common task while they are spatially separated. Team members may be geographically and temporally dispersed; permanent or nonpermanent; members of different organizations, countries, or cultures. They can meet partially or fully in cyberspace. *Virtual teams in high throughput computing* may vary across:

time and geography, domains of science, team size, background or culture, type of task, type of research problems (e.g., applied, basic), computational needs, fluidity of membership in the HTC community, and degree of interdisciplinarity either within their scientific domain and/or across research projects.

Virtual team research has emphasized the study of electronic communication technologies and processes as mediators of virtual team performance [5]. Communication technology and virtuality contribute to the transformation of teamwork in three important ways: (a) they introduce new dimensions of communication among members by breaking down traditional barriers of space and time; (b) they modify traditional group processes; and (c) they enhance a team's ability to access, share, manipulate, retrieve, and store information [6]. Virtual team communication technologies may include videoconferencing, Internet chat rooms, email, and bulletin boards [7]. However, there are other forms of technology that link distributed team members and mediate their performance, such as HTC, giving these teams unique characteristics that distinguish them from teams that use previously studied performance mediating technologies (e.g., communication technologies).

The core characteristic of HTC is its ability to provide large amounts of computing for sustained periods of time. However, HTC has a number of additional characteristics that provide greater access to a wide range of disciplines and toolsets. For example, because Condor runs on many computing platforms and can execute any software that does not require user interaction, a wide range of tools is readily available—from commercial research software to scripting engines and compilers. In addition, the abundance of available scientific tools allows individual scientists or teams to engage with the Condor HTC environment using tools familiar to them. Enabling of existing tools in an HTC setting provides critical social and technological gateways for new adopters of HTC. Access to the HTC environment also exposes new adopters to tools and methods used by others to address similar computational problems. In this way, scientists' skills and knowledge are affected by the capabilities and characteristics of HTC technologies and tools.

We argue that distributed computational cyberinfrastructure generally—and HTC specifically—are enabling technologies for virtual team collaboration. Virtual teams using HTC introduce new and novel ways to accomplish research objectives that were previously unattainable due to limitations of computational power. With HTC, the obstacles of access to computational power are removed, and scientists have access to other scientists using HTC all over the world, linked through their use of HTC.

THE CONDOR PROJECT, HIGH THROUGHPUT COMPUTING, AND SCIENTIFIC RESEARCH COLLABORATION

Currently, there are more than 32,000 known Condor hosts and more than 45 Condor pools in Wisconsin. These figures, which likely underrepresent actual use,[1] account for 32.7% of the known systems in the U.S. and 23.9% of the known systems in the world. Condor pools have been created in 38 U.S. states, including the second largest group in the District of Columbia. Condor is also deployed on 136,730 hosts in 39 countries with the second largest deployment being in Switzerland.

The Condor Project embodies a philosophy of flexibility; this philosophy has allowed the design to flourish in a highly unpredictable distributed operating environment [3]. International distributed systems are heterogeneous in numerous ways: they are composed of many types and brands of hardware; they run various operating systems and applications; they are connected by unreliable networks; they change configuration constantly as old components become obsolete and new components come online; and they have many owners with private policies and requirements that control their participation in the community. Condor has adopted a five-component philosophy of flexibility to address these barriers and enable virtual team collaboration:

1. 1. *Let communities grow naturally.* Given tools of sufficient power, people will organize the computing structures they need. However, human relationships are complex, people invest their time and resources to varying degrees, and relationships and requirements change over time. Therefore, Condor design permits but does not require cooperation.
2. 2. *Leave the owner in control, whatever the cost*. To attract the maximum number of participants to a community, the barriers to participation must be low. Users will not donate their property for the common good unless they maintain some control over how it is used. Therefore, owners of computing resources are given the tools to set policies and retract resources for private use.
3. 3. *Plan without being picky*. It is critical to plan for slack resources as well as resources that are slow, misconfigured, disconnected, or broken. The designers of Condor spend more time and resources

[1] Many users are unable to register their existence with the central Condor administrative site because of security concerns or technical limitations of their local network infrastructure.

contemplating the consequences of failure than the potential benefits of success.

4. 4. *Lend and borrow*. The Condor Project has developed a large body of expertise in distributed resource management and aims to give the research community the benefits of that expertise while accepting and integrating knowledge and software from other sources. It has also instituted a mechanism for collective problem sharing and solving among its users.

5. 5. *Understand previous research*. The Condor Project continually updates its organizational knowledge with previous research to apply well-known fundamentals as well as cutting-edge techniques to emergent problems. The inclusion of current user innovations keeps the work focused on the edge of discovery rather than wasting effort remapping known territory.

Thus, the Condor Project is much more than a complex set of computational resources. The Condor team maintains a close intellectual partnership with computer and domain scientists working together on the challenges of HTC in the context of breakthrough science. Condor has advanced HTC technology via improvements in their software coupled with innovations in computational approaches developed by a wide range of domain scientists. These interactions have made Condor team members privy to numerous sociotechnical problems affecting interdisciplinary virtual team performance.

Experiences at the Condor Project suggest that multidisciplinary and interdisciplinary collaboration around HTC resources has important sociotechnical implications for collaborative research and technology design [8]. Interesting interactions arise when the barriers to computational resource access are removed. For example, some virtual teams using HTC embody what is known as *tool-based specializations*—that is, they view their interaction with HTC as a function of their research areas—whereas other teams exemplify a *perspective-based specialization*—the HTC ceases to be part of their research area, and problem situations define their research space [9, 10]. Those operating in perspective-based specialization may be more likely to share local research methods and solutions with other team members or teams and view HTC as an enabling technology, rather than simply as resource sharing.

As HTC technologies mature, they move out of a tool-based environment and become a part of the research infrastructure. This change can lower perceived risk and enhance trust among users, thereby enabling new models of financial support and social engagement. One important contribution of the PIVOT project will be to examine how the role of HTC will change as users come to see it as a mature production technology. As the tools move out of basic research into applied research and production work, new risks and benefits are likely to appear. For example, computer scientists may perceive a risk of being identified as merely an infrastructure provider. On the other hand, for those contemplating the adoption of HTC, the emergence of HTC as a commodity service may greatly reduce their perception of risk. As HTC systems provide increasingly ubiquitous access to robust computing environments, potential users will see HTC as a relevant and important aspect of their work. The other significant element of HTC system success—fault resistance, a key characteristic of the Condor Project—is another critical enabler of adoption at the margin [2].

Perspective-based collaboration is particularly common in experimental research involving complex instrumentation, such as telescopes, particle accelerators, or CT scanners [11]. For example, high-energy particle physics is a domain that has embodied perspective-based collaboration characterized by collectivism, erasure of the individual epistemic subject, nonbureaucratic mechanisms of work, lack of overbearing formal structures, and an absence of rigid rules [12]. Similar types of collaborations and configurations have been anecdotally found by the Condor team through their experiences working closely with domain scientists. These examples are representative of possible sociotechnical components of a virtual interdisciplinary team model or typology of virtual team scientific collaborations.

VIRTUAL TEAM PERFORMANCE MODEL

Our research model (Figure 1) is a synthesis of sociotechnical systems theory and team performance models. The heart of modeling team performance lies in identifying its underlying mechanisms, such as individual characteristics, task characteristics, work structures, team characteristics, and team competencies. In a comprehensive review of 138 models of team performance, Salas et al. [13] cited the widespread use of *input-process-output* (IPO) to describe and evaluate team performance. The IPO model highlights the importance of throughputs as mediators or moderators of the relationships between input factors (i.e., work system factors) and outputs (e.g., team satisfaction, performance outcomes). The model incorporates the IPO framework, as well as components of the IPO model of virtual team functioning of Martins et al. [5] —specifically, the process categories (planning, action, and interpersonal processes) and the outcome categories (task performance, process, and affective outcomes).

The IPO model categorizes *inputs* into the five work system factors discussed above: organization, individual, task/workload, technology and tools, and external environment [14, 15]. *Processes* include various types of group interaction, such as team coordination and communication. *Output*s include two dimensions of performance: process outcomes (e.g., group dynamics, coordination) and task performance outcomes (e.g., meeting project goals, developing innovative solutions). A key assumption of the IPO model is that the input states affect group outputs via the interactions that take place among members. This model frames the description of the factors that contribute to and hinder virtual team performance.

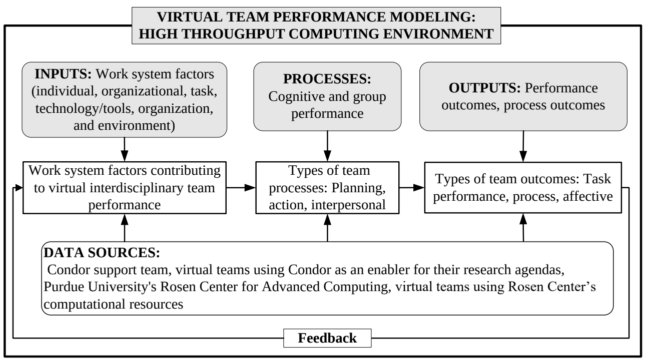

Figure 1. Virtual team performance model

In characterizing the evolution of virtual team performance, we also emphasize the role of feedback loops in the organizational lifecycle(s) of virtual teams. The feedback loop illustrates that virtual team performance is informed by past and current experience, as well as notions of future team states. In this feedback loop, we will capture the stages and causes of virtual team evolution and (re)organization and the ways in which they vary across the various IPO system constructs, such as task, scientific domain, research population, and other stages of the organizational lifecycle.

The IPO model—emphasizing the importance of interactions among inputs, processes, and resulting outputs—is appropriate for capturing the dynamic interactions and emergent states underlying team performance [13]. We chose to adopt the IPO model in order to envelop an exploratory approach to specifying the various dimensions of virtual team performance. The IPO model delivers unconstrained conceptual categories of virtual team performance.

PROPOSED RESEARCH

We are currently proposing research to identify and describe interdisciplinary team performance via team modeling techniques developed in human factors engineering. The main outcome of this research will be a model of interdisciplinary virtual team performance spanning a number of dimensions, such as: space, time, cultural backgrounds, social norms, workflows, and computational environments. This model will focus on interdisciplinary virtual teams working in distributed computing environments across computational research and scientific domains (the Condor Project and Rosen Center will serve as sampling pools). Through iterative sets of interviews and focus groups conducted with HTC computer scientists and the interdisciplinary virtual teams they serve, we aim to: (1) describe the sociotechnical factors that contribute to and hinder virtual team performance; (2) specify the process and task performance outcomes of virtual teams; and (3) model effective virtual team performance (with a specific emphasis on interdisciplinarity and collaboration).

To date, there has not been an in-depth analysis of virtual team collaboration in scientific communities, or of multi- and interdisciplinary virtual teams, specifically. We will address these gaps by exploring the sociotechnical systems aspects of virtual teams conducting scientific research with HTC. An analysis of the social impacts of the technical configurations of HTC software such as Condor will lead to deeper understanding of how HTC is used as an effective enabler of new scientific problem sets, solutions, and collaboration configurations, as well as how the HTC technology can be designed and deployed to meet emerging scientific problems and configurations.

ACKNOWLEDGMENTS

We thank Miron Livny and Todd Tannenbaum of the Condor Project at the University of Wisconsin-Madison for the comments and input into the development of these ideas and for their reviews on earlier drafts of this prose.

REFERENCES

1. Berman, F. and T. Hey, *The Scientific Imperative*, in *The Grid 2: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Editors. 2004, Morgan Kaufmann Publishers: San Francisco, CA. p. 13-23.
2. Livny, M. and R. Raman, *High throughput resource management*, in *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Editors. 1998, Morgan Kaufmann Publishers Inc.: San Francisco, CA. p. 311-337.
3. Thain, D., T. Tannenbaum, and M. Livny, *Distributed computing in practice: The Condor experience.* Concurrency and Computation: Practice and Experience, 2005. **17**(2-4): p. 323-356.
4. Driskell, J.E., P.H. Radtke, and E. Salas, *Virual teams: Effectives of technological mediation on*

team performance. Group Dynamics: Theory, Research, and Practice, 2003. **7**(4): p. 297-323.
5. Martins, L.L., L.L. Gilson, and M.T. Maynard, *Virtual teams: What do we know and where do we go from here?* Journal of Management, 2004. **30**(6): p. 805-835.
6. Godar, S.H. and S.P. Ferris, *Virtual and Collaborative Teams: Process, Technologies, and Practice*. 2004, Hershey: Idea Group Publishing.
7. Avolio, B.J., et al., *Virtual teams: Implications for e-leadership and team development*, in *How People Evaluate Other Organizations*, M. London, Editor. 2001, LEA: Mahwah, New Jersey. p. 337-358.
8. Thain, D., T. Tannenbaum, and M. Livny, *How to measure a large open-source distributed system.* Concurrency and Computation: Practice and Experience, 2006. **18**(5): p. 1989-2019.
9. Schmidt, K., *Cooperative work: A conceptual framework*, in *Distributed Decision-Making: Cognitive Models for Cooperative Work*, J. Rasmussen, B. Brehmer, and J. Leplat, Editors. 1991, Wiley: Chicester, West Sussex, England.
10. Whitely, R., *Cognitive and social institutionalization of scientific specialties and research areas*, in *Social Processes of Scientific Development*, R. Whitely, Editor. 1974, Routledge and Kegan Paul: London p. 69-95.
11. Katz, J.S. and B.R. Martin, *What is research collaboration?* Research Policy, 1997. **26**: p. 1-18.
12. Chompalov, I., J. Genuth, and W. Shrum, *The organization of scientific collaborations.* Research Policy, 2002. **31**: p. 749-767.
13. Salas, E., et al., *Modeling team performance: The basic ingredients and research needs*, in *Organizational Simulation: Capabilities for Assessing and Designing Organizational Responses to Complexity and Uncertainty*, W.B. Rouse and K.R. Boff, Editors. 2004, Wiley: New York.
14. Carayon, P. and M.J. Smith, *Work organization and ergonomics.* Applied Ergonomics, 2000. **31**: p. 649-662.
15. Smith, M.J. and P. Carayon, *Balance Theory of Job Design*, in *International Encyclopedia of Ergonomics and Human Factors*, W. Karwowsky, Editor. 2000, Taylor & Francis: New York.