

# Whither the Grid?

## Future Directions and Issues

Ralph H. Castain

International Space & Response Division  
MS-D448  
Los Alamos National Laboratory  
Los Alamos NM 87545 USA  
rhc@lanl.gov

The computing community has long pursued a vision of large-scale *metacomputers* that could provide a transparent, possibly global, computing resource. A variety of attempts to achieve this vision have been launched over the years [1–3], many stimulated by government funding agencies hoping for another Internet-like success. More recently, these approaches have coalesced into what is loosely termed *the grid* [4].

Despite the technological advances and the enthusiasm of the media, however, adoption of the grid computing paradigm has been slow and spotty due to four primary hurdles:

- Legal liability [5, 7]. Legal issues surrounding liability (for both users and providers) in grid computing environments are quite complex and still evolving. Recent case examples revolve around compromise of intellectual property (IP) while using a grid-enabled computing resource, and may extend to cover protection of backup media and virtual memory files that can potentially contain IP. Further complicating matters is that grid computing activities, by their very nature, cross jurisdictional boundaries and frequently involve international law. Consequently, lawyers within organizations are exercising considerable caution when confronted with requests to open up organizational computing resources to grid users.
- Security [6]. Providers of grid computing resources must trust third-party certificate authorities (CAs), but how reliable are they really? Who is liable when a certificate authority "fails" to do its job, perhaps allowing a malicious user to enter the system using a false or stolen identity? The current grid community's approach to CAs is based on "common reputation" – but that means security problems occur *before* a CA is publicly declared unreliable. Liability and redress of damages in such situations is far from clear.
- Competitive pressures. Access to – and more importantly, ownership of – computing resources has long been a competitive advantage. This is particularly true within the academic community where a team's ability to easily access resources such as high-performance computing clusters is frequently a required element within the proposal review process. Allowing members of a competing organization to access one's computing resources can be – and frequently is – viewed as working against one's own interests.
- Ownership. Although the current grid protocols provide for scheduling priorities, many organizations fear the potential overloading of their internal computing resources by external users, or may cause their own work to be delayed due to an external user's application. In addition, computing resources in many organizations are purchased and/or donated under specific rules of usage that may preclude usage by a broader community.

As a result, the grid paradigm has to-date primarily flourished in two distinct venues:

- Purpose-built, multi-organization systems. Typically operated by universities, these systems are built upon computers specifically purchased with government-provided funding for the purpose of being used on a grid. The fact that the funding agency dictated that the system be made available on a grid *may* afford the operator of the system some legal protection from liability as there can be no expectation of system security beyond that of the grid, though this has yet to be tested in court. Hosting the equipment, however, ensures a flow of operational funds and provides some degree of credibility that appears to translate into a competitive advantage in securing new R&D grants .
- Intra-organizational systems. These grids are almost always implemented as cycle-stealing applications during after-hours operations inside a corporate firewall or within a single academic department . The restriction to reside inside a single organization relieves the system from the legal liability, ownership, and competitive pressures, and provides the organization with some cost savings.

The problems impacting broader adoption of the grid paradigm are largely the result of treating grid computing as primarily a technological challenge – an understandable situation given the heavy scientific focus of the early adopters. However, grid computing is not just a technological problem – it entails legal liability, fair ownership use and other issues that transcend the technological problem of how to make computers work together. Resolving these issues will take some time. Meanwhile, several steps can be taken to encourage further adoption, including returning access control to local authority, operating at the user (and not the current root) level, and creating transparent environments whereby applications can be written to operate on a single machine, in a controlled cluster, or across the grid without any changes to source code.

This talk will present further detail on these issue, the recommended approach to resolving them, and possible interim steps towards that resolution.

## References

1. C. Catlett and L. Smarr, Metacomputing, *Communications of the ACM*, 35(6), pp. 44-52, 1992.
2. D.S. Stevenson and J.G. Rosenman, VISTANET gigabit testbed, *IEEE Journal on Selected Areas in Communications*, 10(9), pp. 1413-1420, Dec 1992.
3. A.S. Grimshaw and W.A. Wulf, The Legion Vision of a Worldwide Virtual Computer, *Communications of the ACM*, 40(1), pp. 39-45, 1997.
4. I. Foster, The grid: Computing without bounds, *Scientific American*, 288(4), pp. 78-85, Apr 2003.
5. J.C. Kesler, Contractual and regulatory compliance challenges in grid computing environments, *IEEE International Conference on Services Computing*, 1, pp. 61-68, July 2005
6. W. Knight, Locking up the grid, *Infosecurity Today*, 1(5), pp. 18-20, Sept 2005
7. R. Keck and D. Satola, Beware gridlock, *Business Law Today*, 13(4), Mar/Apr 2004

Authorized for release as LA-UR-07-0779.