**Cluster Computing – a Heuristic Approach**

James B. Joyce

University of Arkansas Little Rock

**Introduction**

The week spent at the Open Science Grid School at the University of Wisconsin at Madison was illuminating with respect to many areas, including operational mechanics of working with OSG systems, commonly used applications, and areas of research for which cluster computing is beneficial. The TeraGrid 2010 conference strongly reinforced the OSG School. With the opportunity to ask so many experts so many questions over the course of the two weeks, and coming from an applied computer science perspective, there are (at least) three distinct areas in which applied research from the computer science/integrated computing world could increase the productivity of the scientists and researchers who are using the systems. Two of these areas are infrastructure related: security and high availability. The third area is algorithmic in nature: applied heuristics.

The need for enhanced security and high availability should be intuitive; however, system downtime seems to be commonplace, and while some sites have implemented robust security architectures, others have unprotected clusters. The need for applied heuristics may not be quite as intuitive. Each of these three areas will be briefly discussed below.

**Security**

A treatise on the need for strong security of all cluster computing resources is beyond the scope of this paper. To keep it short and simple, of all computer systems, cluster resources are among the most coveted and targeted by the malicious computing community. The attackers range from script-kiddies using clusters as Warez sites, to industrial espionage professionals stealing or modifying research data, to cyber-terrorists using clusters for purposes like attack amplification, redirection, misinformation, and/or obfuscation. While a formalized and comprehensive security architecture *should* be adopted by the cluster community at large, at the very least, clusters should be protected by a firewall and intrusion detection, logs should be monitored, system services should be limited to the minimum set of services required to provide needed functionality, obvious port access should be obfuscated, and intrusion response should be planned out before the fact. Security awareness training should be mandatory for all system users, and administrators should have system specific security training.

**High Availability**

From a system perspective, HA has traditionally meant (at least) a duplication of single-point-of-failure components and enough horsepower and bandwidth to handle service level agreements, customer need, et al. This type of computing is de rigueur in corporate and government enterprise environments, and has developed out of an awareness, understanding, and quantification of the opportunity cost losses associated with system down-time. With an understanding that cluster duplication is a financial impossibility for most research organizations, implementing redundant front-end and login nodes should be much more feasible. With a redundant head node configuration, system maintenance can be performed without bringing down the entire cluster. This represents a bit of HA from the local perspective. Regarding HA from the grid perspective, scheduling of system maintenance can be coordinated across grid sites so as to ensure a maximum of available grid resources at any given time. Having heard several times at the conferences things like, “Our users are used to the downtime”, “There’s a lot of system maintenance”, “This is research computing… it’s not mission critical computing”, and “High availability and security have not been a priority”, it was clear that these were areas where applied computing research could quickly have a significant positive impact.

**Applied Heuristics**

There are numerous problems related to pattern recognition and extrapolation for which conventional programming techniques are less than optimal. These problems usually involve large numbers of variables which are difficult to correlate either programmatically or even intuitively. As well, with many of these problems, an “exact” answer is not always desired; oftentimes, a “fuzzy” answer is more desirable. An example of this type of problem was described at OSG School: looking for DNA sequences for 9mers and 12mers that are closely related to the original sequence, though not an exact match. Another example would be that of identifying all stars or galaxies which exhibit certain general characteristics. These types of problems are ideally suited for neural networks. While neural networks have been used commercially for the past 25 years or so, the industry commercial-based neural network environments have not been ported to run on multiple processors. I find this unusual, since neural networks themselves are examples of parallel processing, albeit on a micro-scale. As such, they have seen extremely limited use in cluster environments to date. As well, this size of problems to which neural networks have been applied has been limited by the amount of available memory and CPU resources on a single host computer running the neural network environment. Threading a commercial neural network application to function in a cluster environment should not prove to be too daunting a challenge, and providing a scaled up version would enable researchers to tackle problems for which there is no reasonable programmatic solution.

**Future Work**

As it pertains to cluster security, I am currently working on a security architecture template and will gladly share it with the group as things progress. Having spent most of my career (at least the last 25 years of it) integrating heuristics and cyber-security, I feel that an elegant and easy-to-implement security architecture can be proposed that will be beneficial for the community. With respect to high availability, I do not believe that any wheels are being recreated. There are numerous examples of HPC environments in which some level of high availability, fault-tolerance, and/or redundancy have been implemented. It comes down to simply making it a priority. Finally, regarding heuristics or artificial intelligence in the cluster environment, I have already met with an associate with whom I have worked closely for the past 20 years or so. He is the owner of the world’s leading neural network company, and he has agreed that he would like to see neural networks running on clusters. He has indicated that he will support a project to scale up his current neural networking development environment.