PROTECTING THE INFORMATION TECHNOLOGY INFRASTRUCTURE

Simulation and Modeling at the Exascale for Energy, Ecological Sustainability, and Global Security
Oak Ridge National Laboratory – Town Hall Meeting
May 18, 2007

Executive Summary

The objective of Breakout Group B4 is to develop tools and methods to protect the distribution information technology infrastructure: ensuring network security, preventing disruption of our communications infrastructure, and defending distributed systems against attacks.

Three primary themes emerged from the Breakout Group discussions: dedicated network channels, the "Sandbox" Model, and coordinated anomaly detection and response. Analysis of user command line sessions is projected to be similar at the exascale to what has already been achieved at the terascale and dedicated and pre-secured network channels help to mitigate the risk of the large exascale data flows. The "Sandbox" Model further makes the security problem scalable by isolating projects from each other and allowing the appropriate security to be applied inside of each "Sandbox" on the system rather than arbitrarily securing all projects on the computer to the highest level of the system. Coordinated anomaly detection and response enhance the ability to more accurately and expeditiously mitigate the risks.

To truly realize the full potential of computational science at the exascale level, there will need to be both progress in the technology, and new leadership strategies and policies to allow flexibility to capitalize not only on the advancements in the technology but the intellectual capital captured through open science. Achieving strong and visionary leadership, crisp policies, and cooperation and mutual understanding among the community will set the foundation for development of the technologies to ensure scalable security to the exascale level.

Introduction

The series of DOE-sponsored town hall meetings entitled, *Simulation and Modeling at the Exascale for Energy, Ecological Sustainability, and Global Security* (E3SGS), engage the scientific and technical community to leverage the knowledge needed to support the DOE strategic initiative to continue the drive in computational science over the next decade toward exascale computer systems.

The focus of the Breakout Group B4 is to identify not only today's environment, but also propose the future environment needed to support the development of tools and methods to protect the distributed technology infrastructure: ensure network security, prevent disruption of our communications infrastructure, and defend distributed systems

against attacks.

The second of three town hall meetings, the working group assembled at ORNL began to build upon the knowledge gained at the LBNL meeting regarding the current challenges, in order to present focused themes to understand how and what to secure to enable the group to draft cyber security requirements upon which to begin to characterize the necessary research focus.

The purpose of this white paper is to summarize the knowledge gained from the ORNL town hall meeting relative to the challenge set forth in meeting the goal of protecting the information technology infrastructure, such that that work may be carried forward at the third town hall meeting.
Within this goal, DOE defined three broad headings:

- Advanced computational and data storage/manipulation capability to support cybersecurity situational awareness,
- Modeling of human and application behavior on the Internet, and
- Providing service-oriented capabilities for multidisciplinary simulation.

ORNL Town Hall Meeting

The goals of the ORNL Breakout Group include the following: generate cyber security requirements in panel discussion format, define state-of-the-art and futures, identify gaps or bottlenecks to achieve, and identify research programs to address the gaps. Topics for the presentation included knowledge discovery through intelligent agents, current state-of-the-art, misuse tolerance, security at scale, and cyberinfrastructure for open science. From the discussions, three primary themes emerged: dedicated network channels, the "Sandbox" Model, and coordinated anomaly detection and response.

Dedicated Network Channels

The first theme identified is the need to build dedicated channels within networks, to allow the separation of data transfers and facilitate the application of graded levels of control. In the exascale environment, dataflow will grow significantly and the cyberinfrastructure is crucial. Cyber security analysis requirements do not necessarily scale up from the terascale environment to the exascale environment. Protection of interactive/control sessions at exascale is similar to what it was at terascale. Yet, the sheer increase in the bulk data transfer for an exascale environment presents significant challenges for cyber security analysis. Currently there are few protocols for bulk data transfers. While there are some that do use specific ports, these are not widely deployed.

Control sessions represent the primary threat. The segregation of the dataflows is one strategy to allow application of appropriate controls commensurate to security risks. System to system data channels may require less stringent security (i.e., less monitoring and analysis), allowing cyber security analysts to focus upon the control channels

where risks are greatest. Conversely, point-to-point networking may allow the needed bulk data transfers from trusted-system to trusted-system, regardless of the size of the transaction.

It is important to ensure consideration of the ease of information exchange as it is important to the open science community. Investing heavily in network monitoring also has the potential for a big payoff with regard to identification of potential attacks by facilitating cyber security analysis to better characterize behaviors to define anomalies for the exascale environment.

Potential Constraints:
- Impose upon the user community the requirement to properly configure clients to use appropriate ports
- Potential need to retool to accommodate service-oriented traffic
- Traffic may move to portals rather than computers which will require logging standards and better log capabilities

<u>"Sandbox" Model</u>

The second theme is the application of a "Sandbox" Model. The classification of DOE-SC HPC systems have changed from "Low" threat to "Moderate" threat as the HPC resource has become unique and desirable by industry, and this trend is expected to continue. The current DOE policy is to certify the system forcing researchers who desire to use the resources to conform to the highest level of risk for the systems, regardless of the risk level of their application. These types of tight controls present real barriers to the science community interested in participation in open science.

Sandboxes are defined, bounded virtual environments in which virtual organizations (VOs) can collaborate and share information. Each VO, regardless of the heterogeneity and physical locations of its members, would have its own sandbox and will be protected based upon its sensitivity and specific constraints (e.g., confidentiality, integrity, availability, etc.).

Application of a "Sandbox" Model allows independent certification of systems and sandboxes, such that a "Moderate" system can provide both "Moderate" and "Low" sandboxes. Within a "Low" security sandbox, the level of rigor of the controls would be relaxed as compared to higher security sandboxes (i.e., appropriate level of control for the level of risk). Basically, it is moving the perimeter from the outside (machine level) to the inside (sandbox level). The "Sandbox" Model makes the security problem scalable.

Constraints:
- Change current DOE Policy change from certification of systems (machine level) to sandbox level
- Applications, OS, memory, etc. will need to be segmented per sandbox
- Require appropriate operational controls associated with the allocation of responsibilities (e.g., common vocabulary, standard schemas for threat models, standardization of user agreements to facilitate understanding)

Coordinated Anomaly Detection and Response

The third theme identified is the need to shift toward coordinated anomaly detection and response and away from single signature-based security.

To do so, automated, intelligent tools are needed to detect anomalies in users, jobs/processes, application, and services behaviors across sites. The idea is for the tool not to replicate data (such as log data), but rather to use it in place via an intelligent index agent and an intelligent search agent. Intelligent agents can communicate in unstable environments and have successfully been applied in today's environment to perform analysis of mega-datasets, significantly reducing analysis time without sacrificing accuracy. These agents enhance the identification of anomalous behavior (e.g., "low and slow") with the application of forensic methodologies, and may help to automate responses (e.g., set frequency for large datasets for analysis vs. looking at small datasets as they come in).

Coupled with the technical capability to detect and respond to these anomalies is the need for better coordination and sharing of incident data and knowledge across sites. This information exchange itself presents potential risks to consider with regard to communication/message passing, information sharing, and encryption. The communication needs to have tight controls as defined by the best management practices and not necessarily driven by bureaucratic constraints.

The "Sandbox" Model previously discussed enhances anomaly detection by providing context for behavior characterization. By minimizing requirements for aliasing, encryption, etc., the amount of information bound to packets is more robust and provides information about context which helps to better identify anomalies.

Constraints:
- Methodologies are dependent upon the cyberinfrastructure
- Bureaucratic constraints present potential barrier to successful coordination

Summary

The continuing theme to leverage from the heterogeneity and autonomy of a community of users organized into VOs is particularly applicable to enhance the cyberinfrasture to accommodate both open science needs as well as higher security needs, and is one advantage of the "Sandbox" Model. Successful application of the model requires technical leadership and a new vision with regard to policy at various echelons from the top down, as well as the development of the appropriate infrastructure and support applications and tools.

Consideration also must be given to not only the technical controls, but also the operational controls associated with an allocation of responsibilities. In addition to appropriate policies, there must be a common vocabulary and appropriate operational controls to standardize discussions, identify standard schemas for threat

models scalable to the appropriate security level, and ensure standardization of user agreements to facilitate understanding.