

## **A Research Agenda for Heterogeneous Hardware and Cloud-based Software Techniques for Integrity in Scientific Computing**

May 4, 2015

Stephen P. Crago, USC / ISI and Department of Electrical Engineering, [crago@isi.edu](mailto:crago@isi.edu), 703-812-3729  
Submitted to: ASCR Cybersecurity for Scientific Computing Integrity Workshop, June 2-3, 2015,  
Area (1), Trustworthy Supercomputing

We are in the midst of two pervasive changes in computing today, both driven by a combination of technological and economic factors. First, processor hardware is becoming heterogeneous, which leads to power-efficiency gains but makes software more challenging. Second, software is moving to the cloud computing paradigm, partly to reduce the cost of obtaining and managing hardware, but, more importantly, to reduce software development costs while providing more scalability and robustness. The move toward heterogeneity has already been embraced by the high-performance computing (HPC) community, but the HPC community has been slower to adopt paradigms from cloud computing. The need to develop scalable, robust, and trusted applications on heterogeneous supercomputers while limited software costs will drive the adoption of cloud-based software technologies. In this whitepaper, we will discuss a research agenda for ensuring integrity for exascale systems through cloud-based software and heterogeneous hardware technologies.

The historic improvement trend for homogeneous, general-purpose microprocessors is coming to an end. One way to overcome these limitations, which the computing industry is already adopting, is by incorporating heterogeneity into processor architectures. Heterogeneity can increase efficiency through specialization and specialized units be turned off to save power when they are not used. Examples of such processing elements include graphical processing units (GPUs), field programmable gate arrays (FPGAs), vector or media functional units, highly parallel coprocessors (e.g. Xeon Phi), and encryption units. ***We believe widespread adoption of heterogeneous computing by almost all applications for which performance improvements are needed is inevitable, and that this adoption of heterogeneity has the potential to improve integrity.***

While this adoption of heterogeneity will lead to performance improvements of an order of magnitude or more for many applications, it will also inevitably lead to an increase in software and programming complexity. We must find ways to amortize the cost of this complexity and to hide it from end application programmers and users, so that those users can continue to focus on science rather than computing. Coincidentally, the emergence of cloud computing and the Platform-as-a-Service (PaaS) paradigm has led to the development of domain-specific programming environments that provide higher level programming interfaces with implementations that provide scalability and robustness while hiding the implementation and resource management from the user. ***We believe that elements of the PaaS model can be leveraged to improve the integrity of scientific applications running on DOE supercomputers.***

We propose a research agenda for integrity for scientific computing based on trusted supercomputing with the following programmatic thrusts:

- 1) **Exploiting heterogeneity to ensure integrity without sacrificing the efficiency at exascale:** The fact that processing elements will be parallel and heterogeneous leads to a variety of opportunities to improve integrity. Providing alternative compute task implementations on different types of processing elements reduces the static attack surface and reduces or eliminates the opportunity to reduce the integrity of an application through a single exploit based on a vulnerability in a single implementation. Accelerator resources can be used to check the computations or behavior of other computing resources while minimizing the overhead of these checks, and the spatial separation of heterogeneous resources provides another layer of integrity. We advocate an ASCR research agenda that includes programs that develop techniques for exploiting heterogeneity to ensure trust.
- 2) **Develop hardware and software co-design and run-time techniques that enable DOE to leverage commercial technology for open science without sacrificing trust:** There have been and are research programs by other agencies that are developing techniques for ensuring that integrated circuits do not contain malicious circuits. Similarly, there are techniques to ensure that boot-up codes for general-purpose processors and field-programmable gate array configurations have integrity. While these hardware-based techniques may uncover some kinds of vulnerabilities, there are other vulnerabilities that may be better detected by software, and we advocate that ASCR include programs that develop both hardware and software trust techniques in its research portfolio.
- 3) **Develop domain-specific programming models and compilation and run-time technologies inspired by the PaaS paradigm that can achieve scalability and efficiency while ensuring integrity:** Data analytic (e.g. Hadoop MapReduce and Spark) and machine learning (Mahout) frameworks have emerged to leverage dynamically provisioned resources in the cloud. The advantage of these frameworks is that they provide a stable, domain-specific programming model to the programmer, who can focus on getting their application written, while targeting the parallelism available in cloud computing through modularity. The domain specific nature of the frameworks allows the run-time system to be specialized to exploit parallelism and to provide other properties (e.g. fault tolerance) efficiently. These frameworks already provide fault tolerance because they are designed for cloud-based environments, where nodes can fail and network connections can be lost, and the techniques could be extended to ensure integrity. Implementations can also use virtualization and a resource management layer to provide isolation between tasks and applications. We believe these techniques can be extended and analogous techniques can be invented that provide a more robust form of integrity for applications, and that ASCR should include such programs in its research portfolio.

The convergence of heterogeneous hardware and cloud-based software technologies offers a unique opportunity to change supercomputer software that can ensure the integrity of scientific applications. In this workshop, we have the opportunity to define the research agenda for the next 10-20 years that will allow ASCR to realize this vision.