FERMILAB-POSTER-22-183-OCIO

Computing Node Scanning Tool for Collaborative Science

Lucas D'Antonio, *Purdue University Northwest*, Under the Mentorship of Jeny Teheran
Office of the CIO, Computer Security, Fermi National Accelerator Laboratory, Batavia, Illinois 60510

OSPool and High Throughput Computing

The Open Science Pool (maintained by the Open Science Grid - OSG) aggregates mostly opportunistic computing resources from highly decentralized contributing clusters at campuses and other organizations, making them available to the US-based open science community. These resources are utilized by virtually all scientific disciplines in some form across the United States. However, auditing these large, distributed clusters for vulnerabilities provides a challenge for OSG security team members, as each site is independently operated.

Open Science Grid sites in the continental United States. 20+ countries in Central and South America also contribute resources.

The OSG security team has little visibility into the packages (and vulnerabilities) deployed at different compute nodes across sites. Without a proper inventory of these systems, mitigation efforts by the security team are severely restricted. This necessitates the creation of a new tool for use by the security team that will be able to analyze the many worker nodes connected to the resource pool.

Computing Node Scanning Tool Architecture

This scanning tool allows the OSG security team to gather information about the computing resources in the Open Science Pool including the operating system, kernel version, any installed packages, and any unapplied security patches to those packages. To accommodate the diverse computing environment, the information collecting process was designed as a set of HTCondor job payloads. These payloads are extensible, can be customized and can be targeted to run on specific resources at will. No installation of additional software, system configuration changes, or elevated privileges are required.

The results are securely transmitted back in the form of a series of files containing the desired targeted asset information. This information can then be organized on a web server for further review by designated security team members. While efforts are made to secure this data, it is important to note that any information collected is available to anyone with access to the Open Science Pool. Therefore this tool is designed to be thorough when gathering information, yet discrete and non-intrusive in implementation.

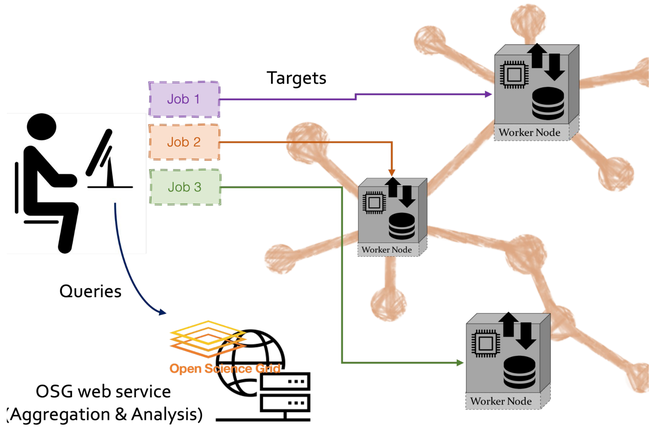

Diagram – how the Compute Node Scanning Tool should operate on the Open Science Pool. A member of the OSG security team queues a series of jobs on HTCondor across different worker nodes and returns information about each machine that runs the script.

Further Considerations

There are a multitude of variables to keep in mind when designing this tool. The working environment is highly heterogenous. Some worker machines on these clusters are containerized to accommodate older programs for instance. Accommodation for these unpatched machines would be permissible since they would be isolated from the machine. A series of tests were designed to adjust the data collection commands to better accommodate the different platforms running on these computing nodes.


```
{
  "Date": "Mon Aug  1 20:10:28 UTC 2022",
  "reportingSite": "login05.osgconnect.net",
  "isContainer": "false",
  "operatingSystem": "Scientific Linux 7.9 (Nitrogen)",
  "Kernel": "3.10.0-1160.59.1.el7.x86_64",
  "Packages": [
    "cyrus-sasl-lib-2.1.26-24.el7_9.x86_64",
    "dwz-0.11-3.el7.x86_64",
    "yum-conf-sl7x-7.9-1.sl7.noarch",
    "globus-xio-gsi-driver-5.4-1.el7.x86_64",
    "bzip2-1.0.6-13.el7.x86_64"
```


Sample data returned from a worker node using the computing node scanning tool, organized as a json object for easy access.

Conclusions

By having this node scanning tool at its disposal, the Open Science Grid has a viable solution for identifying and isolating unpatched systems in the Open Science Pool. This will make the open science community safer from cyber threats overall, while not violating each contributing site's autonomy.

Since the scanning tool is designed to operate as a series of jobs on HTCondor, it can easily be modified to run on other high throughput collection services such as SLURM, and in other high throughput resource pools or collaborations, such as the US-CMS.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

This research was supported in part by an appointment to the U.S. Department of Energy's Omni Technology Alliance Internship Program, sponsored by DOE and administered by the Oak Ridge Institute for Science and Education.