Securing an Open and Trustworthy Ecosystem for Research Infrastructure and Applications (SOTERIA)

Brian Bockelman, Morgridge, on behalf of

Brian Aydemir (Morgridge)
Rob Gardner (U. Chicago)
Fengping Hu (U. Chicago)
Farnaz Golnaraghi (U. Chicago)

NSF Grant #2115148 CICI:UCSS:Securing an Open and Trustworthy Ecosystem for Research Infrastructure and Applications (SOTERIA)

Supply-Chain Security - Near and Present Danger

Tracing the Supply Chain Attack on Android

June 25, 2019 — 46 Comments

Earlier this month, **Google** disclosed that a supply chain attack by one of its vendors resulted in malicious software being pre-installed on millions of new budget Android devices. Google didn't exactly name those responsible, but said it believes the offending vendor uses the nicknames "**Yehuo**" or "**Blazefire**." What follows is a deep dive into the identity of that Chinese vendor, which appears to have a long and storied history of pushing the envelope on mobile malware.

Industry is waking up to the danger of supply-chain attacks.

Big Question: How do we bring industry advancements into the Open Science ecosystem?

Scientific Software Environments

What makes the scientific software different? Is scientific software different?

Unique:

- **Disadvantage**: Poor/non-existent software engineering training provided.
- **Disadvantage**: No reward system for software quality.
  - The software doesn't really need to work or be installable to get credit (publication).
- **Advantage**: Existing trust and reputation network via publication record.
- **Neutral**: Rare use of cloud, distinct set of research computing centers. OCI ('Docker') containers are not the dominant format.

Same:

- Large tree of dependencies.
- Common building blocks (Python, conda), esp. as data science and ML is adopted by industry.
- (Increasing) adoption of containers and related "cloud native" technology.

Containers to the Rescue! (Well, Not Really)

- Containers make it easier to run complex software
  - Just find the thing you want on Docker Hub (and hope it was built by a reputable source, up to date, etc)
- Containers keep the scientific software stack orthogonal to the system software stack
  - Your code lovingly built with GCC 4.1.2 and Python 2.4 on Scientific Linux 5 will work forever … right?
- Container images are immutable, **including the bugs**
  - Heartbleed, Shellshock, your favorite CVE with a T-shirt and website.

Want to share the software to reproduce your publication?
Put it up on DockerHub*

* Oops - DockerHub now deletes unused images after 6 months

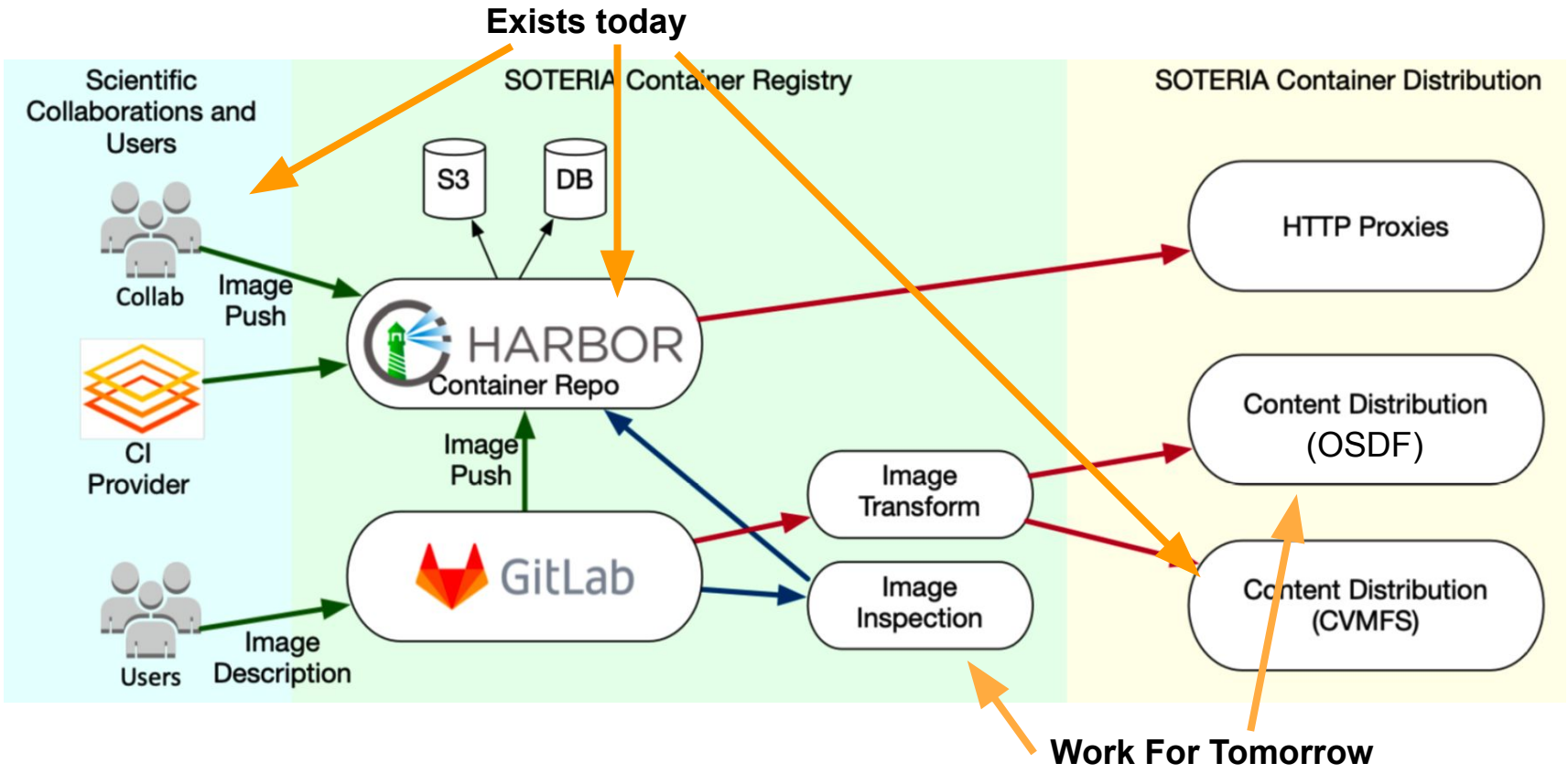
SOTERIA - Filling the Gaps

The SOTERIA project is innovating ways to move the field forward:

- Run a container registry for any US researcher and their collaborators
  - A core part of ‘translational CS’ is having a happy user base to work with.
- Provide tools for discoverability. **Make it easy to cite and discover images**.
  - Generation of Metadata for creation of DOIs and archival
  - Funding Agencies
  - Grant Numbers
- Provide container visibility. **We must understand what’s in the container.**
  - Advertise the discovered software including system RPMs, known vulnerabilities, conda contents.
- Provide environment traceability. **Who uploaded the container?**
  - Associate each artifact with an ORCID.
- Advocate for best practice (immutable tags, CI/CD/autobuilds)

https://soteria.osg-htc.org/

Architecture



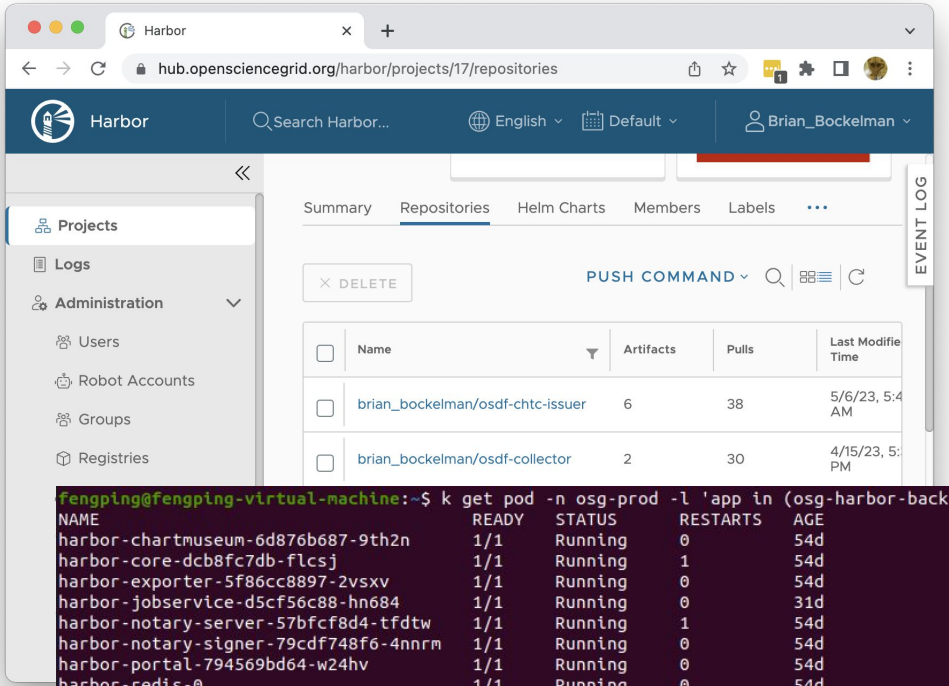
OSG Hub

Powered By

HARBOR

- Implemented by Harbor, an open source container registry.
- Authentication provided by CILogon
  - Use federated identity, not tied to the project!
- Operated on the PATh Kubernetes [platform and tools](#).
- Designed to be high-ish available:
  - Database has an on-site active-standby setup. Incremental snapshots are sent to alternate site and backed up offsite.
  - S3 bucket is similarly replicated.
  - Set to run at the Wisconsin and Chicago sites.

Goal: Failover from primary to secondary site in 30 minutes.

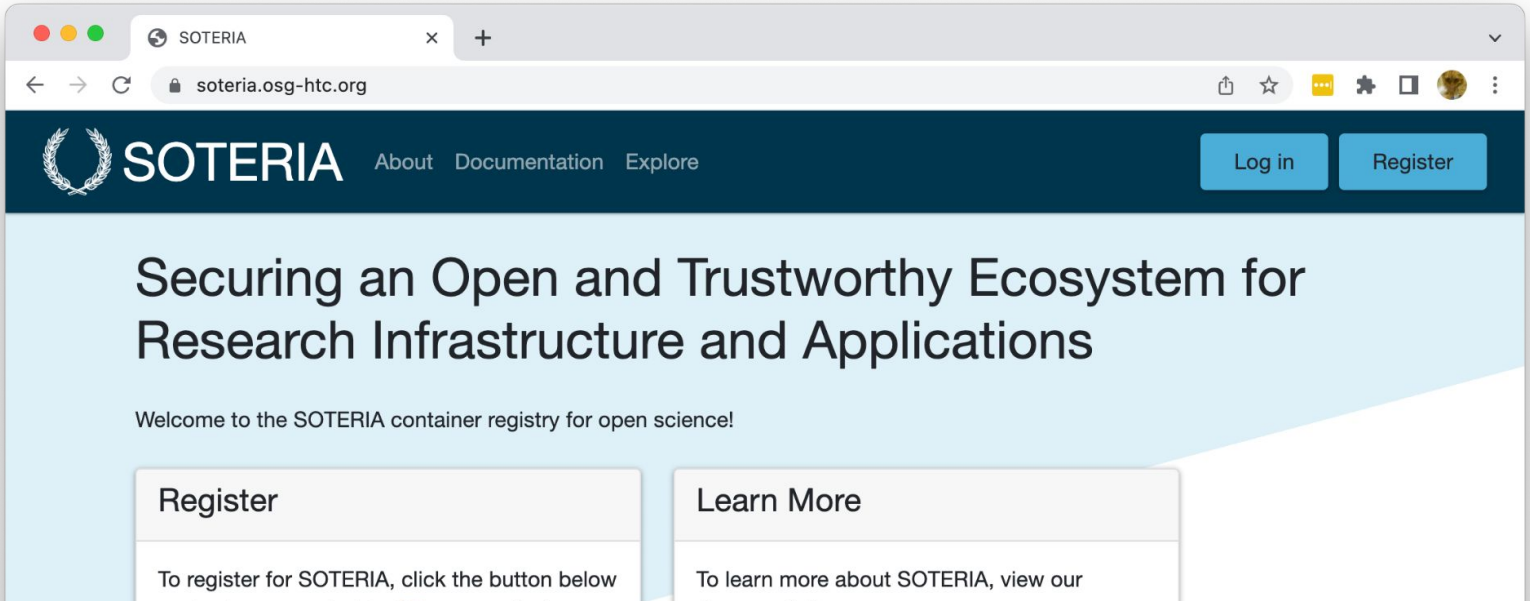



```
fengping@fengping-virtual-machine:~$ k get pod -n osg-prod -l 'app in (osg-harbor-backu
NAME                                  READY   STATUS      RESTARTS   AGE
harbor-chartmuseum-6d876b687-9th2n    1/1     Running     0          54d
harbor-core-dcb8fc7db-flcsj           1/1     Running     1          54d
harbor-exporter-5f86cc8897-2vsxv      1/1     Running     0          54d
harbor-jobservice-d5cf56c88-hn684     1/1     Running     0          31d
harbor-notary-server-57bfcf8d4-tfdtw  1/1     Running     1          54d
harbor-notary-signer-79cdf748f6-4nnrm 1/1     Running     0          54d
harbor-portal-794569bd64-w24hv        1/1     Running     0          54d
harbor-redis-0                        1/1     Running     0          54d
harbor-registry-b456b8b6c-cr82l       2/2     Running     0          54d
harbor-trivy-0                        1/1     Running     0          54d
osg-harbor-backup-28052835-pqms2      0/1     Completed   0          2d16h
osg-harbor-backup-28054275-hn4cr      0/1     Completed   0          40h
osg-harbor-backup-28055715-cd2gm      0/1     Completed   0          16h
fengping@fengping-virtual-machine:~$ k get pod -n osg-prod -l postgres-operator.crunchy
NAME                                  READY   STATUS      RESTARTS   AGE
harbor-prod-1-backup-v722-549gr       0/1     Completed   0          114d
harbor-prod-1-base1-bd9k-0            2/2     Running     0          56d
harbor-prod-1-repo1-diff-27974700-55v8m 1/1   Running     0          56d
harbor-prod-1-repo1-diff-28056240-mfcfl 0/1   Completed   0          8h
harbor-prod-1-repo1-diff-28056420-p98vg 0/1   Completed   0          5h8m
harbor-prod-1-repo1-diff-28056600-qp4d6 0/1   Completed   0          128m
harbor-prod-1-repo1-full-28052700-9xkwf 0/1   Completed   0          2d19h
harbor-prod-1-repo1-full-28054140-pcd5w 0/1   Completed   0          43h
harbor-prod-1-repo1-full-28055580-vssvv 0/1   Completed   0          19h
fengping@fengping-virtual-machine:~$
```

Bootstrapping a Scientific Community

But how does one get access to the OSG Hub in the first place? Our goals:

1. Start using and testing without any staff interaction.
2. Limit “scale use” to US researchers and their collaborators.



Project Onboarding - SOTERIA & COmanage

- COmanage integrates with CILogon to provide a user registry for SOTERIA
  - Roles identify the degree of vetting a user has gone through
  - Group membership controls access to projects on OSG Hub
- The initial registration flow allows us
  - To link a OSG Hub user with their ORCID iD
  - Provision that user a private project so that they can explore the system
- Users can apply for Researcher status
  - Grants the ability to create five projects (three public, two private)
  - Grants the ability to add other SOTERIA users as members of those projects

Key concept:
Each image is tied to the researcher's "social identity" (ORCID).

Registration

Follow the steps below to register for Affiliate status in SOTERIA. This grants you access to a private project on OSG Hub with a 5 GiB quota.

1. Join the SOTERIA Group

SOTERIA is managed as group under the OSG organization on CILogon.

Fill out this form to join the group of SOTERIA users. When prompted by CILogon, select the same identity provider that you used initially to log into SOTERIA.

Check for Group Membership

2. Link Your ORCID iD

ORCID iDs help uniquely identify you amongst SOTERIA's community of researchers.

First, register for an ORCID iD if you do not already have one. Then, link it to SOTERIA.

Check for ORCID iD

3. Create a OSG Hub Account

SOTERIA uses OSG Hub for its container registry.

Create an account there by logging in via OIDC and, when prompted by CILogon, selecting the same identity provider that you chose when logging into SOTERIA. Then come back to this page and click the button below to confirm that SOTERIA can find your OSG Hub account.

Check for OSG Hub Account

4. Create Your Project

Researcher Registration

Researcher status is available for all SOTERIA members that meet certain criteria. You can find these requirements listed out in the documentation.

Institute Affiliated Email

Requirement Met

Active Federal Grant on ORCID Profile

Researcher is listed as having an active grant through that institution from a US federal government funding agency AND the grant shows up in their public ORC ID profile. Staff member should note the grant #, and funding agency.

Grant Number

Funding Agency

Click submit and we will follow up by creating a ticket and sending you an email at the provided address.

Close Submit

Account

Details Actions Projects

Details

Name	ORCiD
BRIAN EMRE AYDEMIR	0000-0001-9048-5408
Status	**Email**
Researcher	BAYDEMIR@MORGRIDGE.ORG

Actions

Create New Project

Researchers can create 3 Public and 2 Private projects. If you require more than the initial allocation please email support@osg-htc.org to request an additional allocation.

Create a new Project

Add Users To My Projects

Soteria leverages Comanage to allow researchers to control who has what permission to their projects. To add users

View Documentation Go To Comanage

Create Project

Researchers can create 3 Public and 2 Private projects.

Name

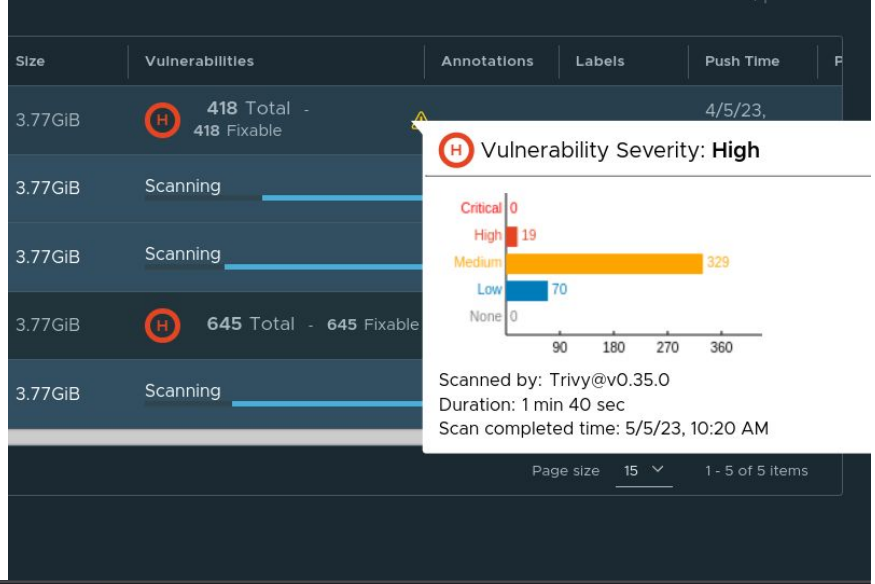
Visibility

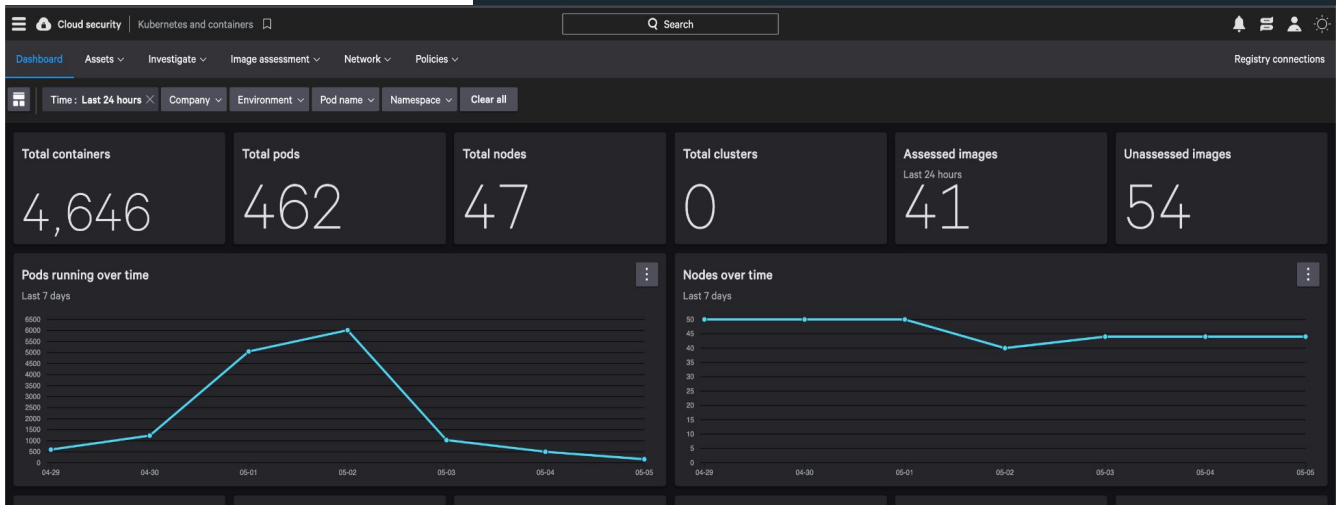
- Public
- Private

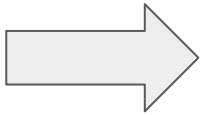
Submit

Vulnerability Scanning

- Vulnerability scanning is an obvious starting point for **image visibility goals**.
- Harbor comes with Trivy to scan images.
- Investigating Crowdstrike Falcon also offers Cloud workload protection and container security.
  - It offers asset management and image assessment for Kubernetes clusters





Vulnerability Scanning  Image Visibility

- We started with the Crowdstrike Falcon image vulnerability analysis as it's a readily-available tool that works **with Harbor but isn't a part of Harbor**.
- Long-term vision:
  - Each artifact uploaded gets registered as a document in an ElasticSearch DB.
  - This triggers a suite of analysis tools that run against the artifact.
  - Some security-related, some visibility. **Can we enumerate and advertise all the software installed in a Conda environment?**
  - Results of analysis tools are sent to the DB, advertised as part of the public webpage of the artifact.

What Comes Next? Image Distribution

‘docker pull’ isn’t the end of the story. For each public image,

- Convert to singularity and upload to the OSDF for distribution.
  - Could even be done for non-public images.
- Convert to flat directory and publish into CVMFS.
  - Planning the successor to the singularity.opensciencegrid.org repo!

Beyond that, there’s a need for:

- **Auditing**: given an ‘interesting’ image, what was used where?
- **Selective replication**: Providing endpoints that *only* mirror images which are signed or without known critical-CVEs.

What comes next? Long-term

SOTERIA is a research project with a beginning, middle – **and end**.

- OSG Hub, as a part of the OSG Consortium, has a much longer lifetime.

There's no universally-accepted way to capture, archiving, and assign a persistent identifier to a **software environment**.

- Minimally, we'd like to build tools to provide a smooth path to archive these containers to Zenodo.
  - Much, much to do in metadata - we're not experts!

Bodleian Library, in Oxford, was founded in 1602. It is believed that 400 years is longer than the typical NSF project.

Looking for partners and friends!

Thanks!

This material is based upon work supported by the National Science Foundation under Grant No. 2115148. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.