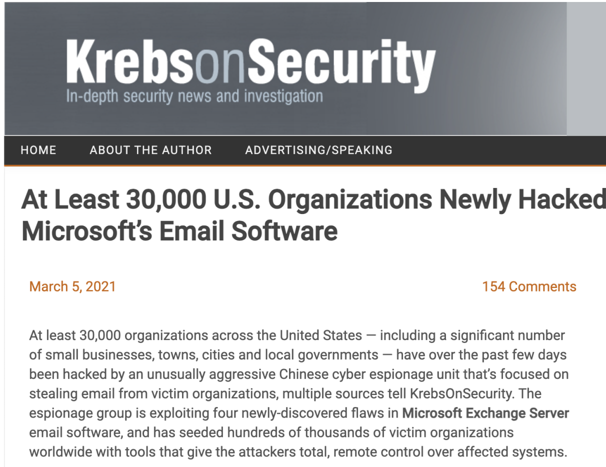
SOTERIA: Empowering Scientific Research with Secure Software Environments.

Farnaz Golnaraghi, University of Chicago, on behalf of

Brian Aydemir (Morgridge)
Cannon Lock (Morgridge)
Brian Bockelman (Morgridge)
Fengping Hu (UChicago)
Lincoln Bryant (UChicago)
Rob Gardner (UChicago)

Throughput Computing 23
Madison, WI
07/11/2023

The Rising Threat of Supply Chain Attacks

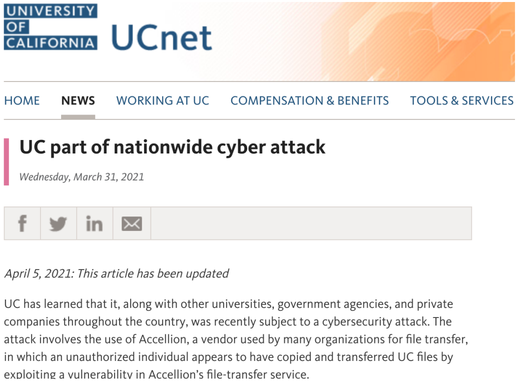


KrebsonSecurity
In-depth security news and investigation

At Least 30,000 U.S. Organizations Newly Hacked Via Holes in Microsoft's Email Software

March 5, 2021 — 154 Comments

At least 30,000 organizations across the United States — including a significant number of small businesses, towns, cities and local governments — have over the past few days been hacked by an unusually aggressive Chinese cyber espionage unit that's focused on stealing email from victim organizations, multiple sources tell KrebsOnSecurity. The espionage group is exploiting four newly-discovered flaws in **Microsoft Exchange Server** email software, and has seeded hundreds of thousands of victim organizations worldwide with tools that give the attackers total, remote control over affected systems.



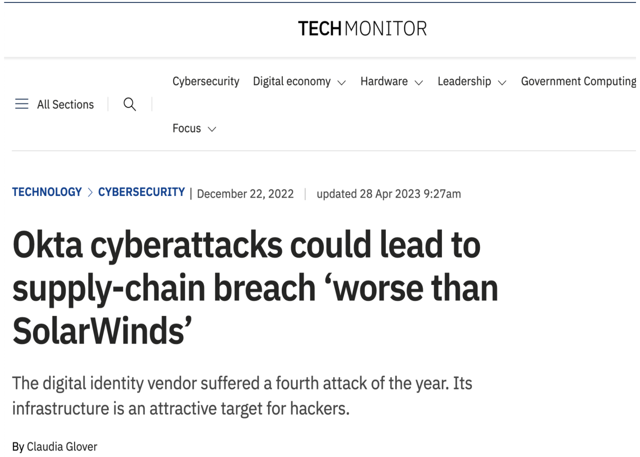
UNIVERSITY OF CALIFORNIA UCnet

UC part of nationwide cyber attack

Wednesday, March 31, 2021

April 5, 2021: This article has been updated

UC has learned that it, along with other universities, government agencies, and private companies throughout the country, was recently subject to a cybersecurity attack. The attack involves the use of Accellion, a vendor used by many organizations for file transfer, in which an unauthorized individual appears to have copied and transferred UC files by exploiting a vulnerability in Accellion's file-transfer service.



TECH MONITOR

TECHNOLOGY > CYBERSECURITY | December 22, 2022 | updated 28 Apr 2023 9:27am

Okta cyberattacks could lead to supply-chain breach 'worse than SolarWinds'

The digital identity vendor suffered a fourth attack of the year. Its infrastructure is an attractive target for hackers.

By Claudia Glover

- Supply chain attacks have been making headlines with increasing frequency.
- Traditionally, they have been associated with sectors such as finance and industry.
- Our scientific community is not immune to their repercussions.
- The repercussions can be severe, leading to data breaches, intellectual property theft, and even wide distribution of compromised research results.

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

Characteristics of Scientific Software

- Scientific software exhibits some unique characteristics that differentiate it from other domains.
  - **Disadvantage:** Insufficient software engineering training and lack of emphasis on software quality.
  - **Disadvantage:** Minimal incentive for robust software functionality due to publication-centric credit systems.
  - **Advantage:** Benefit from an established trust and reputation network through publication records.
  - **Neutral:** Limited use of cloud infrastructure and OCI (‘Docker’) containers not yet dominant.

- Despite these differences, scientific software has some similarities with other domains.
  - Large dependency trees and reliance on common building blocks like Python and Conda.
  - Growing adoption of containers and related "cloud-native" technologies.

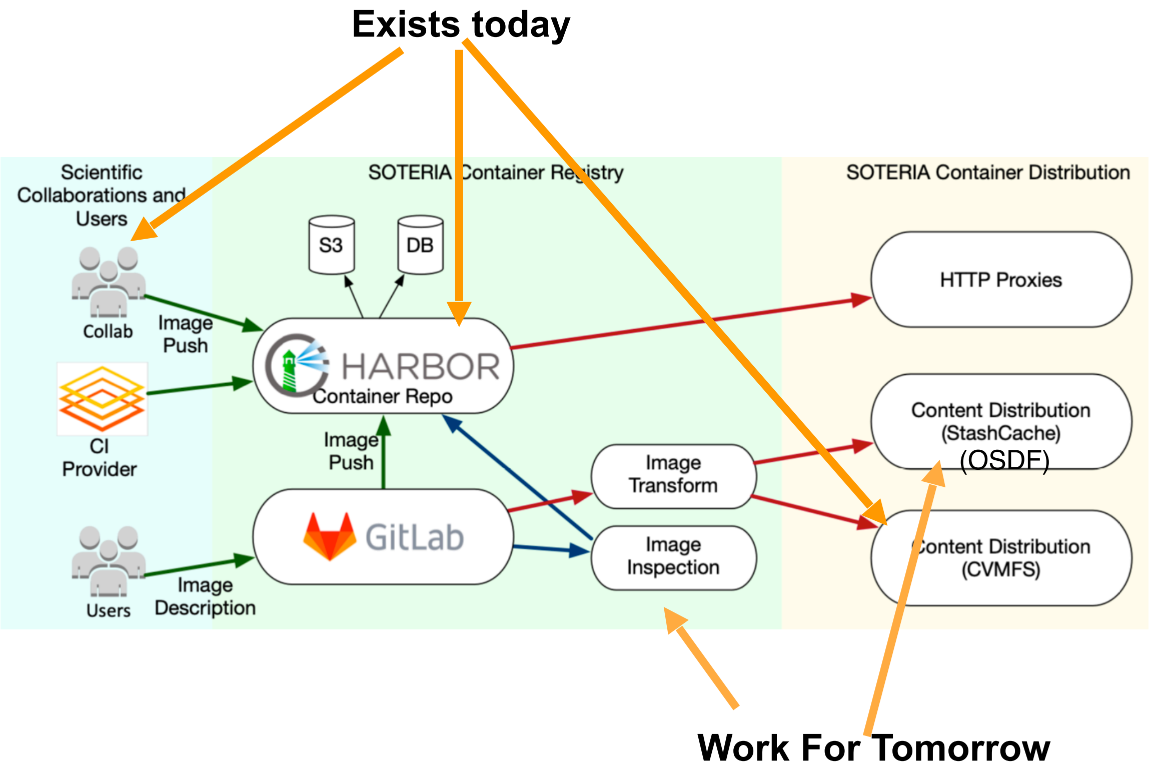
Adoption of Containers:

- Containers streamline the execution of complex software.
  - Simplified access through Docker Hub
    - Caution is necessary regarding reputable sources and up-to-date images.
  - Containers isolate the scientific software stack from the system software stack, enhancing portability.
- Container images offer immutability, but vulnerabilities can persist.
  - Historical software bugs remain within immutable container images.
  - Awareness and mitigation of security vulnerabilities are crucial to ensure robust containerized workflows

How does SOTERIA contributes to filling the gaps and why is it necessary?

- Registry for researchers and collaborations:
  - A service exclusively focused on meeting the needs of secure scientific software environment management.
- Traceability:
  - Where did the container come from? Was it signed by the user?
- Reproducibility:
  - Making the software more reusable and accessible
- Container visibility:
  - Provide information on the systems RPM, vulnerabilities, etc.
- Discoverability:
  - Make it easy to cite and discover images.
- Training and education on the importance of secure software environments

Architecture



Exists today

Scientific Collaborations and Users
SOTERIA Container Registry
SOTERIA Container Distribution
S3
DB
Collab
Image Push
HARBOR
Container Repo
CI Provider
Image Push
GitLab
Users
Image Description
Image Transform
Image Inspection
HTTP Proxies
Content Distribution (StashCache) (OSDF)
Content Distribution (CVMFS)

Work For Tomorrow

Image creation 🟩

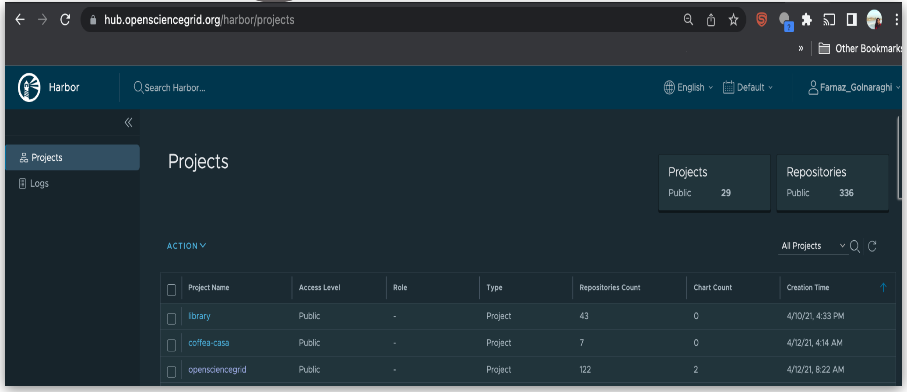
Image introspection 🟦

Image distribution 🟥

SOTERIA Container Registry

- We adapt and utilize open source tools to meet the scientific community’s needs
  - **OSG Hub** powered by **Harbor**
    - Authentication via **CILogon**.
      - Use federated identity, not tied to the project.
    - Operated on the ***PATh Kubernetes platform***
    - 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.

HARBOR



Projects

Projects Public 29

Repositories Public 336

□	Project Name	Access Level	Role	Type	Repositories Count	Chart Count	Creation Time
□	library	Public	-	Project	43	0	4/10/21, 4:33 PM
□	coffea-casa	Public	-	Project	7	0	4/12/21, 4:14 AM
□	opensciencegrid	Public	-	Project	122	2	4/12/21, 8:22 AM

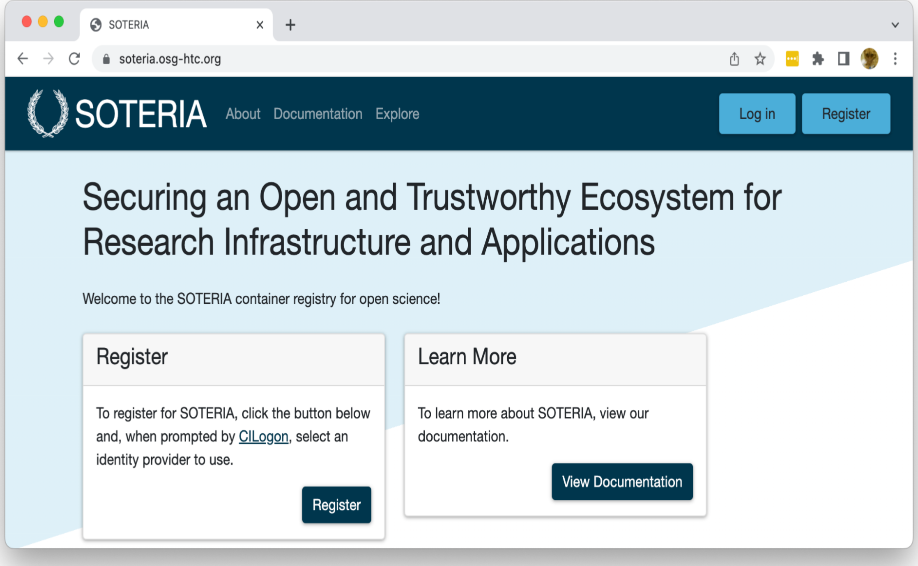
✳ **OSG Hub statistics**:

- Project Count: 52
- Repository Count: 430
- Unique images: 28,261
- Image pulls: ~78.4 million
- Storage allocation: 3.2 TB

Identity and user management, project onboarding

- **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

soteria.osg-htc.org



Securing an Open and Trustworthy Ecosystem for Research Infrastructure and Applications

Welcome to the SOTERIA container registry for open science!

Register

To register for SOTERIA, click the button below and, when prompted by CILogon, select an identity provider to use.

Register

Learn More

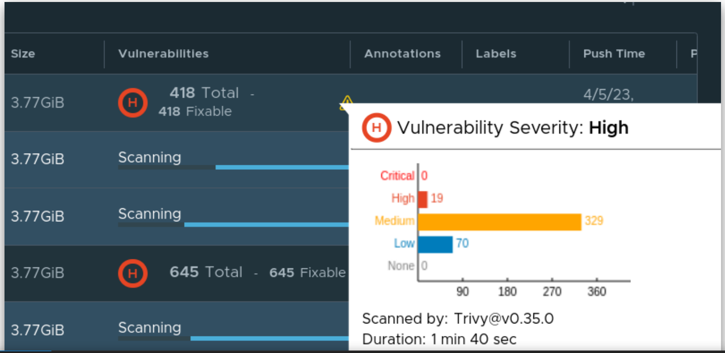
To learn more about SOTERIA, view our documentation.

View Documentation

Key concept:
Each image is tied to the researcher’s “social identity” (ORCID).

Vulnerability Scanning → Image Visibility

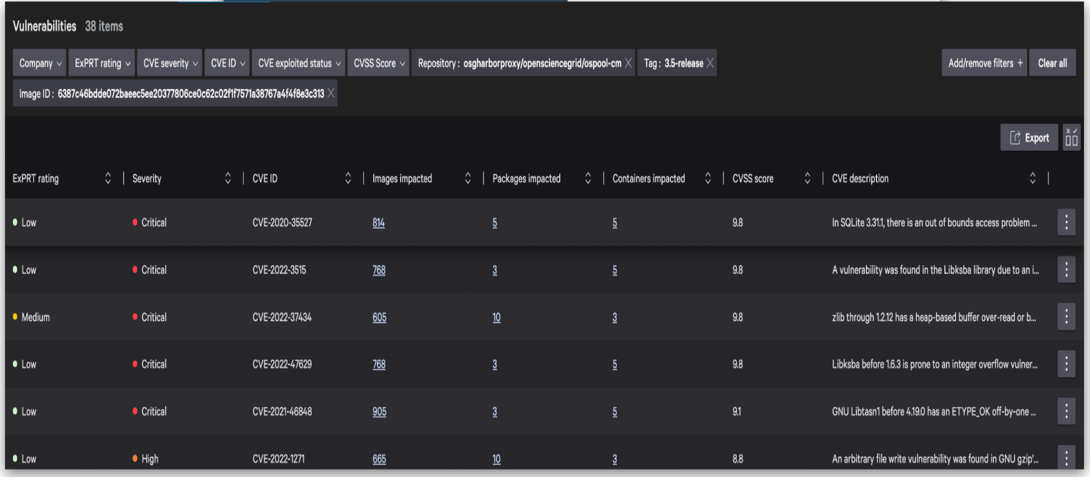
- Vulnerability scanning is an obvious starting point for **image visibility goals**.
- Harbor comes with Trivy.
- We started with the Crowdstrike Falcon - readily-available tool that works with **Harbor but isn't a part of Harbor**.
- Long-term vision:
  - Each artifact uploaded gets registered as a document in an ElasticSearch DB.
  - This triggers a suite of analysis tools that run against the artifact. Some security related and 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.



Size	Vulnerabilities	Annotations	Labels	Push Time
3.77GiB	H 418 Total - 418 Fixable			4/5/23,
3.77GiB	Scanning			
3.77GiB	Scanning			
3.77GiB	H 645 Total - 645 Fixable			
3.77GiB	Scanning			

H Vulnerability Severity: **High**

Scanned by: Trivy@v0.35.0
Duration: 1 min 40 sec



Vulnerabilities 38 items

Repository : osgharborproxy/opensciencegrid/ospool-cm

Tag : 3.5-release

Image ID : 6387c46bdde072baeec5ee20377806ce0c62c02f1f7571a38767a4f4f8e3c313

ExPRT rating	Severity	CVE ID	Images impacted	Packages impacted	Containers impacted	CVSS score	CVE description
Low	Critical	CVE-2020-35527	814	5	5	9.8	In SQLite 3.31.1, there is an out of bounds access problem ...
Low	Critical	CVE-2022-3515	768	3	5	9.8	A vulnerability was found in the Libksba library due to an i...
Medium	Critical	CVE-2022-37434	605	10	3	9.8	zlib through 1.2.12 has a heap-based buffer over-read or b...
Low	Critical	CVE-2022-47629	768	3	5	9.8	Libksba before 1.6.3 is prone to an integer overflow vulner...
Low	Critical	CVE-2021-46848	905	3	5	9.1	GNU Libtasn1 before 4.19.0 has an ETYPE_OK off-by-one ...
Low	High	CVE-2022-1271	665	10	3	8.8	An arbitrary file write vulnerability was found in GNU gzip'...

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.

Thank you!!
Questions?

soteria.osg-htc.org

National
Science
Foundation

This work was supported in part by NSF Grant #2115148 CICI:UCSS:Securing an Open and Trustworthy Ecosystem for Research Infrastructure and Applications (SOTERIA)