Graph for Understanding Artifact Composition

Brandon Lum (@lumjjb), Google
Jeff Mendoza, Kusari
Michael Lieberman (@mlieberman85), Kusari
Mihai Maruseac(@mihaimaruseac), Google
Parth Patel (@pxp928), Kusari
Santiago Torres (@torresariass), Purdue University
The Problem!

- Analyzing our software supply chains are COMPLICATED
  - You are not just impacted by the software you ingest but the software your software ingests
Why?

● Too much data
  ○ SLSA, SBOMs, OSV, VEX, other ITE6
  ○ Vuln Scans, Certification checks
  ○ <enter your document format here>

● Also too little data
  ○ Data quality is often poor
  ○ We don’t know what we don’t know!
  ○ If you don’t know about a threat doesn’t mean it’s not there.

● We need to understand the relationships between the data
  ○ Connections within the supply chain of various ecosystems
  ○ Mapping it to threat intelligence and vulnerability information such as OSV and VEX
  ○ Reconciling identifiers of different documents
  ○ Many more…
Why?

- Too much data
  - SLSA, SBOMs, OSV, VEX, other IT
  - Vuln Scans, Certification checks
- Also too little data
  - Data quality is often poor
  - We don’t know what we don’t know!
- If you don’t know about a threat doesn’t mean it’s not there.
- We need to understand the relationships between the data
  - Connections within the supply chain of various ecosystems
  - Mapping it to threat intelligence and vulnerability information such as OSV and VEX
  - Reconciling identifiers of different documents
- Many more…
What is GUAC

A knowledge graph of software metadata to answer security and supply chain questions
Policy is here!

Policy and insight
Automation, risk management, and compliance throughout the SDLC. Governance, developer assistance, and policy shifted left.

GUAC is here!

Aggregation and synthesis
Smart aggregation turning data into meaning. Intelligent linking of project, resource, developer, artifact, repo, toolchain.

SLSA is here!

Software attestations
Schemas and sources for rich security metadata. SBOM, SLSA provenance, VEX, OSV, security scorecards, developer reputation, plus proprietary data.

Sigstore is here!

Trust foundation
What Supply Chain / Security Questions?

**Proactive**

How do I prevent large scale supply chain compromises?

**Preventive**

Have I taken the right safeguards?

When deciding to use and deploy software, are there sufficient security checks and approvals?

**Reactive**

HOW AM I AFFECTED???

A vulnerability or supply chain compromise is discovered!

- SLSA
- Codecov
- Solarwinds
- trivy
- grype

https://xkcd.com/2347/
Ingest/Process

Public Record
- GitHub
- OSV
- Amazon S3

Org. (Private)
- Azure Blob Storage
- Google Cloud Storage

Vendors
- trivy
- grype

Extract information and relationships
Graph contains:
- Artifacts/identifiers
- Attestations
- Identities
- Relationships (edges)

CISO Inventory Policy
CMDB
Query
2 types of queries

**Informational queries**
- I want to query properties of my software - i.e. which software is affected by vuln X?
- Returns a set of artifact nodes with this information
- “Best effort” (temporal) - given current state of information

**Evidence queries**
- I want to make a policy decision, provide me evidence that the policy is fulfilled
- Returns a subgraph containing all nodes and edges together with the documents/references associated with each node (Attestations, Documents, etc.)
- Assuming monotonic predicates, subgraph acts as a proof
Two major use cases/deployment models

Public Service (e.g. Public-good service run to collect data for public use)
- Ingests all* verifiable documents from public records only
- Queryable by public

Organization (e.g. Private Internal knowledge graph instance consisting of software used within the organization)
- Graph ingests internal documents as well as relevant public documents (e.g. open source libraries)
- Queryable within organization
- May cache parts of the public graphs (more details next!)
We're GUAC Beta v0.1

Provide a deployable service that an organization will be able to answer the vulnerability question (and more) about their software dependencies.

- Provides:
  - E2E Service with a web interface for users to query by software identifiers (/SBOMs)
  - GUAC consumes/ingests organization SBOMs and SLSA documents
  - GUAC augments the graph automatically with Open Source Insights and security metadata (deps.dev, osv.dev, scorecards)
  - Web UI for exploring the graph
Call to action

- Project hosted at: https://github.com/guacsec/guac
- GUAC Community mailing list: https://groups.google.com/g/guac-community
- GUAC Community meetings monthly (Calendar Invite, Recordings: https://www.youtube.com/@guacsec)
- We are looking for devs/contributors! (issue#1)
- The project will have Technical Advisory Members which are to inform the project. It is still a developer first project. (issue#1)
- Kubecon 2022 Talk: https://sched.co/182Jr