Lessons Learned from the OWASP Amass Project

Legal-Entity Driven Outside-In and Bottom-Up Discovery

Speaker: Jeff Foley

Date: October 8-9, 2025

Event: Rochester Security Summit 2025

Speaker

Jeff Foley

Project Leader, OWASP Amass

Over 20 years of experience focused on applied research & development, security assessment, vulnerability management, and attack surface management (ASM):

Current Roles:

- Head of Research at KYND
- Vice Chair of the OWASP Projects Committee
- Passionate Mapper of Unseen Parts of the Internet

Previous Roles:

- Vice President of Research at ZeroFox
- Global Head of Attack Surface at Citigroup
- Global Manager of Vuln Engineering at National Grid
- Program Manager for Offensive Cyber Warfare Research & Development at Northrop Grumman Corporation

GitHub: @caffix
Discord: @caffix
LinkedIn: in/caffix
X/Twitter: @jeff_foley
Mastodon: @caffix@infosec.exchange

What is Bottom-Up Discovery?

- Bottom-Up discovery is typically performed with internet-wide scans.
- These scans trade precision for coverage by attempting each publicly reachable IP address.
- The process is capable of exposing architecture and infrastructure not discovered by top-down methods alone.
- Also, tends to reveal shadow IT that is publicly reachable, yet not assigned public DNS names.

What is Outside-In Discovery?

- **Seeded with corporate data, such as legal names and street addresses for the target organization.**
- **Expanded via legal-entity identifiers and internet resource registration data.**
  - For example, the GLEIF LEI and and company registration jurisdiction and company ID.
  - WHOIS records and RIR registration information (RDAP).
- **When your enumeration workflow returns to the infrastructure, you’re likely to have additional IP addresses to investigate.**

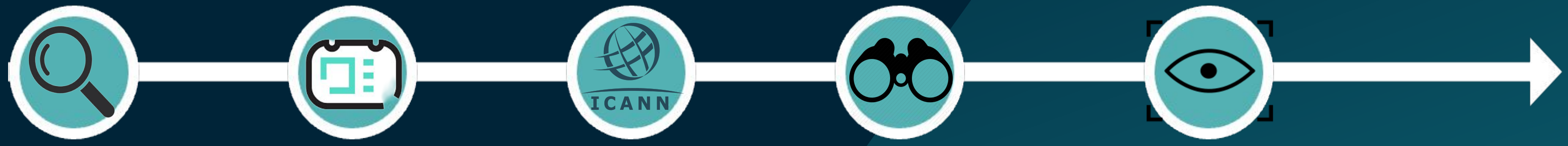
Why Bottom-Up Adds So Much Value?

- Starts from known good data.
  - Corporate legal-entities, IPs, and CIDRs.
- Expands via passive DNS, CT logs, RIR RDAP records, WHOIS, ASNs.
  - Recursively discovers related assets.
- Context-aware filtering.
  - Assets are more likely to be owned or associated with the target organization.
- Typically, uncovers staging environments, old M&A infrastructure, or cloud services spun up without review.

Targeted Internet-Wide Scanning Techniques

- **Tools: ZMap/Masscan sweep 65K ports and all IPv4 addresses can take hours to days!**
  - Successful implementation of internet-wide scanning techniques require dedicated, and properly designed, infrastructure using high-bandwidth cloud instances.
  - Compliance guardrails: rate-limit, abuse email handling, exclusion lists.
- **Legal-entity data narrows the focus to relevant ASNs/CIDRs.**
  - Can be performed from a variety of infrastructures and/or ISPs.
  - Sweeping-with-sampling: probe sample IPs -> density estimate -> selective scan.

Outside-In + Bottom-Up Discovery Process



Associated Organizations

Transform seed data to orgs that registered the resources

Legal Entities

Obtain legal entity IDs for all resource owners

Discover Registered Infrastructure

Query for internet infra using RIR RDAP for all legal entities

Sweep Infrastructure Netblocks

Apply sweeping and sampling with deeper dives around discoveries

Discover Namespaces & Assets

Reverse DNS, cert pulls, content discovery, and port scans reveal assets

Associated Organizations

- The first stage obtains the name (preferably legal name) of the organization that owns the domains of interest.
- Pulling organization names from OV and EV TLS certificates within the namespace.
- Pulling registration information from on-prem networks.
  - Autonomous system and network registrations usually include references to entities that contain legal names.
- Using content discovery from a website in the target namespace.
  - Web page footers often include the company name.
  - About and Company pages can include the legal name.
- Or, simple Google Dorks and AI queries that make the association between websites and the legal-entities that owns them.

Legal Entities

- Brand names do not tell the whole story when attempting to reveal an organization’s presence on the Internet.
- A company registered across the globe can have different legal names in different regions of the world.
- Internet resources are typically registered under an organization’s legal name.
- GLEIF and OpenCorporates both provide search capabilities for registered legal-entities.
- The two services also link legal-entities to a parent organization and subsidiaries.
- This allows the discovery process to map out an entire enterprise of related companies.

GLEIF

- Global Legal-Entity Identifier Foundation - gleif.org
- GLIEF attempts to connect the dots across the universe of entity identification.
- The GLEIF API (https://www.gleif.org/en/lei-data/gleif-api) is freely available to everyone in an attempt to provide transparency.
- GLEIF has less coverage than OpenCorporates, and primarily appeals to organizations utilizing certain payment methods, since LEIs are necessary in some cases.
- The foundation’s coverage will increase as the use of LEIs expands into additional use cases.
- The GLEIF API provides “fuzzy” matching of names and addresses, and the results often require additional checks to reach high confidence levels.

OpenCorporates

- Legal-entity data you can trust - opencorporates.com
- OpenCorporates can founded in 2010 to transform corporate transparency, specifically by making legal-entity data more accessible.
- The OpenCorporates API (https://api.opencorporates.com/) is not free, and starts at a price of ~$1K per month.
- OpenCorporates has data for over 200 million companies, yet has some serious blind spots, such as most countries in Asia.
- Unlike GLEIF, OpenCorporates coverage is not determined by voluntary participation, but instead by their active investigation of legal-entity jurisdictions.
- Similar to GLEIF, OpenCorporates has an API endpoint that provides matching of company names, but also requires scrutiny.

Legal-Entity Discovery Workflow

1. Use brand and company names from the previous stage of the bottom-up discovery process as search criteria in the GLEIF and OpenCorporates fuzzy search APIs.
2. Filter results from Step 1 to ensure they have matching legal names and addresses.
3. Using a legal-entity from Step 2, follow the references to the parent company and all subsidiaries.
4. For each legal-entity, capture its legal name, street address, jurisdiction, and registration number.
5. At this point, you have the tree of organizations making up the enterprise and/or international corporation of interest.

Discover Registered Infrastructure

- The Registration Data Access Protocol (RDAP) is a modern protocol designed to replace the older WHOIS protocol for accessing registration data about internet resources.
- Regional Internet Resources (RIRs) maintain their own data stores for registrations of internet resources, and are required to expose portions of the data set via RDAP.
- The entities that registered the internet resources are included in the data sets, and accessible when querying the RIR data stores.
- The American Registry for Internet Numbers (ARIN) RIR provides a search API.
- The results from the ARIN search API can direct your query to the appropriate RIR RDAP server.

ARIN RIR Search API

- **Use ARIN’s entities endpoint to search for legal-entities discovered from the previous stage of the bottom-up discovery process.**
- **Curl Example:**

  ```
  curl -H “host: rdap.arin.net” -H “origin: https://search.arin.net” -H
  “referer: https://search.arin.net”
  https://rdap.arin.net/registry/entities/?&fn=Salesforce.com
  ```

- **Each of the entities returned has a RDAP handle, which can be further investigated using the entity endpoint.**
- **Each entity query will include related entities**

Sweep Infrastructure Netblocks

- Once at this stage of the process, registered netblock will have been identified.
- Smaller networks can easily be swept in their entirety, while larger (e.g. /16 networks) can still be expensive to scan.
- For these larger netblocks, other effective strategies can be employed for sweeping.

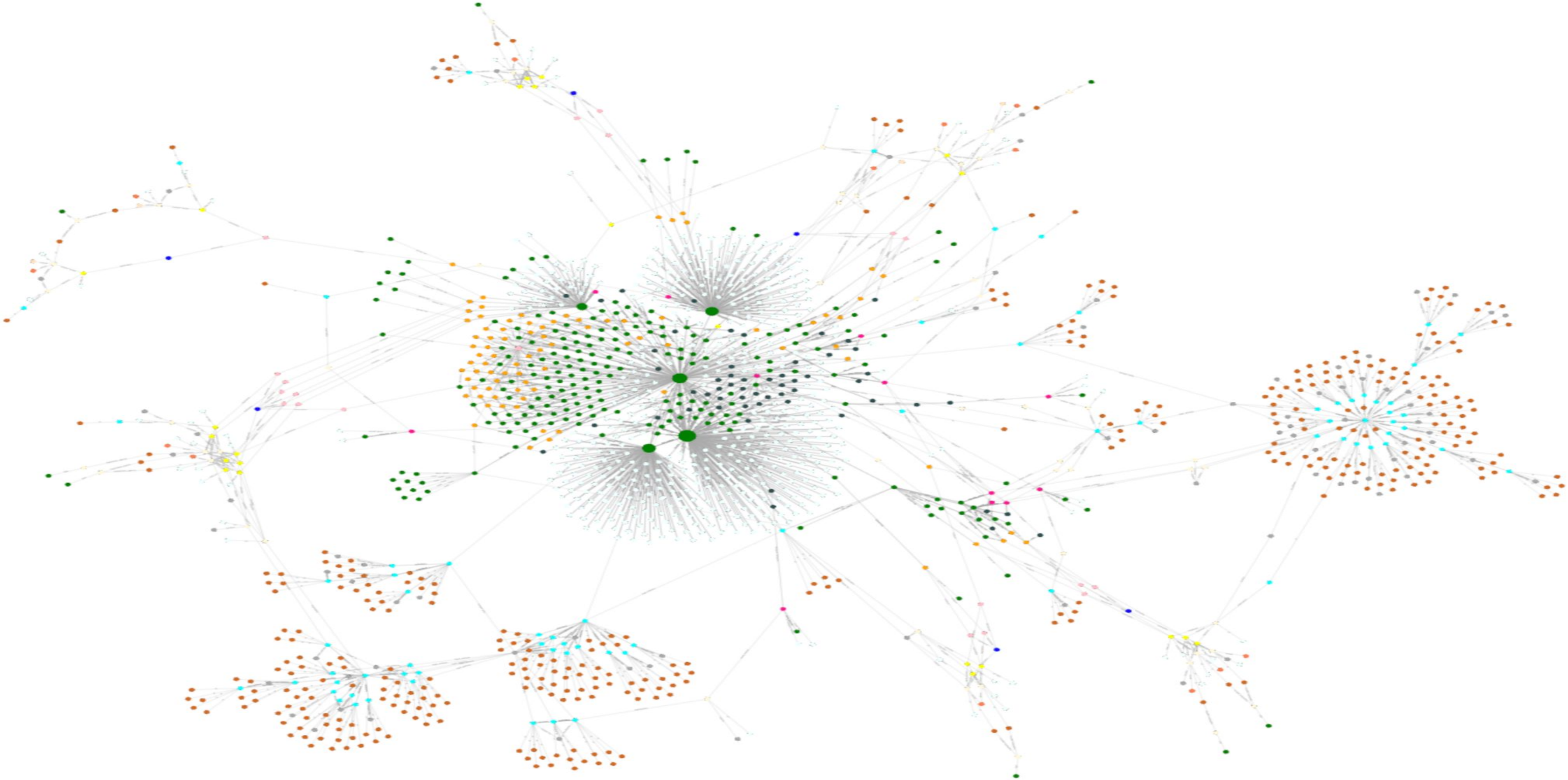
Discover Namespaces & Assets

- For each system that responded during the sweeping and sampling, checks can be made to help reveal information about the discovered asset.

- These methods have been ordered from least to most expensive:
  - Perform reverse DNS
  - Attempt to pull a TLS certificate
  - Execute port scanning with banner grabbing
  - Perform content discovery on web servers.

- Discovered assets will not necessarily have publicly exposed DNS names.

Visualization of the Asset Map



Thank you.
Questions?