GitHub Setup

CBOM Secure · Source Code Sensor (Discover_SourceCode)

Overview

The CBOM Secure Discover_SourceCode sensor scans GitHub repository contents to discover hard-coded cryptographic material in source, configuration, and infrastructure-as-code files:

- Hard-coded private keys, API keys, and access tokens
- X.509 certificates and chains in source or configuration
- Symmetric keys, secrets, and passphrases in code or dotfiles
- SSH keys and PEM material, and .env files containing secrets

Covers both GitHub.com and GitHub Enterprise Server (GHES). All access is read-only via a Personal Access Token (PAT).

Prerequisites

- A GitHub account with access to the org/repositories to scan (org member for org-wide scans).
- For GHES: connectivity from the sensor host to https://<ghes-host>/api/v3.
- The Discover_SourceCode sensor binary on the sensor host.
- Outbound HTTPS (443) to api.github.com or your GHES hostname.
- Permission to create a PAT (or admin approval if restricted).

Step-by-Step Guide

Step 1: Create a Personal Access Token

A fine-grained PAT is recommended (least privilege). Under Settings → Developer settings → Personal access tokens → Fine-grained tokens → Generate new token: select the resource owner, choose Only select repositories (or All), and set Repository permissions Contents: Read-only and Metadata: Read-only. Copy the token immediately. A classic PAT with the repo scope also works where fine-grained tokens are unavailable.

Step 2: (GHES) Identify the API Endpoint

The GHES REST API base URL is https://<GHES_HOSTNAME>/api/v3. Verify connectivity:

```
curl -H "Authorization: token <YOUR_PAT>" \
  https://github.corp.example.com/api/v3/user
```

Step 3: Verify API Access

```
curl -s -H "Authorization: token <YOUR_PAT>" https://api.github.com/user | jq .login
curl -s -H "Authorization: token <YOUR_PAT>" \
  "https://api.github.com/orgs/<YOUR_ORG>/repos?type=all&per_page=5" | jq ".[].full_name"
```

Step 4: Configure the CBOM Secure Sensor

```
sensor:
  name: Discover_SourceCode
  enabled: true
  schedule: "0 2 * * *"
source:
  type: github
  github_token: "${CBOM_GITHUB_TOKEN}"
  # api_base_url: "https://github.corp.example.com/api/v3"   # GHES only
  org: "your-github-org"
  repos: ["your-github-org/backend-api", "your-github-org/infra-terraform"]
  branch: "main"
  include_patterns: ["**/*.pem", "**/*.key", "**/*.crt", "**/*.env",
    "**/*.yaml", "**/*.json", "**/*.tf", "**/*.py", "**/*.js", "**/*.go"]
  exclude_patterns: ["**/vendor/**", "**/node_modules/**", "**/.git/**"]
  max_file_size_bytes: 5242880
  tls_verify: true
output:
  format: cbom-json
  cbom_api_endpoint: "https://cbom.corp.example.com/api/v1/ingest"
export CBOM_GITHUB_TOKEN="ghp_YourPersonalAccessTokenHere"
cbom-sensor start --config /etc/cbom/sensors/discover_sourcecode.yaml
```

Step 5: Validate

```
cbom-sensor run --config /etc/cbom/sensors/discover_sourcecode.yaml --dry-run
tail -50 /var/log/cbom/discover_sourcecode.log
```

Confirm authentication, repository enumeration, and a findings summary; verify findings under Assets → Source Code.

Common Errors

401 Unauthorized — Bad credentials

Cause: The PAT is invalid/expired or not passed correctly.

Resolution: Confirm CBOM_GITHUB_TOKEN is set with no whitespace; check the token's expiry and regenerate if needed.

403 Forbidden — Resource not accessible

Cause: Insufficient scope (missing Contents: Read or repo), or org SAML SSO not authorized for the token.

Resolution: Add the required permission; if the org uses SAML SSO, click Authorize next to the org on the token page.

TLS certificate verify failed (GHES)

Cause: GHES uses an internal CA not in the sensor host trust store.

Resolution: Add the CA to the system bundle or set tls_ca_bundle; never set tls_verify: false in production.

Security Recommendations

- Use a dedicated GitHub service account, not a personal developer account.
- Use fine-grained PATs scoped to only the repositories scanned with Contents/Metadata: Read.
- Rotate tokens every 90 days and automate via your secrets manager.
- Store the PAT in a secrets manager — never plaintext config or shell history.
- Enable GitHub audit log streaming and alert on unexpected API activity.

Conclusion

With a read-only PAT and least-privilege configuration, the Discover_SourceCode sensor maintains an up-to-date inventory of cryptographic assets in GitHub.com or GHES repositories, helping teams find and remediate exposed keys, certificates, and secrets before they are exploited.