Local Git Repositories & Source Archives Setup

CBOM Secure · Source Code Sensor (Discover_SourceCode)

Overview

The CBOM Secure Discover_SourceCode sensor scans local filesystem Git repositories and source archives for hard-coded cryptographic material — private keys, certificates, PEM blocks, API tokens, and symmetric keys — committed to version control or packaged into releases. It operates entirely at the filesystem level, read-only, with no network credentials required.

Prerequisites

- The CBOM sensor agent installed on a host with filesystem access to the repos/archives.
- The sensor service account (cbom-sensor) exists on the host.
- Git installed if branch/history scanning is required.
- Target repos cloned/mirrored locally; archives (.zip/.tar.gz) in a known directory.
- Disk space for --mirror clones if full-history scanning; Python 3.8+ / agent runtime.

Step-by-Step Guide

Step 1: Grant Filesystem Read Access

```
ps aux | grep cbom                          # identify the service account
sudo setfacl -R -m u:cbom-sensor:rX /opt/repositories/
sudo setfacl -R -m u:cbom-sensor:rX /var/source-archives/
sudo -u cbom-sensor ls -la /opt/repositories/my-app     # confirm access
```

Step 2: Prepare Repositories

For current-HEAD scanning, a standard clone suffices. For full-history coverage (material introduced and later deleted), use a mirror clone — the sensor scans bare repos when include_bare_repos: true:

```
git clone --mirror https://github.com/your-org/your-repo.git /opt/repositories/your-repo.git
git -C /opt/repositories/your-repo.git remote update
```

Step 3: Prepare Source Archives

```
sudo mkdir -p /var/source-archives && sudo chown cbom-sensor:cbom-sensor /var/source-archives
cp /home/builds/release-v2.4.1.tar.gz /var/source-archives/
tar -tzf /var/source-archives/release-v2.4.1.tar.gz | head -20    # verify integrity
```

Step 4: Configure the CBOM Secure Sensor

```
sensors:
  - name: Discover_SourceCode
```

```
  enabled: true
  sensor_type: source_code
  schedule: "0 2 * * *"
  source:
    local_paths: ["/opt/repositories/", "/var/source-archives/"]
    recursive: true
    follow_symlinks: false
    include_bare_repos: true
    git_branches: [all]
    scan_archives: true
    archive_extensions: [.zip, .tar.gz, .tar.bz2]
    include_patterns: ["**/*.pem", "**/*.crt", "**/*.key", "**/*.p12",
      "**/*.pfx", "**/*.jks", "**/*.conf", "**/*.yaml", "**/*.env",
      "**/*.json", "**/*.py", "**/*.java", "**/*.go", "**/*.js"]
    exclude_patterns: ["**/node_modules/**", "**/vendor/**", "**/dist/**"]
    max_file_size_bytes: 10485760
  output:
    findings_format: json
    output_path: /var/cbom/output/source_code_findings.json
  service_account: cbom-sensor
sudo systemctl restart cbom-agent
```

Step 5: Validate

```
sudo journalctl -u cbom-agent -f | grep Discover_SourceCode
cbom-agent run-sensor --name Discover_SourceCode --once
cat /var/cbom/output/source_code_findings.json | python3 -m json.tool | head -60
```

Confirm repositories and archives are scanned and that findings include asset_type values (e.g. RSA_PRIVATE_KEY, X509_CERTIFICATE) with file path and line number.

Common Errors

Permission denied accessing repository path

Cause: cbom-sensor lacks read access to the path.

Resolution: Grant rX via setfacl (or group perms) and confirm with sudo -u cbom-sensor ls.

Archive extraction failure — corrupt/incomplete

Cause: The archive is incomplete or was still being written.

Resolution: Verify with tar -tzf; schedule scans after builds complete and files are flushed.

Git repo detected but no branches scanned

Cause: A bare/mirror clone without include_bare_repos, or an empty repo.

Resolution: Set include_bare_repos: true; verify with git -C <repo>.git branch -a.

Security Recommendations

- Grant cbom-sensor read-only access strictly to the configured local_paths.
- Do not run the agent as root; use a dedicated low-privilege account with scoped ACLs.
- Store mirror clones on a dedicated scanning volume, separate from production repos.

- Review the local_paths list regularly as projects are onboarded/decommissioned.
- Restrict the findings output (chmod 640) — it may contain partial key material.

Conclusion

The Discover_SourceCode sensor scans local Git repositories and source archives read-only, with no network credentials — and, using mirror clones for full history, surfaces cryptographic material across the entire codebase for governance in CBOM Secure.