BIT

Dokumentácia projektu

Orchestrator a prioritizačný systém pre autorizované sieťové skenovanie

Samuel Gabriel Galgóci – 121331

23. 11. 2025

Obsah

1 Úvod

Cieľom projektu je vytvoriť bezpečný a konfigurovateľný nástroj na **autorizované skenovanie siete**, ktorý poskytuje kompletnú pipeline od zberu údajov až po ich vizualizáciu a prioritizáciu. Systém je navrhnutý tak, aby bol modulárny, znovupoužiteľný a jednoducho rozšíriteľný. Projekt zabezpečuje najmä spúšťanie Nmap skenov podľa definovaných jobov, normalizáciu výsledkov do jednotného dátového modelu, **obohatenie** zistených služieb o informácie z databáz zraniteľností (CVE), vyhodnocovanie rizika pre jednotlivé služby a hosty a zobrazenie výsledkov cez dashboard.

Celý systém je postavený na Docker Compose infraštruktúre a skladá sa z viacerých kontajnerov:

- **Scan (orchestrator)** – spúšťanie Nmap skenov podľa YAML konfigurácie,
- **Parser (ingest)** – spracovanie Nmap XML a uloženie výsledkov do PostgreSQL,
- **Enricher (CVE enricher)** – doplnenie CVE a EPSS informácií k službám,
- **Dashboard** – vizualizácia hostov, služieb, zraniteľností a rizík.

Dáta sú ukladané do PostgreSQL databázy. Moduly sú navrhnuté tak, aby medzi sebou komunikovali cez jasne definované rozhrania (súbory, databáza, skriptové rozhrania) a dali sa spúšťať samostatne aj ako súčasť orchestrácie.

2 Architektúra systému

Architektúra systému je rozdelená do niekoľkých hlavných komponentov, ktoré sú nasadené ako samostatné kontajnery. Každý komponent má jasne definovanú zodpovednosť.

Orchestrator (Scan modul)

- Hlavný skript orchestrator/scan_orchestrator.py.
- Číta YAML joby (zoznam cieľov, portov, časovanie, výstupné adresáre).
- Spúšťa Nmap nad jednotlivými hostmi alebo rozsahmi a generuje XML a log súbory do `data/raw/`.

Parser / Ingest modul

- Skript ingest/parse_nmap.py.
- Prechádza Nmap XML súbory, extrahuje informácie o hostoch a službách.

- Ukladá údaje do PostgreSQL databázy (tabuľky typu `hosts`, `services` a pod.).

Enricher modul (CVE Enricher)

- Skript enrich/cve_enricher.py.
- Číta služby z databázy a lokálny cache CVE cve_cache.json.
- Na základe portu, protokolu, produktu a verzie priraďuje CVE záznamy.
- Výsledky ukladá do tabuľky vulnerabilities.

Risk scoring

- Skript scoring/risk_score.py.
- Počítanie rizikovej hodnoty pre služby a hosty podľa politiky definovanej napr. v RISK_SCORING.md.
- Výstupom je jednotná číselná metrika rizika vhodná na prioritizáciu.

Dashboard

- Streamlit aplikácia dashboard/app.py.
- Pripája sa na databázu a poskytuje prehľadné zobrazenie hostov, služieb, zraniteľností a rizík.
- Umožňuje ovládanie pipeline (spúšťanie skenov, parsovanie, enrichment, scoring).

Databáza a lab hosty

- PostgreSQL kontajner so schémou definovanou v db/schema.sql.
- Sada testovacích hostov (host01 – host20) s rôznymi službami a zraniteľnosťami.

3 Scan modul (Orchestrator)

3.1 Účel

Scan modul je zodpovedný za **autorizované sieťové skenovanie**. Jeho cieľom je:

- spúšťať Nmap s jednotnou a auditovateľnou konfiguráciou,
- ukladať výstup v štandardizovanom formáte (XML),
- evidovať jednotlivé joby a skeny tak, aby boli reprodukcibilné a dohľadateľné.

3.2 Vstupy a výstupy

Vstupy

- YAML joby (napr. `jobs/internal_quickscan.yaml`),
- parametre ako rozsahy IP alebo názvy hostov, porty a Nmap switche.

Výstupy

- Nmap XML súbory v priečinku `data/raw/<job_name>/`,
- logy zo skenovania (štandardný výstup a chybový výstup),
- prípadné záznamy o priebehu jobu v databáze alebo log súboroch.

3.3 Implementačné detaily

Scan modul beží v samostatnom kontajneri. Každý job sa vykonáva ako sekvencia skenov na úrovni definovaných hostov alebo rozsahov, čo umožňuje:

- lepšie riadenie časovania a retry mechanizmov,
- jednoduchšie logovanie a analýzu chýb,
- možnosť ľahko rozšíriť funkcionalitu o ďalšie typy skenov.

4 Parser modul (Ingest)

4.1 Účel

Parser slúži na transformáciu Nmap XML výstupov do relačného dátového modelu. Hlavné ciele sú:

- extrahovať z XML len relevantné údaje,
- uložiť ich do PostgreSQL v normalizovanej podobe,
- pripraviť dáta pre ďalšie spracovanie (CVE a EPSS enrichment, scoring, dashboard).

4.2 Vstupy a výstupy

Vstupy

- XML súbory generované Nmapom (`data/raw/<job_name>/*.xml`),
- DSN na databázu `postgresql://user:pass@db:5432/scans`.

Výstupy Parser zapisuje dáta do viacerých tabuliek:

- `hosts` – informácie o hostoch (IP, hostname, OS, `last_seen`),
- `services` – informácie o službách (port, protokol, služba, produkt, verzia,

4.3 Implementačné detaily

Skript `ingest/parse_nmap.py`:

- prechádza súbory podľa glob patternu,
- validuje XML a ošetruje základné chyby (napr. neúplné reporty),

- používa SQL schému definovanú v `db/schema.sql`.

5 Enricher modul (CVE Enricher)

5.1 Účel

Enricher rozširuje základné informácie o službách o **zraniteľnosti (CVE)**. Vďaka tomu vzniká centrálne miesto, kde sú zoskupené:

- nájdené služby,
- známe zraniteľnosti (CVE) viazané na tieto služby,

5.2 Vstupy a výstupy

Vstupy

- tabuľky `hosts` a `services` v databáze,
- lokálny cache CVE `enrich/cve_cache.json`.

Výstupy

- tabuľka vulnerabilities, ktorá obsahuje:
  - väzbu na `service_id`,
  - identifikátor `cve_id`,
  - hodnotenie `cvss`,
  - doplnkové metadáta (popis, zdroj a pod.).

5.3 Implementačné detaily

Skript `enrich/cve_enricher.py` sa spúšťa samostatne, typicky ako `docker compose run -rm enricher`. Matching zraniteľností prebieha podľa kombinácie:

- port a protokol,
- produkt a verzia,

6 Dashboard

6.1 Účel

Dashboard je používateľské rozhranie postavené na **Streamlit**, ktoré spája technickú pipeline s používateľom. Slúži na:

- spúšťanie jednotlivých krokov pipeline (scan, parse, enrich, score),
- vizuálny prehľad o hostoch, službách, zraniteľnostiach a rizikách,
- jednoduchšiu analýzu výsledkov a ich prioritizáciu.

6.2 Funkcionalita

Dashboard obsahuje viacero typov pohľadov a ovládacích prvkov.

Ovládanie pipeline

- tlačidlá typu *Run Scan*, *Parse Nmap XML*, *CVE Enrich*, *Risk Scoring*,
- zobrazenie konzolových výstupov (stdout/stderr) pre lepšie debugovanie,
- možnosť opakovaného spustenia jednotlivých krokov podľa potreby.

Dátové pohľady

- **Hosts** – zoznam hostov, počet služieb, maximálne riziko,
- **Services** – zoznam služieb, porty, bannery, verzie, risk skóre,
- **Vulnerabilities** – filtrovanie a vyhľadávanie CVE, EPSS, možný export do CSV,
- **Top Risks** – najrizikovejšie služby alebo hosty,
- **Scan Jobs** – história spustených jobov s časom a konfiguráciou.

Údržba a administrácia

- možnosť čistenia raw súborov v `data/raw/**`,
- `truncate` vybraných tabuliek v databáze (napr. pre lab scenáre),

7 Záver

Projekt poskytuje kompletnú pipeline od autorizovaného skenovania siete až po prioritizáciu zraniteľností a ich vizualizáciu. Modularita systému (scan, parser, enricher, scoring, dashboard) umožňuje:

- samostatné nasadenie a škálovanie jednotlivých komponentov,
- jednoduché pridávanie nových zdrojov dát (ďalšie skenery, nové CVE zdroje),
- postupné vylepšovanie bez zásadných zásahov do architektúry.

Vďaka kontajnerizácii a zdieľanému dátovému modelu je možné systém nasadiť v izolovanom lab prostredí s testovacími hostami, ale aj v reálnej infraštruktúre. Projekt môže slúžiť ako základ pre pokročilejší vulnerability management alebo pre integráciu do SOC/SIEM prostredia, kde pridanou hodnotou je jednotné spracovanie skenov, zraniteľností a rizík v jednom konzistentnom nástroji.