Spracovanie sieťových logov v ElasticSearch

Obsah

1. Úvod do spracovania sieťových logov 3

2. Elasticsearch a jeho úloha pri spracovaní sieťových logov 4

3. Alternatívy k Elasticsearch pre spracovanie sieťových logov 5

3.1 Splunk 5

3.2 Graylog 6

3.3 Loggly 6

3.4 Fluentd + Grafana + Loki 6

4. Porovnanie výhod a nevýhod jednotlivých riešení 6

4.1 Splunk pri spracovaní sieťových logov 6

4.2 Elasticsearch pri spracovaní sieťových logov 7

4.3 Graylog pri spracovaní sieťových logov 7

4.4 Loggly pri spracovaní sieťových logov 8

5. Vlastnosti a požiadavky multitenantných riešení 8

5.1 Požiadavky na multitenantné riešenie 9

5.2 Elasticsearch v multitenantnom režime 9

5.3 Multitenancia v ostatných riešeniach 9

6. Záver a odporúčania 10

6.1 Kľúčové zistenia 10

6.2 Odporúčania 10

1. Úvod do spracovania sieťových logov

V dnešnom digitálnom svete je nepretržité sledovanie a analýza sieťovej prevádzky nevyhnutnou súčasťou správy IT infraštruktúry. Každý systém pripojený do siete generuje množstvo údajov vo forme sieťových logov – záznamov o sieťovej komunikácii, pripojeniach, požiadavkách či udalostiach súvisiacich so zabezpečením. Tieto logy pochádzajú napríklad z firewallov, smerovačov, serverov, IDS/IPS systémov alebo prevádzkových monitorovacích nástrojov.

Spracovanie sieťových logov umožňuje:

- **Detekciu a prevenciu hrozieb** – včasná identifikácia útokov, ako sú DDoS, port scanning či neautorizovaný prístup;
- **Monitorovanie siete** – sledovanie stavu a zaťaženia infraštruktúry;
- **Audity a forenznú analýzu** – rekonštrukciu udalostí po bezpečnostných incidentoch;
- **Splnenie legislatívnych požiadaviek** – záznamy o činnostiach potrebné pre GDPR, ISO 27001 a podobné normy.

Vzhľadom na rastúce objemy sieťových dát je efektívne spracovanie logov výzvou. Tradičné databázy často nestačia kvôli nedostatočnému výkonu pri indexovaní a vyhľadávaní veľkého množstva údajov. Preto sa stále častejšie využívajú špecializované nástroje ako **Elasticsearch**, ktoré sú určené na distribuované spracovanie a analýzu logov v reálnom čase.

Cieľom tejto práce je preskúmať možnosti použitia Elasticsearch pri spracovaní sieťových logov, popísať architektúru riešenia, spôsob spracovania a vizualizácie údajov, ako aj výhody a obmedzenia tohto prístupu v porovnaní s alternatívami.

2. Elasticsearch a jeho úloha pri spracovaní sieťových logov

Elasticsearch je výkonný, distribuovaný vyhľadávací a analytický engine, založený na knižnici Apache Lucene. V oblasti spracovania sieťových logov zohráva kľúčovú úlohu pri ukladaní, indexovaní a analyzovaní dát v reálnom čase. V rámci **Elastic Stacku** (ELK: Elasticsearch, Logstash, Kibana) slúži ako centrálne úložisko a vyhľadávací engine pre sieťové udalosti pochádzajúce z firewallov, IDS/IPS systémov, routerov, prepínačov, proxy serverov a ďalších zariadení.

Kľúčové vlastnosti Elasticsearch pri spracovaní sieťových logov:

- **Rýchle vyhľadávanie a analýza**: Elasticsearch umožňuje okamžité vyhľadávanie na základe IP adries, portov, protokolov, krajiny pôvodu a časových značiek, čo je dôležité pri detekcii sieťových hrozieb.
- **Horizontálne škálovanie**: Vďaka distribuovanej architektúre je schopný spracovávať milióny sieťových udalostí denne, čo je zásadné pre väčšie siete a organizácie.
- **Flexibilný dátový model**: Umožňuje ukladať semištruktúrované dáta (napr. Netflow, syslog, Suricata JSON) bez potreby pevnej schémy.
- **Integrácia so zberovými nástrojmi**: V spolupráci s nástrojmi ako Filebeat, Logstash alebo Packetbeat dokáže ingestovať a transformovať rôznorodé sieťové logy do použiteľného formátu.
- **Vizualizácia cez Kibana**: Umožňuje vytvárať prehľadné dashboardy zobrazujúce počet pripojení, anomálie, geografické rozdelenie zdrojov útokov a ďalšie.

Výzvy a obmedzenia:

- **Zložitosť konfigurácie a správy**: Riadenie veľkých Elasticsearch klastrov s viacerými indexmi pre rôzne logové zdroje si vyžaduje odborné znalosti.

- **Náročnosť na infraštruktúru**: Pri veľkých objemoch dát (napr. Netflow z celopodnikovej siete) sú potrebné silné diskové a pamäťové zdroje.
- **Bezpečnosť a izolácia dát**: Bez použitia X-Pack alebo OpenSearch Security môže byť potrebné samostatne riešiť autentifikáciu, autorizáciu a šifrovanie.

Príklady využitia pri sieťových logoch:

- **Detekcia anomálií**: Zaznamenanie veľkého počtu spojení z jedného zdroja (napr. DDoS, port scanning).
- **Vyhľadávanie podľa IP alebo subnetu**: Rýchle identifikovanie škodlivej aktivity podľa známej IP adresy.
- **Geolokácia spojení**: Identifikovanie podozrivých spojení z krajín, ktoré sa bežne nepripájajú.
- **Forenzná analýza**: Retrospektívne preskúmanie udalostí podľa časovej osi a kombinácie viacerých filtrov.

3. Alternatívy k Elasticsearch pre spracovanie sieťových logov

Aj keď je Elasticsearch jedným z najpoužívanejších nástrojov na spracovanie a analýzu logov, existuje viacero alternatívnych riešení, ktoré poskytujú podobnú funkcionalitu – niektoré s vyššou mierou jednoduchosti, iné s rozšírenými bezpečnostnými možnosťami či integrovanými analytickými funkciami.

Pri výbere technológie na spracovanie sieťových logov je dôležité zohľadniť faktory ako:

- schopnosť pracovať s formátmi ako **Syslog**, **Netflow**, **Suricata** alebo **Zeek**,
- výkonnosť pri spracovaní veľkého množstva logov v reálnom čase,
- úroveň vizualizácie a korelácie udalostí,
- podporu multitenantného alebo distribuovaného prostredia.

Medzi hlavné alternatívy k Elasticsearch patria najmä:

3.1 Splunk

Splunk je komerčný nástroj zameraný na spracovanie a analýzu veľkého množstva dát vrátane sieťových logov. Podporuje ingestovanie logov z rôznych zdrojov, detekciu hrozieb,

vizualizáciu a tvorbu korelačných pravidiel. Disponuje výkonným query jazykom (SPL – Search Processing Language).

3.2 Graylog

Graylog je open-source riešenie navrhnuté pre centralizovanú správu logov. Umožňuje zber, normalizáciu, analýzu a alertovanie na základe logov. Je vhodný pre sieťové logy vďaka natívnej podpore formátu Syslog a integrácii s nástrojmi ako Suricata alebo Zeek.

3.3 Loggly

Loggly (od spoločnosti SolarWinds) je cloudová služba na spracovanie logov. Podporuje široké spektrum vstupov vrátane sieťových zdrojov, jednoduché vyhľadávanie a vytváranie vizualizácií. Je vhodný najmä pre menšie tímy a rýchle nasadenie.

3.4 Fluentd + Grafana + Loki

Alternatívou ku klasickému ELK Stacku je kombinácia nástrojov **Fluentd** (na zber logov), **Grafana Loki** (na indexovanie a ukladanie) a **Grafana** (na vizualizáciu). Tento prístup poskytuje ľahšie nasaditeľné a menej náročné riešenie pre cloud-native infraštruktúry.

4. Porovnanie výhod a nevýhod jednotlivých riešení

V tejto kapitole sa zameriame na porovnanie výhod a nevýhod štyroch hlavných riešení na spracovanie transakčných logov: **Splunk**, **Elasticsearch**, **Graylog** a **Loggly**.

4.1 Splunk pri spracovaní sieťových logov

Splunk je robustné komerčné riešenie určené na analýzu logovacích a telemetrických dát. V oblasti sieťovej bezpečnosti a spracovania sieťových logov sa používa najmä na:

- detekciu útokov,
- správu udalostí (SIEM),
- koreláciu údajov z viacerých zdrojov (firewally, IDS/IPS, proxy, VPN, atď.),
- tvorbu prispôsobených detekčných pravidiel.

Výhody:

- Výkonný query jazyk (SPL) s možnosťou korelácie viacerých typov údajov.
- Intuitívne používateľské rozhranie s pokročilými dashboardmi.

- Vysoká miera prispôsobiteľnosti – alerty, detekcie, Machine Learning add-ony.
- Podpora pre veľké prostredia, integrácie s cloudom, AD, syslog, Netflow, atď.

Nevýhody:

- Licencovanie na báze objemu dát – môže byť veľmi nákladné pri veľkých sieťach.
- Vyššie hardvérové nároky pri veľkom množstve indexovaných logov.
- Uzavretý ekosystém – obmedzená otvorenosť v porovnaní s open-source riešeniami.

4.2 Elasticsearch pri spracovaní sieťových logov

Elasticsearch je často voľbou pre open-source riešenia na spracovanie sieťových logov. Používa sa v kombinácii s **Logstash**, **Filebeat**, **Packetbeat** a **Kibana**. Je vhodný na:

- centralizovaný zber logov (syslog, firewall, proxy, IDS...),
- real-time analýzu útokov,
- vizualizáciu tokov dát a podozrivých aktivít.

Výhody:

- Rýchle indexovanie a vyhľadávanie veľkých objemov logov.
- Horizontálne škálovateľný, vhodný pre distribuované systémy.
- Otvorený a modulárny – množstvo pluginov, Beats, integrácií.
- Vhodný pre real-time analýzu a vizualizáciu pomocou Kibany.

Nevýhody:

- Vyžaduje pokročilú konfiguráciu a znalosti pri väčších nasadeniach.
- Menej priamočiara korelácia medzi logmi rôznych typov bez vlastného vývoja.
- Bez X-Pack alebo OpenSearch Security má obmedzené zabudované zabezpečenie.

4.3 Graylog pri spracovaní sieťových logov

Graylog je open-source riešenie na centralizovanú správu logov, ktoré je obľúbené najmä pre svoju jednoduchosť, rozšíriteľnosť a dobrú podporu pre sieťové logy. Umožňuje prijímať údaje prostredníctvom **Syslog**, **Beats**, **REST API** a iných protokolov. Vhodný je pre analýzu logov z firewallov, routerov, VPN gateway, IDS/IPS či proxy serverov.

Výhody:

- Jednoduché nasadenie a správa – vhodný pre stredne veľké a menšie organizácie.

- Webové UI s možnosťou vyhľadávania, filtrácie a dashboardov.
- Podpora pre **stream-based alertovanie** – prichádzajúce logy sa vyhodnocujú v reálnom čase.
- Možnosť obohatenia logov (napr. o geolokačné údaje, DNS rozlíšenia).

Nevýhody:

- Menej výkonný pri veľmi veľkých objemoch logov v porovnaní s Elasticsearch alebo Splunk.
- Niektoré pokročilé funkcie dostupné iba v enterprise verzii.
- Menej rozsiahla dokumentácia a komunita v porovnaní s Elastic Stack.

4.4 Loggly pri spracovaní sieťových logov

Loggly je cloudové riešenie na centralizovaný zber a analýzu logov, ktoré je súčasťou portfólia SolarWinds. Je orientované najmä na jednoduchosť použitia, rýchle nasadenie a integráciu s aplikáciami a infraštruktúrou. Pre sieťové logy podporuje prijímanie údajov cez Syslog protokol alebo REST API.

Výhody:

- Rýchla konfigurácia bez potreby vlastnej infraštruktúry.
- Intuitívne UI so základnými grafmi, filtrami a vyhľadávaním.
- Podpora integrácií s rôznymi platformami (AWS, Heroku, Docker, atď.).

Nevýhody:

- Obmedzená funkcionalita pri pokročilej analýze alebo korelácii dát.
- Výrazne menej prispôsobiteľný ako Elasticsearch či Splunk.
- Cena závisí od množstva prijatých logov a požadovanej retencie

5. Vlastnosti a požiadavky multitenantných riešení

Multitenantné spracovanie logov znamená, že viacero nezávislých používateľov (tenants) – napríklad oddelení, pobočiek, zákazníkov alebo projektov – zdieľa jednu infraštruktúru pre zber, spracovanie a analýzu logovacích dát. V kontexte **sieťových logov** sú multitenantné riešenia dôležité najmä pre:

- **poskytovateľov spravovaných bezpečnostných služieb (MSSP)**,
- **organizácie s oddelenými bezpečnostnými zónami alebo infraštruktúrami**,
- **SaaS poskytovateľov**, ktorí potrebujú oddeliť prístupy k logom jednotlivých klientov.

5.1 Požiadavky na multitenantné riešenie

Účinné multitenantné spracovanie logov musí zabezpečiť:

- **Izoláciu dát** – každý tenant má prístup len k svojim vlastným logom;
- **Zdieľanie výpočtových zdrojov** – centrálna správa infraštruktúry pre všetkých tenantov;
- **Auditovateľnosť** – možnosť spätne skontrolovať, kto s dátami pracoval;
- **Bezpečný prístup** – detailné riadenie oprávnení (napr. pomocou RBAC/DLS);
- **Prispôsobiteľné vizualizácie** – každý tenant má vlastné dashboardy, alerty a filtre.

5.2 Elasticsearch v multitenantnom režime

Elasticsearch natívne nepodporuje multitenanciu ako izolovaný kontajnerový systém, ale umožňuje viacero prístupov:

- **Samostatné indexy pre každého tenanta** – každý tenant má vlastný logický index alebo index pattern (napr. customer1-*, deptA-*).
- **Zdieľaný index so štítkom (tenant_id)** – všetky dáta v jednom indexe, no prístup je filtrovaný pomocou **Document Level Security (DLS)**.
- **Integrácia s OpenSearch Security / Elastic Security** – tieto vrstvy umožňujú definovať oprávnenia až na úroveň poľa alebo typu záznamu.

Výzvou v prípade Elasticsearchu je správa veľkého množstva indexov a zabezpečenie škálovateľnosti bez ovplyvnenia výkonu.

5.3 Multitenancia v ostatných riešeniach

- **Splunk**: podporuje multitenantný prístup pomocou tzv. "apps" a "roles", pričom prístup k dátam sa obmedzuje na úrovni sourcetype, indexu alebo konkrétnej aplikácie.
- **Graylog**: poskytuje viacero "streams", pričom každému používateľovi je možné prideliť prístup len k vybraným streamom (filtrovaným podľa pravidiel).
- **Loggly**: ponúka obmedzenú multitenanciu – vhodnejší pre jedného klienta alebo jednoduché oddelenie dát.

6. Záver a odporúčania

V tejto práci sme sa zamerali na porovnanie možností spracovania **sieťových logov** s dôrazom na použitie nástroja **Elasticsearch** a jeho alternatív. Sieťové logy, ktoré pochádzajú z firewallov, IDS/IPS systémov, smerovačov, proxy serverov či iných bezpečnostných zariadení, tvoria kľúčový zdroj informácií pre zabezpečenie infraštruktúry, analýzu incidentov a dodržiavanie predpisov.

6.1 Kľúčové zistenia

- **Elasticsearch** sa ukazuje ako výkonné a flexibilné riešenie na spracovanie sieťových logov v reálnom čase, s podporou horizontálneho škálovania, rýchleho vyhľadávania a integrácie s vizualizačnými nástrojmi ako Kibana.
- **Alternatívy ako Splunk, Graylog či Loggly** ponúkajú rôzne stupne funkcionality – od enterprise-grade bezpečnostnej analytiky až po jednoduché cloudové log manažment systémy.
- **Multitenantné riešenia** sú čoraz dôležitejšie pre veľké organizácie a poskytovateľov služieb. Elasticsearch umožňuje viaceré prístupy k logickej separácii dát, no vyžaduje premyslený dizajn oprávnení a architektúry.

6.2 Odporúčania

- **Pre menšie tímy alebo jednoducho škálovateľné riešenie** odporúčame Elasticsearch v kombinácii s Filebeat/Logstash a Kibana ako otvorený a dobre dokumentovaný základ.
- **Pre enterprise prostredia so zložitými požiadavkami na audit a compliance** môže byť vhodnejší Splunk alebo Graylog Enterprise, najmä ak sa vyžaduje SIEM funkcionalita.
- **Pri návrhu multitenantného prostredia** je dôležité vopred definovať spôsob izolácie dát, plánovanie indexov, životného cyklu dát a systém oprávnení (napr. pomocou DLS/RBAC).
- **Bez ohľadu na použitý nástroj** by malo byť spracovanie sieťových logov súčasťou širšej bezpečnostnej stratégie organizácie – vrátane detekcie incidentov, forenznej analýzy a reakcie na hrozby.