FluentBit + OpenSearch + Dashboards

Acabem de veure com Falco pot ser una solució perfecta per implementar un **HIDS** (un "Host Intrusion Detection System"). Aquests tipus de programes permeten tenir una visibilitat en temps real de tots els events de seguretat que estan passant a un determinat sistema i alertar sobre comportaments que es detectin fora d'un patró reconegut com correcte. En aquest sentit, podríem utilitzar Falco per observar (en alguna de les seves sortides configurades: pantalla, *journal*, un fitxer de registre personalitzat, etc) els seus logs generats en temps real per comprovar la integritat de fitxers i carpetes (i notar canvis sospitosos), per detectar "rootkits" i la propagació de software malició (malware), per adonar-se de canvis fora de l'horari habitual en el funcionament del serveis, per detectar l'alteració de fitxers sensibles com els de contrasenyes (o detectar-hi atacs per força bruta), per monitoritzar escanejos als ports oberts, etc, etc.

Però ¿què passa si tenim diversos sistemes a monitoritzar? No podem estar enfront de cadascun a la vegada per observar els missatges que Falco generen en local. A més, instal.lar el Falco a cada sistema monitoritzat tampoc és una solució gaire elegant: és molt més òptim tenir instal.lat només un recolector de dades a cada màquina monitoritzada i llavors implementar algun sistema d'enviament per xarxa de la informació per fer-la arribar a un únic Falco centralitzat que recollís, reconegués i filtrés ell sol els events dels diferents sistemes. D'aquesta manera, tot es tindria molt més endreçat, amb un únic sistema que recolliria els events dels sistemes remots i els aniria repassant en temps real per tal de generar, si s'escaiés, les alertes corresponents.

Seguint amb aquesta idea, podríem fins i tot emmagatzemar en disc la informació recollida dels diferents sistemes monitoritzats per tenir-ne un històric centralitzat i, per què no, visualitzar-la en forma de gràfiques estadístiques per fer-ne un estudi més exhaustiu i a la vegada còmode de totes aquestes dades.

Per implementar aquesta sol.lució integral de *recollida i enviament de dades → filtratge → emmagatzematge → visualització* , més enllà de Falco, que seria només un cas concret, es poden fer servir diferents eines genèriques. Les que estudiarem a continuació són de les més utilitzades a l'hora de recollir" logs" (del tipus que siguin) de diferents fonts i emmagatzemar-los de forma centralitzada per tal de accedir a ells gràfica i còmodament des d'un navegador, que és just el que volem. Concretament, els diferents components que estudiarem, instal·lables de forma independent però que normalment treballen junts (de fet, dels quatre components, tres formen part del mateix projecte, https://opensearch.org), són:

*"**FluentBit**" (https://fluentbit.io) : Aquest programa fa dues coses:

1) "Absorbeix" les línies que es vagin afegint al final de fitxer/s de **logs** local/s (prèviament indicats) o directament del "Journal" del sistema (o provinents d'una altra font qualsevol prèviament configurada...en general, la quantitat d'orígens possibles és força diversa i variada; de fet, FluentBit no només pot absorbir línies de cadenes sinó que també és capaç, tal com aviat estudiarem, de recollir dades numèriques que es corresponen a **mètriques** de funcionament del sistema, com ara la temperatura de la CPU, l'espai lliure de disc, l'espai lliure de memòria, la quantitat de connexions establertes, etc)

2) ...i les reenvia -per la xarxa- a un destí central que pot ser o bé un servidor DataPrepper o bé directament un servidor OpenSearch (o un altre programa similar compatible) per a què allà es processin i, eventualment, s'emmagatzemin

NOTA: Altres softwares similars en funcionalitat que podem destacar (tot i que n'hi ha molts més) són:
*__Fluentd__ (https://www.fluentd.org) -Desenvolupat pel mateix equip que FluentBit, és un software més complex i pesat; està més enfocat a fer d'agregador a gran escala de dades que FluentBit
*__LogStash__ (https://www.elastic.co/logstash)
*__Vector__ (https://vector.dev)

*"**OpenSearch DataPrepper**" (https://opensearch.org/docs/latest/data-prepper/index) : Aquest programa -opcional!- s'utilitza per rebre de forma centralitzada totes les línies/mètriques absorbides per "Fluentbit" (o per qualsevol altre "recolector" que reemeti les dades rebudes mitjançant HTTP) i així fer dues coses més en un únic punt (el que se'n diu "agregació" dels logs) de forma controlada:

1) "Parsejar" i/o filtrar i/o enriquir determinats logs segons diferents criteris que s'hagin especificat i ...

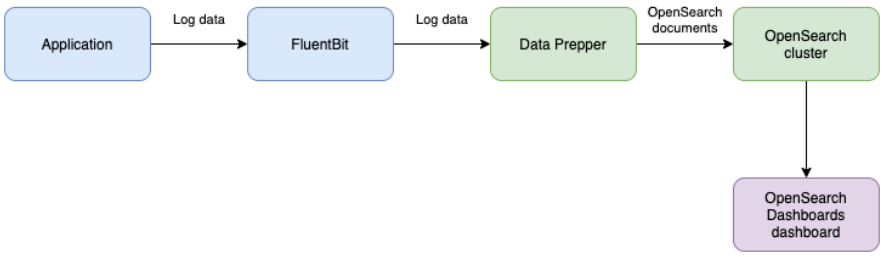
2) ...enviar -per la xarxa o en local- els logs ja "netejats", "enriquits" i filtrats a un servidor OpenSearch (o a un fitxer o a la sortida estàndard, etc)

NOTA: Aquest programa diem que és opcional perquè és perfectament possible fer que "FluentBit" enviï els logs directament a OpenSearch sense passar per DataPrepper. A més, "FluentBit" disposa de prou mecanismes per "netejar"/"enriquir"/"filtrar" les dades abans d'enviar-les al destí, així que en la majoria de vegades no caldria interposar l'intermediari DataPrepper. Només en casos de processament de dades molt concrets, d'agregació de dades provinents de diversos orígens o quan sigui necessària tenir prou capacitat de "back-pressure" en cas d'una ingesta massiva de dades puntual pot ser necessari la implementació d'un servidor DataPrepper.

*"**OpenSearch**" (https://opensearch.org/docs/latest/about/): Aquest programa és el responsable d'emmagatzemar en disc les dades recollides de "DataPrepper" (o obtingudes de "FluentBit" directament). Està optimitzat per gestionar cadenes (la majoria de logs són d'aquest tipus) i la seva especialitat és la de realitzar recerques (d'aquí el seu nom) perquè indexa tot el contingut que guarda i, a partir d'aquí, també fer-ne agregacions (és a dir, càlculs estadístics agrupats, com ara mitjanes, màxims/mínims, comptadors, etc)

*"**OpenSearch Dashboards**" (https://opensearch.org/docs/latest/dashboards/index/) : Aquest programa és un client d'OpenSearch que permet fer-li consultes sobre les dades que tingui emmagatzemades i, a la vegada, un servidor web que permet mostrar a l'usuari diferents perspectives d'aquestes mitjançant un panel de control molt gràfic i visual.

El següent diagrama il.lustra l'acabat d'explicar: "FluentBit" recull les dades, "DataPrepper" les processa ("parseja", neteja, enriqueix, filtra, agrega...és a dir, transforma), "OpenSearch" les indexa i emmagatzema i "Dashboards" les visualitza i interacciona amb l'usuari.



NOTA: Al diagrama també sindica, per colors, la possible ubicació de cada component: en color blau estan marcats els que haurien d'estar instal·lats sobre la màquina (o màquines!) de la/les qual/s se'n volen extreure les dades d'interès; en color verd estan marcats els que haurien d'estar instal·lats en una màquina única central i en color morat està marcat el que podria estar instal·lat en una màquina a banda.