Getting Started - Install EdgeLake (with Open Horizon) to manage KubeArmor events

Overview

KubeArmor analyzes telemetry data to understand application behavior for container/node forensics.
With thousands of nodes deployed (using Open Horizon), sending events streams to a centralized node is not a viable option.

This document details an installation of EdgeLake instances to manage the KubeArmor's event data to extract real time insight from the data, enable real-time alerts and monitoring and service the data to analysis and AI applications, all of that without cloud contracts and costs.

An overview of the deployment is with the following link: EdgeLake - KubeArmor Integration
Notes:

1. EdgeLake is the Open-Source release of the AnyLog Edge Software Platform.
2. Below are the links to the technical documentation of the platform.
3. The EdgeLake-KubeArmor setup detailed in this doc is using the EdgeLake code base regardless

if the documentation or scripts reference AnyLog.

AnyLog Documentation

The documents listed below provide basic training on AnyLog as an Edge Platform. They review the basic concepts, usage of the CLI and AnyLog commands.

- Install Presentation
- Product Documentation
- Getting Started Document
- Prerequisites
- Training Documentation
    - Session I
    - Session II
    - Fast Deployment (Cheatsheet)

Deployment setup

The setup requires the following deployments:

- KubeArmor instances to monitor events on pods, containers, and virtual machines.
- An EdgeLake Network with the following components:
- One or more **Operator Nodes -** These nodes host the KubeArmor events data.
- One or more **Query Nodes** - These nodes service the KubeArmor data to applications that need the data.
- One **Master Node** - The Master Node hosts the shared metadata. The shared metadata facilitates the EdgeLake operations.

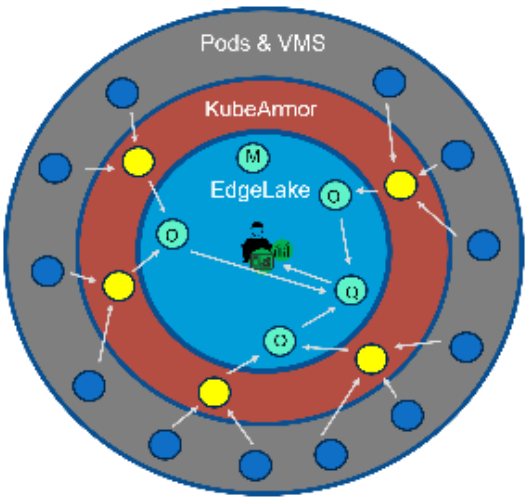
Note: Data is transferred between KubeArmor and an EdgeLake Node using a gRPC connector. Details on the EdgeLake gRPC connector are available here

In the diagram below, the gray (outer) circle represents the Pods and VMs that are monitored.
KuberArmor is in the middle (brown) circle, monitors the Pods and VMs and generates event logs that are pulled (using gRPC connector) by EdgeLake instances (in the innermost, blue, circle).
The EdgeLake nodes are of 3 types:

- 1 Master Node (M) - hosts the shared metadata.
- 1 Query Node (Q) - Interacts with the applications to service the data.
- 3 Operator Node (O) - host the KubeArmor data.

Queries are processed by issuing a query to the Query Node, the Query Node is using the shared Metadata to determines which are the target Operators that host the data. It transfers the query to the target Operators, the replies from all target Operators are aggregated and returned as a unified reply to the application.

This setup hosts the KubeArmor data at the edge and satisfies queries without centralizing the data.


Pods & VMS
KubeArmor
EdgeLake