1 Objetivos

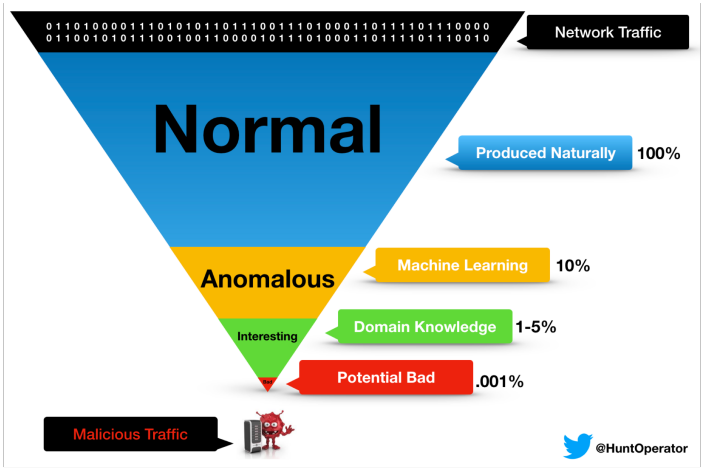
- Aplicar los conocimientos sobre threat hunting, data science y dominio experto para la detección de dominios maliciosos en el tráfico de red.

2 Preámbulo

¿Qué es Threat Hunting?

Austin Taylor lo define como "*el proceso de buscar en la red, de forma proactiva e iterativa, amenazas avanzadas que evaden las soluciones de seguridad existentes, con el fin de detectarlas y aislarlas*."

En el siguiente flujo, Taylor muestra como solo una muy pequeña porción del tráfico es realmente maliciosa, y las técnicas que deben seguirse para identificar dicho tráfico:



- Network Traffic: todos el tráfico de la red.
- Produced naturally: tráfico natural, producto por usuarios y dispositivos.
- Machine Learning: los clasificadores basados en ML son capaces de filtrar el 90% del tráfico no malicioso.
- Domain Knowledge: aquí entra el conocimiento experto de un profesional. Es capaz de separar tráfico que no pasa ciertas pruebas. Por ejemplo: ¿Un dominio Web se encuentra en la lista Alexa 1 millón o en la lista de Cisco Umbrella? ¿Hace cuánto se registró un dominio? Un dominio malicioso generalmente tiene un período corto de registro o no tiene registro directamente.
- Potential bad: en base a los filtros del dominio experto se analiza el tráfico restante para confirmar o no si es sospechoso.

La búsqueda y caza de amenazas en el tráfico de red es una tarea desafiante: del tráfico total solamente un segmento muy pequeño es realmente malicioso, por ejemplo, algunas conexiones de una computadora infectada tratando de conectarse con un servidor remoto en espera de instrucciones (C2), a través de dominios DGA, entre miles de conexiones benignas.

Domain Generation Algorithm son algoritmos que generan dominios pseudo aleatorios, y son usados para comunicar computadoras infectadas con un servidor de comandos. Los dominios se generan con la misma semilla de inicialización en las víctimas y en el servidor, es una técnica moderna a diferencia de malware tradicional donde los dominios están "hardcoded" dentro del código fuente, y con ingeniería inversa es posible obtener dicha lista de dominios y darlos de baja.

La idea del Threat Hunting es aplicar el conocimiento sobre tráfico de red para depurar la cantidad de tráfico a analizar, y encontrar las conexiones realmente maliciosas.

3 Desarrollo

El laboratorio consiste en analizar un archivo con registros de red, con más del 99% de registros como benignos, y encontrar y detectar conexiones hacia dominios DGA.

Se deberá desarrollar de forma individual, y se debe entregar un repositorio con el jupyter notebook con el desarrollo.

Parte 1 – Filtrado y preprocesamiento

Para este ejercicio se utilizará el archivo large_eve.json que se encuentra en Canvas, en el módulo de la semana. Este archivo contiene miles de paquetes capturados mediante el IDS Suricata. Pasos:

1. Cargue la información del archivo large_even.json en una lista, muestre la cantidad de registros total (deben ser 746, 909). Este es nuestro tráfico inicial!
2. Debido a que estamos buscando dominios web, del total de registros, solamente estamos interesados en los registros DNS. Cargue únicamente aquellos registros que sean DNS.
3. Muestre la nueva cantidad de registros filtrados. Deben ser 21484. Esta es una cantidad mucho más manejable, pero aún se debe seguir depurando la información a buscar.
4. Muestre la información de 2 registros cualesquiera.
5. Debido a que la data consiste en estructuras JSON anidadas, utilice la característica json_normalize para normalizar la información y asignarla en un dataframe. Muestre el shape del dataframe, debería obtener (21484, 163).
6. Como estamos buscando dominios DGA, debemos filtrar los registros DNS para aquellos registros tipo A (son aquellos que mantienen una dirección IP asociada a un dominio). Después de filtrar debería obtener 2849 registros.

7. Filtre los dominios únicos. Debe obtener 177 registros únicos
8. Escriba una función que obtenga el TLD para un dominio. Por ejemplo, para api.wunderground.com el TLD es wunderground.com, para safebrowsing.clients.google.com.home, el TLD es home. Utilice un LLM para ayudarle a escribir esta función, verifique que obtiene correctamente el TLD, incluya el prompt utilizado en su notebook.
9. Del dataframe de dominios únicos de tipo A, obtenga el TLD (top level domain) utilizando la función anterior para crear una columna nueva llamada domain_tld, y elimine todas las demás columnas.

Parte 2 – Data Science

10. Utilice Gemini para clasificar los dominios como DGA (1) o legítimos (0).
11. Filtre aquellos considerados como DGA (valor 1) y muéstrelos. Recuerde que los resultados de los modelos pueden incluir falsos positivos y falsos negativos, por lo que no podemos fiarnos por completo de esta clasificación y debemos seguir indagando. Después de eliminar duplicados, debe obtener 61 registros únicos.

Parte 3 – Dominio experto

12. Ahora ya tenemos un listado de dominios reducido y considerado como sospechoso, por lo que debemos aplicar dominio experto para encontrar los verdaderos registros maliciosos. Escriba una función que utilice la lista de un millón de TLD proporcionada en Canvas, y devuelva 0 si el TLD se encuentra en la lista y 1 si no está. Utilice un LLM para crear dicha función, verifique que no se carga la lista cada vez que se busca un TLD. Incluya el prompt en su notebook.
13. Utilice la función para determinar si los TLD se encuentran en dicha lista. Filtre aquellos que si se encuentran. Después de eliminar duplicados, debería obtener 13 dominios.
14. Finalmente, para confirmar los dominios maliciosos podemos buscar la fecha de creación del TLD. Cree una función qué en base al TLD, devuelva la fecha de creación de este. Utilice un LLM para escribir dicha función, incluya el prompt utilizado en su notebook.
15. Muestre la fecha de creación para cada uno de los 13 dominios finales ¿Cuáles son los dominios que podemos confirmar como sospechosos?
16. Recuerde que los dominios DGA son conocidos por formarse de forma aleatoria: secuencias aleatorias de caracteres, no palabras. Indique que dominios sospechosos tienen este patrón y que pueden confirmarse como dominios DGA.