Metodología de Etiquetado Manual para la Construcción de un Sistema de Detección de Anomalías para una Aplicación Web

Introducción

En el ámbito de la detección de ataques/anomalías y análisis de seguridad en aplicaciones web, la disponibilidad de datos etiquetados es un requisito fundamental para el desarrollo y validación de modelos. Sin embargo, en entornos reales, los logs de acceso web no suelen incluir etiquetas que indiquen si una solicitud es legítima o maliciosa. Por ello, se implementó una metodología de **Etiquetado Manual Asistido** que combina la inspección individual de registros con técnicas de agrupación por patrones, permitiendo generar un conjunto de datos anotado para el entrenamiento de modelos de clasificación.

Este capítulo describe la estrategia de etiquetado llevada a cabo sobre los logs de acceso preprocesados, detallando los enfoques utilizados, los criterios de clasificación y las técnicas implementadas para optimizar el proceso ante un volumen significativo de datos.

Marco Conceptual del Etiquetado

El etiquetado manual se realizó sobre un conjunto estructurado de logs de Apache en formato combinado (Combined Log Format), previamente parseados. Inicialmente, todos los registros se marcaron con el valor son

etiquetados con un valor de -1, indicando un estado de **no-etiquetado**. El objetivo fue reemplazar este valor por:

- 0 : indica que el registro corresponde a una solicitud **normal**
- 1 : indica que el registro presenta características de **anomalía**

Estrategia de Etiquetado Implementadas

Etiquetado Individual

Se desarrolló una función interactiva que permitía al analista revisar registro por registro, mostrando campos clave como:

- Dirección IP del cliente
- Código de estado HTTP
- Tamaño de la respuesta
- Agente de usuario (User-Agent)
- Método HTTP
- URL solicitada
- Protocolo utilizado

El analista podía asignar una etiqueta (0 o 1) o finalizar el proceso. Si bien esta aproximación permite un análisis detallado, resultaba poco escalable para grandes volúmenes de datos, por lo que se utilizó principalmente para familiarizarse con la naturaleza de los registros.

Etiquetado por Lotes basado en Patrones Comunes

Para agilizar el proceso, se diseñó una estrategia basada en la agrupación de registros con características similares, asumiendo que solicitudes idénticas en ciertos atributos pueden clasificarse de manera uniforme.

Agrupación por N-Tuplas de Atributos: Se generaron n-tuplas a partir de la combinación de atributos seleccionados, inicialmente considerando:

- Método HTTP
- URL solicitada

- Protocolo
- Código de estado

Posteriormente, se optimizó la agrupación utilizando solo método y URL, dado que estas dos características capturan de manera más efectiva los patrones de solicitud recurrentes.

Proceso de Etiquetado por Tuplas: Para cada tupla única:

1. Se calcula su frecuencia en el conjunto no etiquetado.
2. Se muestran al analista:
   - La frecuencia de la tupla
   - Los valores únicos más comunes de otros atributos relevantes (como User-Agent o protocolo)
3. El analista asigna una etiqueta para todos los registros pertenecientes a esa tupla, permitiendo etiquetar cientos de registros con una sola decisión.

Orden de Revisión basado en Frecuencia: Se implementó la opción de revisar las tuplas en orden inverso de frecuencia (de menos a más frecuentes), bajo la hipótesis de que las anomalías suelen ser menos comunes que el tráfico normal, lo que permite priorizar la revisión de patrones potencialmente sospechosos.

Integración de Resultados de Análisis Automático

El proceso de etiquetado manual se benefició de una etapa previa de análisis automático realizada sobre el dataset completo. Este análisis, ejecutado antes de iniciar la revisión manual, identificó y etiquetó automáticamente registros con patrones claramente maliciosos, sirviendo como un filtro inicial y reduciendo el volumen de datos a revisar manualmente.

Identificación Automática de Logs Problemáticos: Se utilizó la biblioteca `lars` para detectar líneas que no seguían el formato estándar de Apache. Registros con cadenas en hexadecimal o ausencia de User-Agent se marcaron automáticamente como anómalos (etiqueta 1)

Detección Automática de Métodos HTTP Inusuales: Métodos distintos a GET y POST (como DELETE, TRACE, REQMOD) se consideraron sospechosos, se analizaron detenidamente y se etiquetaron automáticamente como anomalías

Análisis Automático de Códigos de Estado HTTP: Códigos como `400` (Bad Request), `408` (Request Timeout) y `499` (Client Closed Request) se analizaron en contexto y aquellos provenientes de un conjunto limitado de IPs y con agentes de usuario vacíos se clasificaron automáticamente como actividad maliciosa automatizada

Esta capa de análisis automático preexistente permitió que el etiquetado manual se enfocara en casos menos evidentes y en la validación de los criterios automáticos, optimizando así el tiempo del analista

Resultados del Proceso de Etiquetado

Tras aplicar las estrategias descritas, se obtuvo un conjunto de datos con tres categorías:

- *No-Etiquetados* (-1): Registros que no fueron revisados (40%)
- *Normales* (0): Solicitudes legítimas (45%)
- *Anómalos* (1): Solicitudes con características de ataque o comportamiento sospechoso (15%)

La distribución resultante permitió avanzar hacia una fase de modelado supervisado, con un conjunto de datos etiquetado que refleja tanto el conocimiento del dominio como los patrones identificados automáticamente.