Extracción de Características de la URL para un Modelo de Detección de Anomalías

Introducción

La URL (Uniform Resource Locator) constituye una de las fuentes más ricas de información para la detección de comportamientos maliciosos en tráfico web. Dado que muchos ataques, como inyecciones SQL, Cross-Site Scripting (XSS) y path traversal, se manifiestan directamente en la estructura y el contenido de las URLs solicitadas, la extracción sistemática de características a partir de este campo es fundamental para entrenar modelos de clasificación efectivos.

Se diseñó e implementó un conjunto de 26 funciones de extracción de características que transforman la URL cruda en un vector numérico interpretable para el entrenamiento de modelos de ML. Estas características se agrupan en tres categorías principales: **Contenido Semántico**, **Estructura Sintáctica** y **Indicadores de Ofuscación**.

Características basadas en Contenido y Keywords

Estas características buscan identificar la presencia de términos, patrones o cadenas comúnmente asociados a técnicas de ataque conocidas. Se implementaron mediante expresiones regulares insensibles a mayúsculas/minúsculas para capturar variantes.

- *Detección de Intentos de Inyección SQL*: Se cuantifica la aparición de palabras reservadas de SQL y comandos de manipulación de bases de datos:
  - **Términos**: `SELECT`, `FROM`, `WHERE`, `DELETE`, `DROP`, `CREATE`, `TABLE`, `UNION`, `INSERT`, `UPDATE`, `ALTER`, entre otras.
- *Identificación de Patrones de Cross-Site Scripting (XSS)*: Se busca terminología asociada a la ejecución de scripts maliciosos en el navegador de la víctima. Esta función indica posibles intentos de inyectar código malicioso
  - **Términos**: `script`, `alert`, `javascript`, `onerror`, `onload`, `eval`, `iframe`, `cookie`.
- *Detección de Ejecución de Comandos del Sistema*: Se identifican términos relacionados con la interacción con el sistema operativo subyacente. Esta función permite revelar intentos de ejecutar comandos arbitrarios en el servidor (Command Injection).
  - **Términos**: `cmd`, `shell`, `exec`, `/bin/`, `/etc/`, `/tmp/`, extensiones como `.php`, `.exe`.
- *Búsqueda de Access Logs a Recursos Sensibles*: Se analiza la presencia de términos asociados a paneles de administración, credenciales o archivos críticos:
  - **Términos**: `.env`, `.env/config`, `admin`, `password`, `login`, `credential`, `secret`, `token`, `config`.
- *Identificación de Patrones de Error y Depuración*: Se buscan términos que podrían explotarse para obtener información de diagnóstico:
  - **Términos**: `error`, `debug`, `exception`, `trace`, `stack`, `warning`.
- *Detección de Terminología Asociada a Malware*: Se rastrean palabras vinculadas a amenazas cibernéticas conocidas:
  - **Términos**: `malware`, `ransomware`, `phishing`, `exploit`, `virus`, `trojan`, `backdoor`.

Características Estructurales y Sintácticas

Estas características cuantifican propiedades formales de la URL, como su longitud, composición de caracteres y uso de símbolos especiales, que pueden desviarse de lo esperado en tráfico legítimo.

- *Longitud y Composición de Caracteres*: Longitud total de la URL, conteo de dígitos y letras, y cantidad de caracteres especiales (símbolos como `'`, `"`, `;`, `--`, `<`, `>`, `&`, `|`, `\\` son típicos en payloads de inyección)
- *Frecuencia de Símbolos Específicos*: como `.`, `%`, `?`, `=`, `-`, `_`. Estas pueden ser interpretables como: un número elevado de `.` puede sugerir intentos de path traversal o dominios sospechosos; `?` y `=` reflejan la presencia de parámetros de consulta; `-` y `_` pueden ser parte de nombramientos inusuales o ofuscación
- *Proporción de Caracteres Inusuales*: Se calcula la razón entre caracteres no alfanuméricos (excluyendo guiones, puntos y guiones bajos) y la longitud total. Un valor alto sugiere una URL con posible contenido ofuscado o malicioso.

> **Ofuscación**: técnica mediante la cual se modifica intencionalmente el código, los datos o las cadenas de texto para dificultar su comprensión, análisis o detección, manteniendo intacta su funcionalidad original. En seguridad informática, es un método utilizado por atacantes para evadir sistemas de protección como firewalls, sistemas de detección de intrusiones (IDS) y filtros de aplicaciones web
> **Ejemplos**:
>
> - *Codificación de Caracteres*: Codificación URL (reemplazar caracteres especiales por `%` seguido de su valor hexadecimal)
> - *Ofuscación de Comandos SQL*: Romper comandos SQL en partes concatenadas

Características de Ofuscación y Codificación

Los atacantes suelen ofuscar sus payloads para evitar la detección. Estas características buscan identificar técnicas comunes de enmascaramiento.

- *Detección de Técnicas de Codificación*: Identificación de patrones como `\\xXX` o `%XX` (codificación hexadecimal)
- *Funciones de Evaluación de Código*: Detectar intentos de ejecutar código dinámicamente generado. Se buscan palabras en la URL como: `eval(`, `chr(`, `char(`, `fromCharCode`, `exec(`, `popen(`.
- *Referencias a Directorios Sensibles y Path Traversal*: Identificar intentos de acceder a archivos del sistema o navegar fuera del directorio web permitido. Rutas comunes: `/etc/`, `/bin/`, `/tmp/`, `/root/`, `/home/`, `../`, `..\\`.

Integración en el Conjunto de Datos y Justificación de las Funciones de Extracción

Estas características, aplicadas a la URL, junto con las demás variables del log (como código de estado, método HTTP, tamaño de respuesta), forman un espacio de características multidimensional que captura tanto el contexto de la solicitud como la naturaleza potencialmente maliciosa de su contenido.

ADL-WAF es uno de los trabajos que utiliza algunas de estas funciones para crear un sistema de detección de anomalías. Las funciones que este modelo emplea son:

- **Alphanumeric Character Ratio**: Característica que cuenta la proporción de caracteres alfanuméricos en una carga útil. Las solicitudes regulares tienen una mayor proporción de caracteres numéricos y alfabéticos en comparación con los símbolos especiales. Como resultado, esta característica tiende a tener más valor en solicitudes normales que en las de un ataque
- **Badwords Ratio**: Representa la relación entre el número de malas palabras (términos comúnmente utilizados como partes de consultas de ataque) y la longitud de los caracteres alfanuméricos. Las consultas normales no tienen malas palabras, pero las solicitudes anómalas suelen tener una proporción mayor
- **Special Character Ratio**: Característica que cuantifica la proporción de caracteres especiales (no alfanuméricos) con respecto a la longitud total de entrada. A menudo, las solicitudes anómalas tienen más caracteres

especiales, como símbolos, numéricos y alfabéticos. Por tanto, esta propiedad tendrá un valor superior en solicitudes anómalas a lo normal

- **Illegal Special Character Ratio**: Característica que representa la proporción de caracteres especiales ilegales con respecto al número total de caracteres especiales en la carga útil. Las solicitudes normales tienen una proporción baja o ausencia de caracteres especiales ilegales, mientras que las solicitudes anómalas exhiben una proporción más alta