UNIVERSIDAD DE LA CIUDAD DE AGUASCALIENTES

MAESTRÍA EN CIENCIA DE DATOS

GESTIÓN DE PROYECTOS DE CIENCIA DE DATOS

“Detección de anomalías de tráfico en servidores web”

Alumno:

E23S-18014: MITSIU ALEJANDRO CARREÑO SARABIA

Periodo Enero 2024 - Junio 2024, Aguascalientes, Ags

Resumen.

[Resumen conciso del proyecto que incluya los objetivos principales, el alcance y la importancia del estudio.]

Contenido

Introducción

[Contextualización del problema o la pregunta de investigación que se aborda en el proyecto.]

1. Propuesta Científica

Se propone desarrollar una solución integral de monitoreo de tráfico en servidores web así como la detección automatizada de tráfico anómalo mediante la implementación de técnicas de análisis topológico así como aprendizaje automático las cuales en conjunto permitan por una parte el constante monitoreo de los usos y la toma de decisiones preventivas y correctivas.

Aplicando estas metodologías, es posible evaluar nuevas peticiones basado en el tráfico histórico del servidor y obtener un índice de similitud respecto a solicitudes pasadas, con ello es posible detectar anomalías o contenido malicioso y tomar tanto acciones correctivas (protección ante uso anómalo) como preventivas (detectar picos o valles de tráfico y ajustar la infraestructura acorde).

1.1 Antecedentes

Con la expansión del acceso a servicios de internet, así como la creciente disponibilidad de dispositivos de distintas categorías para conectarse a la red, la demanda y tráfico de servicios web se encuentra en constante aumento. Mucho se ha desarrollado en términos de escalabilidad de infraestructura así como adopción de soluciones distribuidas para dar servicio a la creciente demanda.

La cantidad de tráfico que recibe un servidor web con acceso a internet es inmenso, el cuál recibe peticiones de usuarios reales, peticiones de bots, peticiones automatizadas y peticiones de usuarios con intenciones maliciosas, solo analizar la cantidad de información generada de manera manual es una tarea imposible.

El tráfico de servidores web tiene claras tendencias como recursos solicitados, región geográfica de donde se solicita, hora en que se solicitó, cantidad de bytes enviados, por lo que identificar las tendencias y detectar las anomalías es un trabajo que puede ser automatizado.

Dada la tendencia de alojar datos sensibles aumenta la importancia a evaluar qué y cómo se están accediendo a los recursos solicitados, así como desarrollar herramientas que faciliten filtrar las anomalías para tomar acciones correctivas.

1.2 Objetivos del Proyecto

El objetivo del sistema es desarrollar una solución integral de monitoreo y detección de tráfico anómalo mediante la implementación de técnicas de aprendizaje automático para decidir las acciones preventivas y/o correctivas necesarias.

Para ello es necesario realizar múltiples acciones de soporte como:

- Analizar las técnicas y procesos tanto tradicionales como de aprendizaje automático mediante los cuales se analiza tráfico web actualmente
- Enumerar las características y casos de uso de sistemas de monitoreo y alerta efectivos

- Desarrollar un sistema detección de anomalías basado en aprendizaje automático
- Evaluar el sistema desarrollado implementandolo en un entorno controlado

1.3 Preguntas de investigación

¿De qué manera se analiza el tráfico web actualmente?

¿Qué elementos debe tener un sistema de alertas para ser útil (falsos negativos/falsos positivos, canales de comunicación, protocolos extras)?

¿Actualmente cómo se ha implementado el aprendizaje automático en análisis de tráfico web?

1.4 Justificación

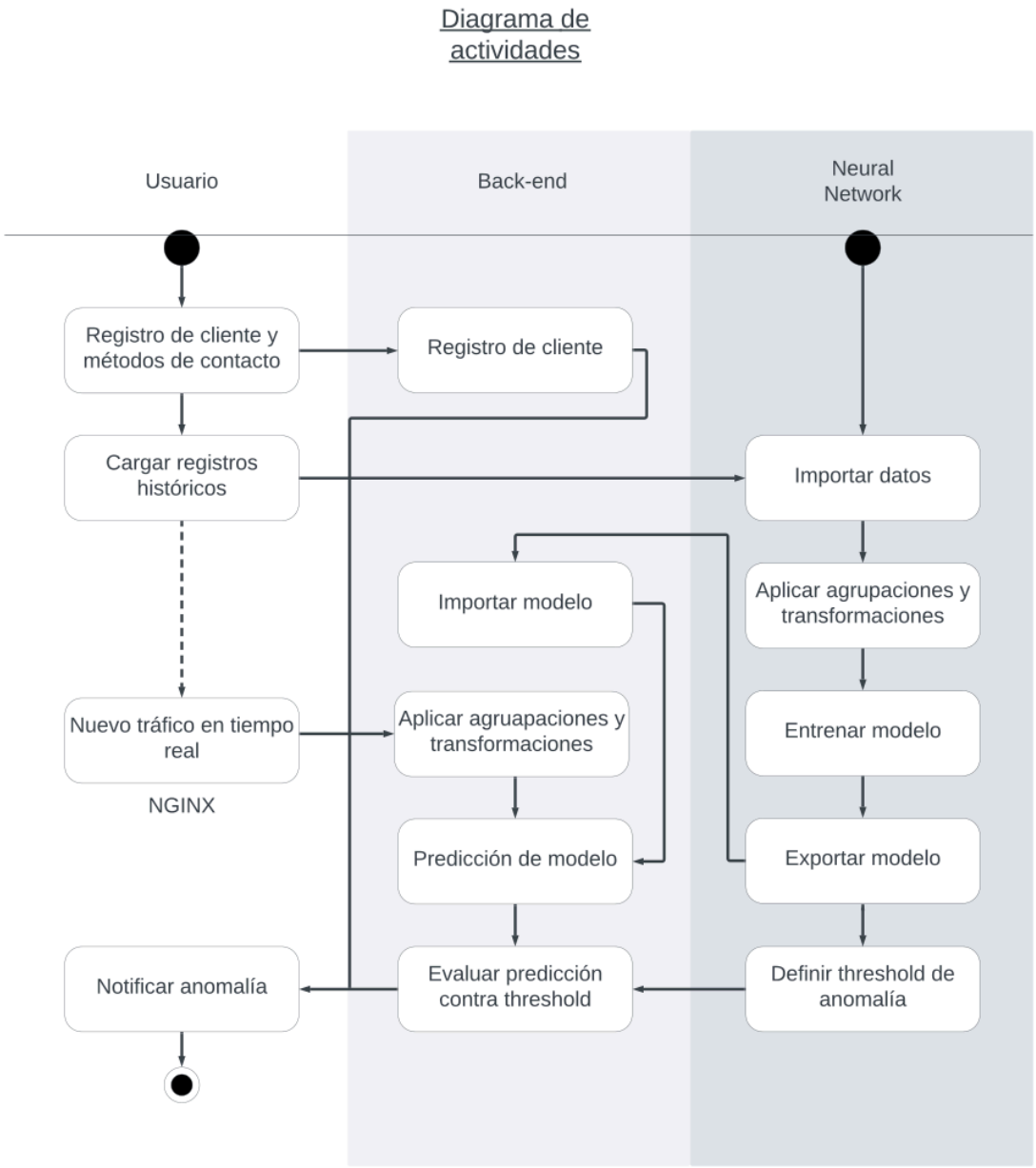
El tráfico a un servidor web provee datos confiables sobre la información y el contexto bajo el que se usan sus recursos, pero la cantidad de información generada es tan grande que un análisis manual no es viable. Entender los usos típicos y diferenciarlos de los atípicos es una herramienta poderosa que aplicada en tiempo real permitirá mejorar la calidad, y resguardo de la información contenida.

Analizar los registros de tráfico web permite no solo entender la manera en que se consume la información que contiene un servidor, sino también detectar si el uso generalizado se transforma, o si existen anomalías que pueden ser malintencionadas. Dado el volumen de información que se genera, y la creciente sensibilidad de los datos alojados, aplicar herramientas de aprendizaje automático permitirá agilizar y perfeccionar cualquier proceso manual.

1.5 Viabilidad

Para la realización del proyecto se cuenta con acceso a fuentes de información necesarias para realizar el análisis y desarrollo correspondiente con datos confiables y en cantidad suficiente para simular condiciones reales.

Respecto a la estructura e infraestructura se realizó el siguiente diagrama en los que se establecen las distintas entidades involucradas en el sistema así como su interrelación, dependencia y relevancia. Se considera que se cuenta con la suficiente experiencia técnica para construir el sistema de manera exitosa.



Por otra parte es necesario realizar una inversión inicial para rentar los equipos de cómputo adecuados para realizar el entrenamiento, transformación y manejo de datos adecuado, por lo que es necesario considerarlo tanto en las características de los equipos necesarios como en su financiamiento.

2. Propuesta Financiera

[Presenta una evaluación de los recursos necesarios para llevar a cabo el proyecto de manera eficiente y efectiva]

2.1 Presupuesto

[Detalle de los costos estimados para llevar a cabo el proyecto, incluye costo estimado de recursos humanos, recursos materiales, y de otros recursos requeridos (licencias, permisos, etc.)]

2.2 Justificación Económica

[Explicación detallada de cómo se utilizarán los fondos solicitados y cómo contribuirán al éxito del proyecto]

2.3 Fuentes de Financiamiento

[Identificar las posibles fuentes y explicar por qué son adecuadas para el proyecto, pueden incluir fondos privados o públicos]

3. Propuesta de Gestión del Proyecto

[La propuesta de gestión del proyecto establece un marco organizativo y operativo para garantizar la ejecución exitosa y el cumplimiento de los objetivos establecidos.]

3.1 Equipo de Trabajo y estructura organizativa

[Descripción de los miembros del equipo, sus roles y responsabilidades. Detalle de la estructura de gestión del proyecto, incluyendo roles, comunicación y toma de decisiones]

3.2 Plan de trabajo

[Establece de manera detallada las actividades, tareas, recursos y plazos necesarios para alcanzar los objetivos del proyecto de manera eficiente y efectiva.]

3.3 Riesgos y Mitigación

[Identificación de los posibles riesgos del proyecto y estrategias para mitigarlos.]

3.4 Plan de Comunicación

[Recomendaciones para la comunicación y cooperación entre los tres equipos de trabajo del proyecto, incluyendo reuniones regulares, informes de progreso, etc.]

3.5 Ética y Cumplimiento

[Consideraciones éticas y legales relevantes para el proyecto, incluyendo la protección de datos y la conformidad con regulaciones aplicables.]

4. Siguientes Pasos

[Detalle de las actividades que permitirán comenzar la implementación del proyecto, una vez que este ha sido autorizado]

5. Conclusiones

6. Referencias