**Estrategia para Escalar el Desarrollo de una Plataforma de Modelado de Amenazas Asistida por IA.**

Para definir esta esta estrategia se tuvo en cuenta varios factores los cuales , son base para todo ciclo de vida de desarrollo de app , software etc. Metodologías DevSecOps la cual permite tener controles en las diferentes fases del ciclo. Sobre el modelado de amenazas hay varias formas de evaluarlo, las metodologías mas conocidas y aplicadas en el mercado con los temas de STRIDE, DREAD y PASTA, los cuales van a estar en cada uno de los puntos de la estrategia sugerida. Con el apoyo de la IA, hoy día se automatiza mucho trabajo procesos que ayudan a obtener información más rápido, sin embargo, es un modelo que toca entrenar y pulir para evitar las llamadas alucinaciones, en el desarrollo de la estrategia se tuvo en cuenta lo siguiente:

**1. Validación de Inputs y Estándares de Documentación**

Dada la falta de documentación estandarizada en muchos entornos, se implementará un módulo de validación de entradas que verificará que los datos ingresados cumplan con un formato coherente y sean relevantes para el análisis.

Los inputs clave que se considerarán incluyen:

* **Descripción del caso**
* **arquitectura del sistema.**
* **activos críticos.**
* **actores involucrados**
* **controles de seguridad.**

Se aplicarán reglas específicas y modelos de aprendizaje automático para identificar inconsistencias.

Si se detecta una anomalía, se rechazará la entrada y se notificará al usuario con una explicación clara. Esta validación garantiza que solo se procesen datos fiables y relevantes, lo que mejora la precisión del análisis de amenazas.

**2. Incorporación del Feedback de Usuarios**

Se crea un sistema de retroalimentación continuo, donde los analistas de seguridad puedan evaluar la calidad de los resultados. Si un resultado parece incorrecto, los usuarios podrán marcarlo y dejar un comentario. Se aplicará procesamiento de lenguaje natural (NLP) para categorizar y priorizar los comentarios, permitiendo ajustes iterativos en los algoritmos de IA.

**3. Clasificación de Amenazas mediante STRIDE**

La clasificación de amenazas se basará en el modelo **STRIDE** (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege).

**Por ejemplo**: si el sistema detecta un intento de ingreso sospechoso, lo marcará como una amenaza de suplantación (Spoofing) y lo mostrará en un panel de control sencillo.

**4. Gestión y Priorización de Hallazgos con DREAD**

Los hallazgos serán evaluados según la metodología **DREAD** (Damage, Reproducibility, Exploitability, Affected Users, Discoverability). Se asignará una puntuación del 1 al 10 en cada categoría y se calculará un promedio. Por ejemplo, una vulnerabilidad que puede explotarse fácilmente y afectar a muchos usuarios tendrá una puntuación alta y será tratada con prioridad. Los resultados se integrarán en paneles interactivos que usarán colores para indicar riesgos: verde para bajo, amarillo para medio y rojo para alto.

**5. Simulación y Análisis de Amenazas con PASTA**

Se aplicará el marco PASTA para simular posibles ataques y su impacto en los procesos de negocio. Por ejemplo, la IA puede simular un ataque de phishing para ver cómo afecta a los empleados y los sistemas. Usando datos históricos, el sistema podrá predecir qué áreas son más vulnerables y sugerir acciones preventivas sencillas, como el cambio de contraseñas.

**6. Integración Continua y Pruebas Automatizadas**

Se integrará la plataforma en un pipeline de CI/CD, utilizando herramientas DevSecOps de CI para realizar pruebas de seguridad automáticas (SAST, DAST y SCA).

**Por ejemplo**: cada vez que un desarrollador suba un nuevo cambio al sistema, se ejecutarán pruebas automáticas para detectar posibles vulnerabilidades, como contraseñas expuestas o configuraciones incorrectas.

**7. Mecanismo de Ajuste y Mejora Continua (STRIDE)**

Finalmente, se implementará **STRIDE** (Spoofing Tampering Repudiation, Information Disclosure, Denial of Service y Elevation of Privilege), un mecanismo que correlaciona datos de amenazas conocidas con nuevos patrones descubiertos.

Esta estrategia, al combinar múltiples marcos de modelado de amenazas y una infraestructura de IA adaptativa, garantiza una mejora constante en la calidad y precisión de los resultados generados, alineándose con las necesidades de seguridad de la organización.

**Caso de Uso: Modelado de Amenazas en una Aplicación.**

**Contexto:** Una entidad quiere proteger su aplicación core contra posibles ciberataques, garantizando la calidad y precisión de los resultados generados por la plataforma de modelado de amenazas asistida por IA.

**Flujo del Caso de Uso:**

**Ingreso y Validación de Datos:**

El usuario que requiere validar el modelado tiene los datos de los siguientes inputs:

"case\_description": "Evaluar riesgos relacionados con Spoofing y Tampering en las comunicaciones",

"system\_architecture": "Arquitectura en microservicios",

"critical\_assets": "Base de datos de usuarios",

"data\_flows": "Transferencia de datos entre frontend y backend",

"involved\_actors": "Usuarios, administradores, servicios externos",

"security\_controls": "Autenticación multifactor, cifrado TLS"

**Análisis y Clasificación de Amenazas STRIDE:**

La IA identifica que el comportamiento es anómalo, lo clasifica como manipulación de datos (**STRIDE**: Tampering).

**Evaluación de Riesgos con DREAD:**

El sistema evalúa la amenaza utilizando DREAD:

Damage (Daño): 7 (puede afectar las transacciones)

Reproducibility (Reproducibilidad): 5 (requiere ciertos parámetros específicos)

Exploitability (Explotabilidad): 6 (el ataque es posible si se obtienen credenciales)

Affected Users (Usuarios Afectados): 4 (solo el usuario involucrado)

Discoverability (Descubribilidad): 3 (es poco visible en los registros normales)

**Promedio DREAD: 5 (Amenaza Moderada)**

**Simulación y Evaluación de Impacto (PASTA):**

El sistema ejecuta una simulación de ataque, evaluando cómo una explotación exitosa podría afectar las transacciones.

Determina que el impacto sería el acceso no autorizado a cuentas bancarias específicas, afectando un 5% de las cuentas si no se mitiga la vulnerabilidad.

**Acciones de Mitigación:**

La IA recomienda acciones inmediatas: bloqueo temporal de la cuenta comprometida, notificación al usuario y al equipo de seguridad.

Además, sugiere habilitar un mecanismo de validación adicional para transacciones superiores a cierto monto.

**Validación del Resultado:**

Se realiza una revisión posterior utilizando el módulo de retroalimentación de usuarios. Se encuesta al equipo de seguridad sobre la calidad de la detección y las acciones tomadas.

La IA ajusta sus parámetros en función de las respuestas, asegurando una mejora continua en la precisión y calidad de los resultados generados.

**Conclusión:**

Este flujo de modelado de amenazas garantiza una evaluación integral de las amenazas, utilizando PASTA para el análisis del contexto, STRIDE para la clasificación, DREAD para la priorización. El resultado es un sistema que aprende continuamente, mejorando su capacidad de detección y respuesta, de acuerdo con la información suministrada.

Holber Hernández

POC – Modelo de amenazas con IA