Universidad Nacional de General Sarmiento

Proyecto Profesional I
Reunión Formal 2

Datacript

Docentes:

- Juan Carlos Monteros
- Francisco Orozco De La Hoz
- Leandro Dikenstein

Integrantes:

- Benitez Micaela - DNI: 43 473 284
- Benitez Yamila - DNI: 41 199 878
- Clauser Nahuel - DNI: 39 803 927
- Gómez Federico - DNI: 40 743 800
- Perez Giannina – DNI: 43 729 769
- Prieto Lucas - DNI: 43 626 494
- Torrico Franco – DNI: 42 370 140

2023

Índice

Introducción

Este documento contiene información detallada sobre los pasos que llevamos a cabo durante este Sprint, con sus correspondientes investigaciones.
En primera instancia, describimos las etapas que seguimos para definir nuestra Estrategia de Seguridad de la Información. Luego, realizamos una clasificación de datos con respecto a la ley de protección de datos personales, con el fin de otorgar la política de seguridad correspondiente a cada dato.

Por otro lado, además de visualizar el diseño definido para la base de datos junto a los pasos seguidos para realizarlo, comentamos los aspectos y dificultades relevantes que surgieron al momento de implementar la base de datos. A su vez, agregamos funcionalidades a nuestra API, con el fin de utilizarla como prueba de vinculación con los demás equipos. Por último, llegamos a un acuerdo con los demás equipos para poder definir los roles y permisos de usuario de la plataforma.

Estrategia de Seguridad de la Información (ESI)

La seguridad de la información es una disciplina estratégica que se encarga de garantizar la confidencialidad, integridad, disponibilidad y trazabilidad. La Estrategia de Seguridad de la Información (ESI) se implementa para proteger procesos, tecnología y personas. En conclusión, todos los entes que se encuentran expuestos a vulnerabilidades y amenazas que los atacantes tratarán de explotar para comprometer las propiedades de la seguridad de la Información. Estos son los pasos a seguir para definir la ESI:

1. **Clasificación de activos:** Evaluar los activos asegura que a aquellos que son los más importantes para la continuidad del proyecto, se les dé la máxima prioridad cuando se implementen las medidas de seguridad.
2. **Identificación de amenazas:** Identificación de entes que podrían querer robar o dañar los activos identificados en el paso uno, así como por qué y cómo podrían hacerlo. A cada amenaza identificada, se le asigna un nivel de amenaza basado en la probabilidad de que ocurra.
3. **Identificación de vulnerabilidades:** Una vulnerabilidad es una debilidad que una amenaza puede explotar para abrir una brecha en la seguridad y robar o dañar activos clave. Durante este paso, las pruebas de penetración y las herramientas automatizadas de escaneo de vulnerabilidad pueden ayudar a identificar las vulnerabilidades.

4. **Perfil de riesgo:** El proceso de perfil de riesgo mide el riesgo para cada combinación de activo-amenaza-vulnerabilidad, para luego asignar un puntaje de riesgo. Estos puntajes están basados en el nivel de amenaza y el impacto en la compañía si el riesgo de verdad ocurriese. Esto nos permite priorizar correctamente las vulnerabilidades identificadas y concentrar sus esfuerzos en los riesgos más significativos.
5. **Tratamiento de riesgos:** Luego de la evaluación, el riesgo debe ser tratado, transferido, tolerado o eliminado. Cada decisión debe ser documentada junto con sus justificaciones. Este proceso se debe realizar para cada escenario de amenaza, para que de esta forma se prioricen los riesgos más significativos.

Para la **clasificación de activos** consideramos varias categorías clave como:

a. **Hardware:** Servidor.
b. **Software:** Base de datos.
c. **Datos:** Datos de médicos y entidades médicas, contraseñas, resultados e imágenes.
d. **Personal:** Administrador, médicos y auditor.
e. **Procesos:** Cifrado de datos.

En cuanto a la identificación de **amenazas** de los activos mencionados anteriormente, hay varios puntos a tener en cuenta como el fallo del suministro eléctrico, que pueden causar interrupciones en los sistemas y pérdida de datos, la presencia de virus, malware y otros tipos de software malicioso que pueden infectar y causar daño a sistemas operativos, aplicaciones y bases de datos, el acceso no autorizado, robo o manipulación de los datos por parte de atacantes que intentan acceder a información confidencial, la divulgación no autorizada de información, la mala gestión de contraseñas, el phishing y la negligencia en la seguridad de la información y como último, el incumplimiento de políticas de seguridad, lo que podría dejar sistemas y datos vulnerables a ataques y brechas de seguridad.

A partir de estas amenazas, identificamos **vulnerabilidades** que deben atacarse cuanto antes para evitar futuros problemas, ellas son:

- Fallos en hardware que pueden causar interrupciones en el servicio.
- Configuraciones incorrectas de seguridad, como contraseñas predeterminadas o débiles.
- Errores de programación, como desbordamientos de búfer o inyección de código SQL.
- Malware y software malicioso que puede infectar sistemas y robar datos.
- Acceso no autorizado a datos sensibles.
- Fuga de datos debido a errores humanos o fallos técnicos.
- Falta de cifrado en datos almacenados.
- Mal uso de privilegios por parte de empleados o administradores de sistemas.
- Falta de capacitación en seguridad para el personal.

- Incumplimiento de políticas de seguridad por parte de empleados.
- Falta de revisión y actualización de políticas obsoletas.

Con respecto a los **perfiles de riesgo** de nuestros activos, los medimos de la siguiente manera tomando en cuenta cada categoría:

a. **Personal → Perfil de Riesgo: Medio**

   El riesgo asociado con el personal puede variar según su rol y nivel de acceso a sistemas y datos. Los administradores de sistemas y otros usuarios privilegiados pueden representar un mayor riesgo si sus credenciales son comprometidas o si cometen acciones maliciosas. La capacitación y la gestión de accesos son importantes para mitigar estos riesgos.

b. **Procesos → Perfil de Riesgo: Medio**

   Si bien los procesos en sí mismos pueden no ser directamente vulnerables a ataques, la falta de procesos adecuados o la violación de políticas y procedimientos pueden conducir a brechas de seguridad.

c. **Hardware → Perfil de Riesgo: Alto**

   Los activos de hardware a menudo almacenan y procesan datos críticos para la organización. Su pérdida o compromiso puede tener un impacto significativo en la continuidad del negocio y la confidencialidad de la información. También son objetivos atractivos para robos y ataques físicos.

d. **Software → Perfil de Riesgo: Alto**

   El software es esencial para el funcionamiento de los sistemas de información de la organización. Las vulnerabilidades en el software pueden ser explotadas por atacantes para comprometer la seguridad de los datos y los sistemas. La pérdida de integridad o disponibilidad de estos activos puede ser perjudicial.

e. **Datos → Perfil de Riesgo: Muy Alto**

   Los datos son a menudo el activo más valioso de una organización. La confidencialidad, integridad y disponibilidad de los datos son fundamentales. La pérdida o filtración de datos confidenciales puede tener graves repercusiones legales, financieras y de reputación.

Por último, para **tratar estos riesgos** consideramos las siguientes medidas a tener en cuenta:

1. Promover buenas prácticas de seguridad cibernética.

2. Implementar políticas y procedimientos de gestión de contraseñas sólidas y requerir cambios periódicos de contraseñas.
3. Utilizar cifrado para proteger los datos.
4. Mantener el software y los sistemas operativos actualizados con parches de seguridad.
5. Realizar copias de seguridad regulares y almacenarlas de forma segura para poder recuperar los datos en caso de un ataque.
6. Implementar controles y procedimientos para hacer cumplir las políticas de seguridad. (corresponde al auditor y podríamos realizar automatizaciones)

Clasificación de datos según Ley 25326: Protección de datos personales

El primer paso que llevamos a cabo para poder realizar la clasificación de los datos maestros incluídos en nuestro proyecto, fue investigar sobre la ley de protección de datos personales.

“En la República Argentina se dictó la ley 25.326, instituyendo un mecanismo de protección de los datos personales, que incluye los datos sensibles y de salud; así también estableció principios de licitud de tratamiento de datos personales en general, que resultan aplicables a los datos relativos a la salud.”

En resumen, la ley de protección de datos personales te protege si tus datos de identidad, de salud o de crédito son usados sin tu consentimiento.

Durante nuestra investigación, nos encontramos con que existen cuatro tipos de datos.

- **Datos públicos:** Datos que pueden ser obtenidos y ofrecidos sin reserva alguna y sin importar si se trata de información general, privada o personal. Por ejemplo, datos obtenidos por fuentes de acceso público.
- **Datos semiprivados:** Datos que no tienen naturaleza íntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no sólo a su titular sino a cierto sector o grupo de personas o a la sociedad en general.
- **Datos privados:** Datos que por su naturaleza íntima o reservada sólo son relevantes para el titular. La información privada es aquella que por versar sobre información personal y por encontrarse en un ámbito privado, sólo puede ser obtenida y ofrecida por orden de autoridad judicial en el cumplimiento de sus funciones.
- **Datos sensibles:** Datos que afectan la intimidad del titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas o la pertenencia a sindicatos. Por ejemplo, datos biométricos y datos genéticos.

A partir de esto, realizamos la siguiente clasificación de datos, con el fin de priorizar y determinar qué datos necesitarán cifrado u otra política de seguridad:

- **Hospitales, clínicas y sanatorios:**
    - ID: Público
    - Nombre: Público
    - Dirección: Público
- **Profesionales médicos:**
    - ID: Privado
    - DNI: Público
    - Correo electrónico: Privado (política de seguridad)
    - Especialización: Público
    - Entidad donde atiende: Público
    - Medio de contacto preferente (mail, whatsapp, telegram, etc): Privado
    - Contraseña: Privado (cifrado)
- **Pacientes:**
    - Nombre: Público
    - Edad: Público
    - Sexo: Público
    - Altura: Privado (política de seguridad)
    - Peso: Privado (política de seguridad)
    - Índice de Masa Corporal (IMC): Privado (política de seguridad)
    - Antecedentes: Sensible (cifrado)
    - Imagen a predecir: Sensible (cifrado)

Roles y permisos de usuario

- **Administrador / Auditor:**
    - Cargar imágenes.
    - Cargar informes.
    - Consultar historial.
    - Acceso a datos encriptados para gestionar claves de acceso.
- **Administrador:**
    - Acceso a ABM de usuarios.
- **Profesional de la salud:**
    - Cargar imágenes.
    - Cargar informes.

- **Médico:**
    - Acceso al resultado de clasificación.
    - Agregar feedback sobre el resultado de la clasificación.
    - Consultar historial.

Guía de buenas prácticas de seguridad cibernética

- Mantener actualizado el sistema operativo y todo el software del dispositivo utilizado.
- Instalar y mantener actualizado el software antimalware.
- Realizar copias de seguridad de los datos sensibles almacenados.
- No dejar material confidencial o sensible desatendido.
- Ser precavidos con los mensajes que se reciban por correo electrónico, redes sociales o mensajería instantánea.
- Prestar especial atención para evitar acceder a archivos o sitios web que podrían tener contenido malicioso.
- Solamente descargar archivos, software y aplicaciones de fuentes o sitios de confianza.
- Siempre mantener tu equipo resguardado y seguro mientras no lo estés utilizando.

Definición de políticas para contraseñas

Es fundamental establecer políticas de creación de contraseñas sólidas para los usuarios de un sistema, ya que estas políticas ayudarán a garantizar la seguridad de las cuentas y la integridad de los datos. Nuestro equipo tendrá en cuenta las siguientes políticas:

1. La longitud de la contraseña será mínimo de 8 caracteres.
2. La contraseña estará compuesta por una combinación de letras mayúsculas, minúsculas, caracteres numéricos y símbolos especiales como los siguientes: ` ~ ! @ # $ % ^ & * ( ) _ + - = { } | [ ] \ : " ; ' < > ?
3. No se repetirán los mismos caracteres en la misma contraseña. (ej.: “111222”).
4. Se cambiará con una periodicidad de 60 días.
5. No contendrá el DNI, ni datos personales.

Diseño e Implementación de Base de Datos

La base de datos se encuentra armada con el propósito de realizar pruebas antes de registrar una base de datos completamente nueva y funcional.

Descripción general de la base de datos:

- Nombre BD: Pruebas

- Propósito: PruebasBD será una base de datos relacional diseñada para almacenar datos de pruebas que posteriormente van a ser levantados con API's. Su propósito principal será para que tanto el equipo de front pueda utilizar los datos y los mismo hacia los equipos de los modelos.
- Scope: La BD contendrá datos de los médicos y usuarios, imágenes médicas de prueba, información del historial de todo lo realizado.
- Partes clave interesadas: Equipo de front y los equipos de los modelos.
- Meta: PruebasDB busca alcanzar lo más posible a como sería la relación con ambos equipos.
- Fuente de los datos: Momentaneamente toda será inventada por nosotros.
- Frecuencia de las actualizaciones: La BD se actualiza en tiempo real a medida que los clientes interactúan con nuestra plataforma.
- Accesibilidad: El acceso a PruebasBD es controlada a través de roles de usuario, con diferentes niveles de acceso dependiendo del equipo.
- Dependencias: Actualmente no cuenta con ningún tipo de dependencia, en un futuro la base de datos se basará en los datos recibidos en la plataforma de ingreso de datos diseñada y también se usará un sistema de reportes para las analíticas.
- Arquitectura de alto nivel: La base de datos usa PostgreSQL como el DBMS y está hosteado a través de Ngrok. Estará integrado con el front y los modelos a través de APIs.
- Planes a futuro: Como se menciona al principio esto será la base para poder armar la base de datos principal.

Actualizaciones de API REST para Pruebas

Para continuar con las pruebas de integración con los demás equipos agregamos los siguientes endpoints:

1. **"/historial/{id}"**, proporciona información sobre un diagnóstico médico identificado por su ID. Al acceder a este endpoint con un ID válido, el servicio responderá con un objeto JSON que contiene detalles del diagnóstico asociado a ese ID.

   Los atributos que se pueden esperar en la respuesta son:

   - **id:** identifica de manera única el diagnóstico médico.
   - **descripción:** contiene una cadena de texto que describe la información relevante del diagnóstico médico.
   - **estudio**: indica el tipo de estudio médico realizado en el diagnóstico.
   - **fecha:** representa la fecha en la que se registró el diagnóstico médico.
   - **hora:** indica la hora en la que se registró el diagnóstico médico.

2. **“/predicciones”**, proporciona un JSON que contiene detalles sobre la predicción del modelo con respecto a la clasificación de una imagen.

   Los atributos que se pueden esperar en la respuesta son:
   - **precisión:** indica la precisión de la predicción y puede tener valores en el rango de 0.00 a 1.00.
   - **predicción:** describe el nivel de confianza de la predicción. Se divide en tres rangos:
     - De 0.00 a 0.60: Clasificación no certera.
     - De 0.60 a 0.65: Clasificación poco certera.
     - De 0.65 a 1.00: Clasificación muy certera.

3. **POST: “/Diagnosticos”** En este caso se creará un nuevo diagnóstico con las siguientes entradas en formato JSON llamado desde la api. Este llamado trabaja con un archivo JSON, donde se hace el guardado de los datos del diagnóstico antes de la integración a la base de datos.

   Los atributos que se pueden esperar en el envío de datos JSON son:

   - **Usuario ID:** es el usuario al que pertenece el diagnóstico
   - **Edad:** indicar la edad de la persona
   - **Peso:** indicar el peso de la persona
   - **Altura CM:** ingresar la altura registrada que tuvo la persona.
   - **Sexo:** indicar el sexo de la persona
   - **Sección del cuerpo:** ingresar la parte del cuerpo sometida al diagnóstico
   - **Condiciones Previas:** ingresar una cadena de texto que describa la información relevante del diagnóstico médico.
   - **Imagen:** en este caso imagen ingresar la palabra imagen, ya que esperamos trabajar con imágenes los próximos pasos.

4. **“/Diagnosticos/all”** proporciona un JSON con una lista con información de todos los diagnósticos que hayan incluidos en la base, en este caso en un guardados en archivo json, el que cumple el rol de guardado antes de la integración a la base.
   - **id:** identifica de manera única el diagnóstico
   - **Usuario ID:** es el usuario al que pertenece el diagnóstico
   - **id rol:** es el rol del usuario

   Lista de diagnósticos/resultados:

- **Edad:** indica la edad que tuvo al realizarse el diagnóstico.
- **Peso:** indica el peso de la persona.
- **Altura CM:** representa cuál fue la altura registrada que tuvo la persona.
- **Sexo:** indica el sexo de la persona.
- **Sección del cuerpo:** indica la parte del cuerpo que fue sometida al diagnóstico.
- **Condiciones Previas:** contiene una cadena de texto que describe la información relevante del diagnóstico médico.
- **Imagen:** en este caso imagen devuelve la palabra imagen, ya que esperamos trabajar con imágenes los próximos pasos.

5. **"/Diagnosticos/{id_diagnostico}"** proporciona un JSON con la información del diagnóstico indicado por ID que haya incluido en la base, primero se verifica que exista, en este caso en un archivo json el que cumple el rol de guardado antes de la integración a la base.
    - **id:** identifica de manera única el diagnóstico
    - **Usuario id:** es el usuario al que pertenece el diagnóstico
    - **id rol:** es el rol del usuario

    lista de diagnósticos/resultados:

    - **Edad:** indica la edad que tuvo al realizarse el diagnóstico
    - **Peso:** indica el peso de la persona
    - **Altura CM:** representa cuál fue la altura registrada que tuvo la persona.
    - **Sexo:** indica el sexo de la persona
    - **Sección del cuerpo:** indica la parte del cuerpo que fue sometida al diagnóstico
    - **Condiciones Previas:** contiene una cadena de texto que describe la información relevante del diagnóstico médico.
    - **Imagen:** en este caso imagen devuelve la palabra imagen, ya que esperamos trabajar con imágenes los próximos pasos.

6. **"/Diagnosticos/deleteAll"** cumple la función delete el cual funciona provisoriamente para vaciar o limpiar el archivo JSON trabajado para probar la funcionalidad de la misma.

Evaluación de Vulnerabilidades en el Sistema

La seguridad de las aplicaciones web es una preocupación crítica en el entorno cibernético actual. Las amenazas a la seguridad, como la inyección de SQL, el cross-site scripting (XSS) y otros ataques, son persistentes y pueden tener consecuencias devastadoras para las organizaciones. En esta sección, se discutirá la decisión de utilizar la herramienta OWASP ZAP (Zed Attack Proxy) para probar y evaluar las vulnerabilidades en el sistema.

1. **Amplia Cobertura de Vulnerabilidades:** OWASP ZAP es ampliamente reconocida por su capacidad para identificar y evaluar una amplia gama de vulnerabilidades en aplicaciones web. Esto incluye ataques comunes como inyección de SQL, XSS, CSRF y muchos otros. La herramienta ofrece una cobertura integral que nos permitirá identificar y corregir problemas críticos de seguridad en nuestro sistema.

2. **Interfaz de Usuario Intuitiva:** La interfaz gráfica de usuario (GUI) de OWASP ZAP es intuitiva y fácil de usar. Esto nos permite utilizar la herramienta de manera efectiva sin requerir un conocimiento profundo en seguridad informática. Además facilita la incorporación de pruebas de seguridad en el ciclo de desarrollo.

3. **Comunidad Activa de Usuarios y Desarrolladores:** OWASP ZAP es respaldada por la Open Web Application Security Project (OWASP), lo que significa que está respaldada por una comunidad activa de usuarios y desarrolladores. Esto garantiza actualizaciones regulares, soporte y la disponibilidad de recursos en línea para resolver problemas y compartir conocimientos.

4. **Personalización y Automatización:** La herramienta permite la personalización de pruebas de seguridad para adaptarse a las necesidades específicas de nuestro sistema. Además, ofrece la capacidad de automatizar el proceso de escaneo de vulnerabilidades, lo que facilita la integración en flujos de trabajo de desarrollo y pruebas continuas (CI/CD).

5. **Reportes Detallados:** OWASP ZAP genera informes detallados sobre las vulnerabilidades detectadas. Estos informes son esenciales para comunicar hallazgos al equipo de desarrollo y para la priorización de correcciones. Ayudarán a garantizar que nuestro sistema sea más seguro antes de su implementación.

Próximos pasos

En nuestro próximo Sprint, nos enfocaremos en la implementación de los diversos aspectos que hemos definido hasta este punto. Estas implementaciones abarcan el cifrado con AES, la prueba de penetración y las políticas para contraseñas. Además, nos ocuparemos de la gestión de roles y permisos para diferentes tipos de usuarios, incluyendo administradores/auditores, administradores, profesionales de la salud y médicos.

Por otro lado, continuaremos trabajando en la definición de aspectos cruciales de ciberseguridad que serán implementados en fases posteriores. Esto incluye la planificación para implementar actualizaciones de software y un backup para la base de datos.