EXTERNALIZACIÓN DE CÓMPUTO SEGURO: CRIPTOGRAFÍA Y SUS APLICACIONES

1. INTRODUCCIÓN

En los últimos años hemos visto cómo nuestra forma de guardar y gestionar la información ha evolucionado. Un claro ejemplo es la toma de apuntes en la Universidad. Hace no muchos años los apuntes se tomaban a mano y se compartían realizando fotocopias. Hoy en día está cada vez más extendido el uso de medios digitales tanto para tomarlos: a través de fotografías, tablets, etc. como para compartirlos: correo electrónico, aplicaciones de mensajería o proveedores cloud.

En el mundo de la empresa, se ha dado una evolución similar. Los datos que poseen y recolectan las empresas cada vez son más numerosos y tienen un alto valor, pero además, en muchos casos, son sensibles o privados. Pensemos en grandes sectoriales como banca o seguros, donde gracias a los datos de sus procesos y de sus clientes son capaces de tomar decisiones orientadas a mejorar su negocio. Estos datos, por un lado son privados -no queremos que nadie sepa cuándo y dónde hemos tenido un accidente de tráfico o a qué hora salimos a hacer deporte- pero también tienen que estar seguros, ya que la empresa que los ha generado puede no querer compartirlos con otras de su competencia.

Debido al gran volumen de información que estas empresas generan, muchas tienen que recurrir a Proveedores cloud que les ayudan a gestionar estos volúmenes de información de una forma ágil y escalable sin necesidad de reorganizar toda su infraestructura.

El gran reto para estas empresas que externalizan sus datos es la de tener la certeza de que sus datos son almacenados de forma segura. El mundo de la seguridad es muy amplio, y aquí nos vamos a centrar en la seguridad en el uso, es decir, cuando los datos se están procesando (por ejemplo, cuando se cargan en memoria para realizar un cálculo). Este tipo de seguridad es además la más vulnerable, ya que los datos hay que leerlos de algún repositorio, procesarlos, y volverlos a escribir, y todo esto garantizando la seguridad del dato y de los cálculos que se realicen.

2. EL PROBLEMA

Dado un conjunto de datos, **el objetivo es poder analizarlo de una forma segura**, y que en ningún caso nadie pueda ver los datos, exceptuándome a mí, que soy el responsable y encargado de ejecutar estos análisis.

Este problema lo podemos dividir en dos fases:

1. Encontrar una forma de **guardar los datos** de forma segura. En esta fase, tengo los datos y yo sí puedo verlos. Guardar varios dataset en mi ordenador supondría a la larga una falta de espacio, por lo que una solución cloud sería la adecuada. De esta manera, antes de enviar los datos a un proveedor cloud tengo que encargarme de que nadie con acceso a ese servidor remoto pueda verlos (ni siquiera yo o los administradores del servidor).
2. Encontrar una forma de **procesar los datos**. En esta fase, una vez que los datos están almacenados de forma segura, tengo que poder realizar los análisis sobre los mismos, siendo yo el único que pueda ver el resultado.

Para este problema vamos a usar el **dataset de cáncer** de pulmón disponible en https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer

Como podemos observar, son datos realmente sensibles cuya seguridad debe estar garantizada, pero sin embargo el análisis que queremos realizar es sencillo, queremos contar cuántos casos de cáncer existen en función de si los individuos son fumadores o no (invitamos al estudiante a aumentar/simplificar la complejidad de esta consulta como crean oportunos).

En GMV llevamos varios años trabajando en Técnicas de Preservación de la Privacidad (PET, por sus siglas en inglés), donde empleamos distintos esquemas de criptografía, como la **Encriptación Homomórfica** (y sus variantes, SomeWhat y Partial) para poder llevar a cabo este tipo de soluciones.

3. PLAN DE TRABAJO Y OBJETIVOS DE APRENDIZAJE

Una vez que se hayan creado los equipos de trabajo, el procedimiento es el siguiente:

1. Al comenzar el evento, se realizará una reunión inicial con los equipos inscritos en el problema, orientándoles sobre el camino a seguir.
2. A mitad del evento se contactará con los equipos para revisar los avances y resolver las posibles dudas.
3. Al finalizar, se sintetizan y preparan los resultados antes de enviar la solución.

Los objetivos de aprendizaje son:

- Familiarizarse con los problemas que tienen las empresas y cómo las matemáticas pueden ayudar a solucionarlos.
- Familiarizarse con las Técnicas de Preservación de la Privacidad.
- Plantear y formular una solución a un problema dado.
- Análisis, discusión y presentación de resultados.

4. REFERENCIAS

- Moreno, P. (2022). How to Secure the 3 States of Data: At Rest, In Motion, In Use. Retrieved 27 September 2022, from https://accelerationeconomy.com/cybersecurity/how-to-secure-the-3-states-of-data-at-rest-in-motion-in-use/

- Homomorphic encryption - Wikipedia. (2022). Retrieved 27 September 2022, from https://en.wikipedia.org/wiki/Homomorphic_encryption

- Paillier cryptosystem - Wikipedia. (2022). Retrieved 27 September 2022, from https://en.wikipedia.org/wiki/Paillier_cryptosystem