Nociones de Privacidad

La privacidad puede tener varias definiciones:

- La capacidad de un individuo o grupo de individuos de ocultarse u ocultar información sobre ellos.
- El derecho de individuos, grupos o instituciones de determinar cuando, como y que información sobre ellos puede ser comunicada a otros.
    - Otros: Adversaros
        - Sociedad
            - Compañías
            - Otros individuos
        - Estados

Privacidad y Seguridad

- La Seguridad es un medio para alcanzar la privacidad. La seguridad coincide con la privacidad en lo siguiente:
    - Existencia de adversarios estrategicos
    - Muchos principios de diseño de seguridad también se aplican a la seguridad
- Por otro lado, la privacidad:
    - Transciende el dominio digital
    - Modelo de amenaza: Muchas veces actores débiles, en ocasiones, adversarios poderosos.
    - No se puede asumir la existencia de terceras partes de confianza

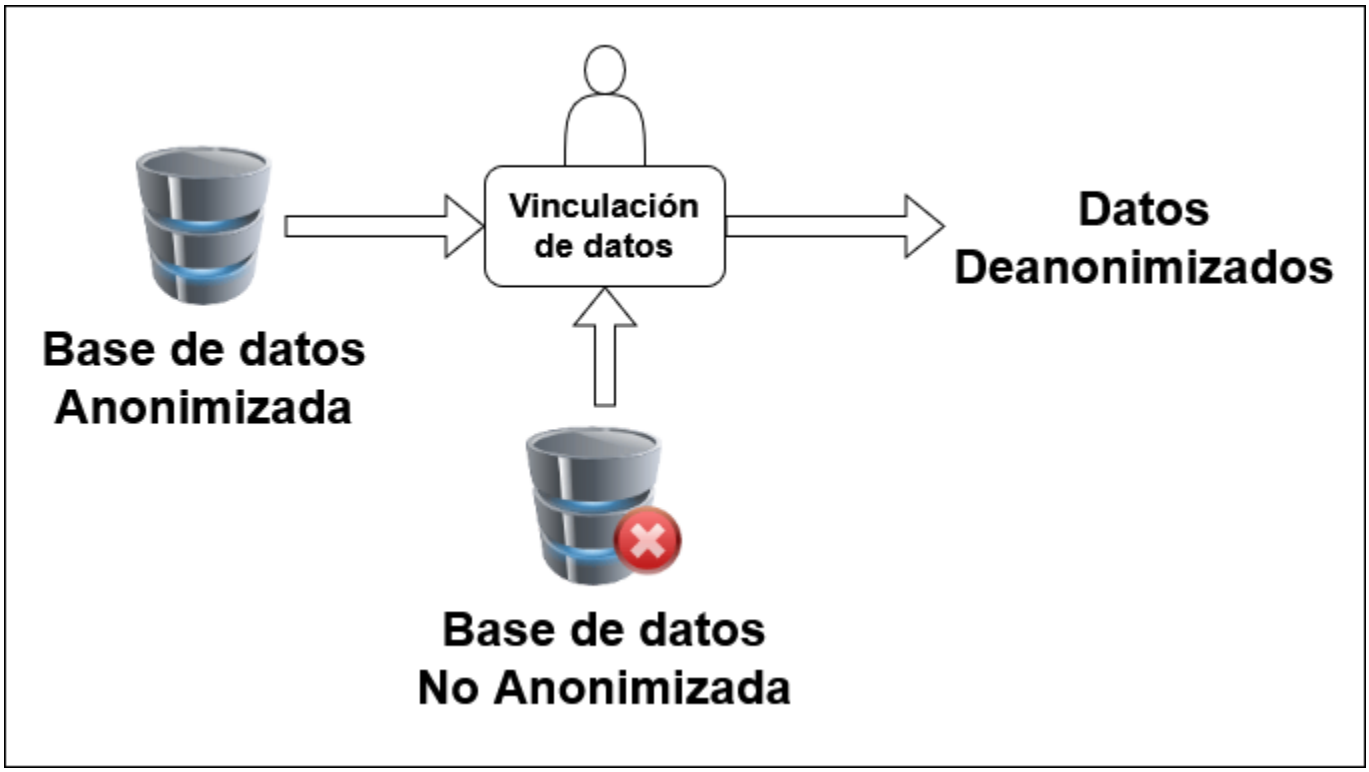
Ataques de inferencia

- Tratan de averiguar toras cosas a través de los datos disponibles.
- Inferencia estadística de datos observados:
    - Los datos oservados pueden decir mucho más de lo esperado ya que somos predecibles
    - Ejemplo: Likes de Facebook
        - Los Likes de Facebook son buenos predictores de atributos que la gente considera privados.
        - Los Likes pueden ser vistos por los adversarios.
        - Predictores de inteligencia alta:
            - Thunderstorms
            - Colert Report
            - Science
            - Curly Fires
        - Predictores de baja inteligencia:
            - Sephora
            - Harley Davidson
            - Lady Antebellum
    - Cambridge Analytica:
        - Extracción de perfiles psicológicos de la huella digital de los usuarios para influenciar sus emociones o comportamiento.

Deanonimización por vinculación de datos (data linking)

Usando bases de datos externas es posible desanonimizar a personas con atributos externos. El adversario puede acceder por cualquier clase de medio a una base de datos no anonimizada y cruzando los datos es capaz de desanonimizar a cualquier individuo.

- Los datos anonimizados pueden ser deanonimizados enlazandolos con atributos externos.
- Cuantos más escasos sean los datos, más únicos son, por lo que es más fácil vincularlos.



Deanonimización de Netflix

Netflix lanzó un concurso para buscar herramientas de recomendación para sus usuarios. Lo que hicieron fue ofrecer un premio a quien pudiera crear el mejor sistema de recomendaciones. Cada uno de los datasets estaba conpuesto por el ID de usuario anonimizado, un ID de una película, la puntuación que le dió dicha persona y la fecha de la puntuación. Se pensaba que esta información era completamente privada y segura.

La base de datos era bastante dispersa por los miles de atributos que habían, el conjunto de películas que una persona había visto era una variable casi única. Algunas películas habían sido vistas por solo un pequeño grupo de personas, siendo esto atributos únicos. Lo que se hizo fue medir la similitud entre los registros de datos con cierta tolerancia entre la puntuación y la fecha. Se empezaron a utilizar 1 y 0 para indicar si una película gustaba o no. Se tomó la base de datos de netflix y se cruzón con IMDb, buscando la gente que dio puntuaciones parecidas en las mismas fechas consitentemetne.

Con este método se pudieron identificar 2 de cada 50 personas cruzando los datos con IMDB. Las consecuencias de esto fue una denuncia contra netflix por parte de las personas cuyos datos estaban en eld ataset al ser esto una violación de la privacidad. Netflix tuvo que pagar 9 Millones de dólares a los usuarios afectados.

Deanonimización basada en la localización

Se tomaron datos GPS de vehículos en el área de Detroit con un minuto de diferencia. Cuando los coches estaban apagados no enviaban datos. Se eliminaron sitios donde no se trabajaba por la tarde y se eliminaron los coches fuera de las áreas residenciales. Con estos datos fue posible localizar la casa de múltiples individuos. Durante 2 semanas, con los datos de 172 personas, mediante el uso de heurística (Las 3 Am estaba en casa, Luegar donde pasaban más tiempo, Uso de geolocalización reversa y páginas amarillas) para separar los vehículos se logró desanonimizar a un 5% de los sujetos.

GDPR

Datos personales: cualquier información relacionada con un individuo que puede ser directa o indirectamente identificada. Nombres y direcciones de correo son datos personales obvios. Los datos pseudoanónimos también pueden caer bajo esta definición.

Principios de la GDPR

- El procesamiento debe ser legal, justo y transparente para el dueño de los datos.
- Se debe especificar claramente para que son los datos
- Minimización de datos: solo se deben pedir los datos absolutamente necesarios para el servicio que se ofrece
- Límite de almacenamiento: Solo se puede alamacenar la información que identifica a una persona si es necesaria para el propósito especificado
- Integridad y confidencialidad: El procesamiento se debe hacer de tal manera que asegure la seguridad, integridad y confidencialidad.
- Responsabilidad: El controlador de los datos es responsable de demostrar que se cumple con la GDPR y todos sus principios.