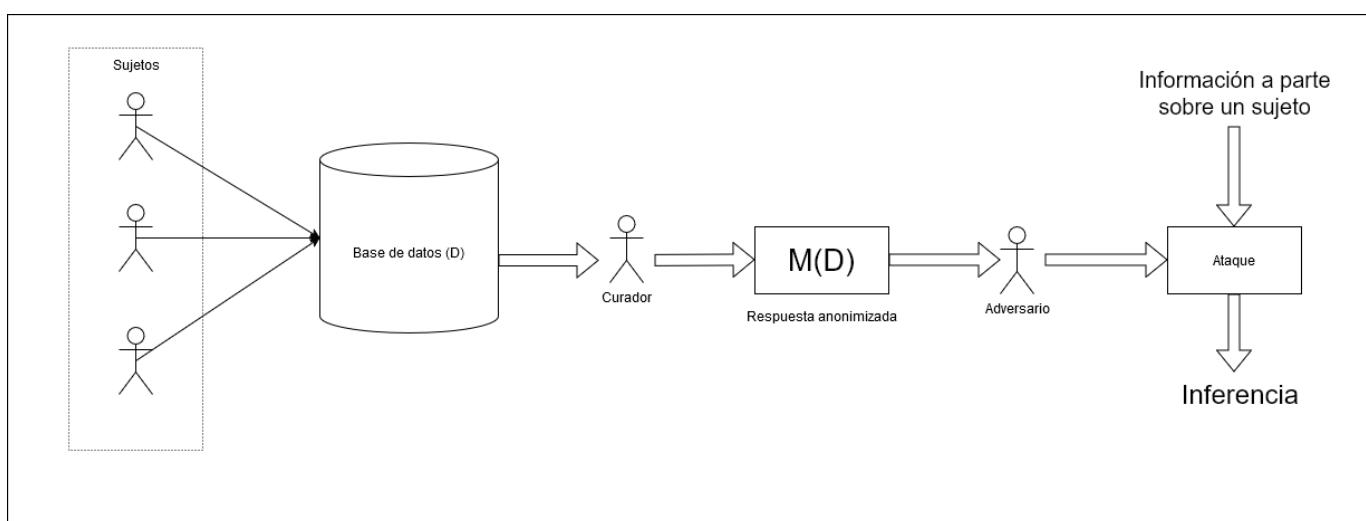


# [PAN] Privacidad Diferencial (Resumen)

## Caso base

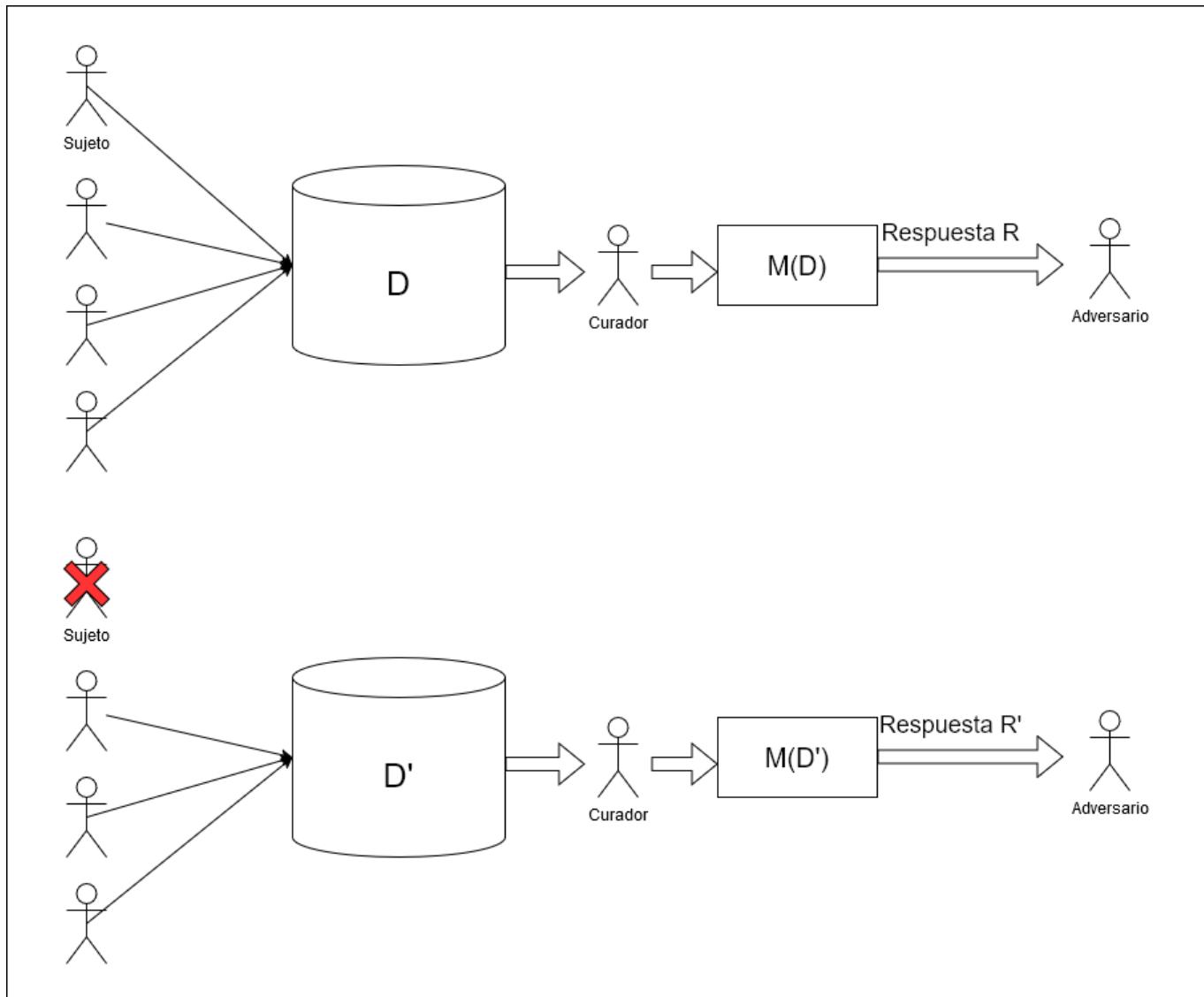
Tenemos un dataset  $D$  que contiene datos de usuarios, siendo cada fila los datos de un usuario. El Curador, que es una entidad de confianza para los usuarios, publica algunos datos usando un mecanismo  $M$  que da como resultado  $R=M(D)$ . El adversario trata de realizar inferencias sobre los datos  $D$  contenidos en  $R$ .

## De que protege la privacidad diferencial



La privacidad diferencial protege contra el riesgo del conocimiento de información sobre un sujeto usando inferencia con información obtenida a parte. De esta forma, observando la respuesta  $R$  no se puede cambiar lo que el adversario puede saber.

La clave para dificultar a un adversario identificar datos sobre un sujeto es poder crear dos salidas  $R=M(D)$  y  $R=M(D')$ , siendo  $D$  y  $D'$  dos datasets diferenciados por que el primero contiene al sujeto en cuestión y el segundo no, las cuales no puedan ser distinguibles la una de la otra. Para hacer esto se diseña el mecanismo  $M$ , el cual no puede ser determinístico, si no probabilístico.



La distribución de los datasets debe ser similar, es decir, dada una probabilidad  $R$  de que un dato viene del dataset  $D$ , esta tiene que ser similar a la probabilidad de que un dato venga del dataset  $D'$ . Los datasets que difieren en una fila son conocidos como vecinos. En resumidas cuentas, la probabilidad de que  $M(D)=R$  debe ser muy similar a la de que  $M(D')=R$ .

## Como definir distribuciones similares

### Definición tentativa de privacidad con parámetro $P$

Un mecanismo  $M$  es privado si para todas las salidas posibles de  $R$  un todos los pares de datasets vecinos ( $D, D'$ ):

$$|Pr(M(D')=R) - Pr(M(D)=R)| < P$$

El problema de esta definición es que existen ciertas salidas de  $R$  que solo pueden ocurrir cuando la entrada es  $D'$ , lo que permite al adversario distinguir entre  $D$  y  $D'$ .

## Definición tentativa de privacidad 2 con parámetro P

$$\Pr[\mathcal{M}(D') = R] \leq \Pr[\mathcal{M}(D) = R] \leq \Pr[\mathcal{M}(D) = R] * p$$

## Definición de Privacidad Diferencial (PD)

Un mecanismo  $M: D \rightarrow R$  es  $\epsilon$ -diferencialmente privado ( $\epsilon$ -PD) si para todas las posibles salidas  $R \in R$  y los datasets vecinos  $D, D' \in D$ :  $\Pr[M(D) = R] \leq \Pr[M(D') = R] * e^\epsilon$ . Se usa  $e^\epsilon$  en vez de  $\epsilon$  porque facilita la formulación de ciertos teoremas útiles. OJO: Si el dominio de salida del mecanismo no es discreto el sistema no funciona.

### A tener en cuenta

- Cuanto más pequeño es el valor de  $\epsilon$  Más privacidad
- La privacidad perfecta se da cuando  $\epsilon=0$ , el problema de esto es que la salida va a ser prácticamente inutil
- No existe un consenso sobre como de pequeño debe ser  $\epsilon$ , pero debe tener un valor que evite que la salida del mecanismo sea inútil.

### Sobre la privacidad diferencial y rendimiento de ataques empíricos

La privacidad diferencial asegura la protección incluso contra adversarios poderosos que conocen los inputs de  $D$  o  $D'$ . En la práctica, un algoritmo que provee  $\epsilon=10$  puede proveer una protección empírica contra ataques existentes bastante alta.

## Privacidad diferencial aproximada

Esta definición de la privacidad Diferencial permite algo más de tolerancia. Un mecanismo  $M: D \rightarrow R$  es  $(\epsilon, \delta)$ -Diferencialmente Privado si para todas las posibles salidas de  $R \subset R$  y para parejas de datasets vecinos  $D, D' \in D$ :  $\Pr[M(D) \in R] \leq \Pr[M(D') \in R] * e^\epsilon + \delta$

## Configuración de privacidad Diferencial

Dependiendo de donde se ejecuta el mecanismo hay 2 tipos de modelos:

- Privacidad diferencial Central: Hay un agregador centralizado de confianza que ejecuta el mecanismo  $M$
- Privacidad diferencial Local: Cada usuario ejecuta el mecanismo  $M$  y reporta el resultado al adversario

Existen dos definiciones sobre como se pueden definir dos datasets vecinos en un modelo central:

- Privacidad diferencial acotada:  $D$  y  $D'$  tienen el mismo número de entradas, pero se diferencian

en el valor de una de ellas.

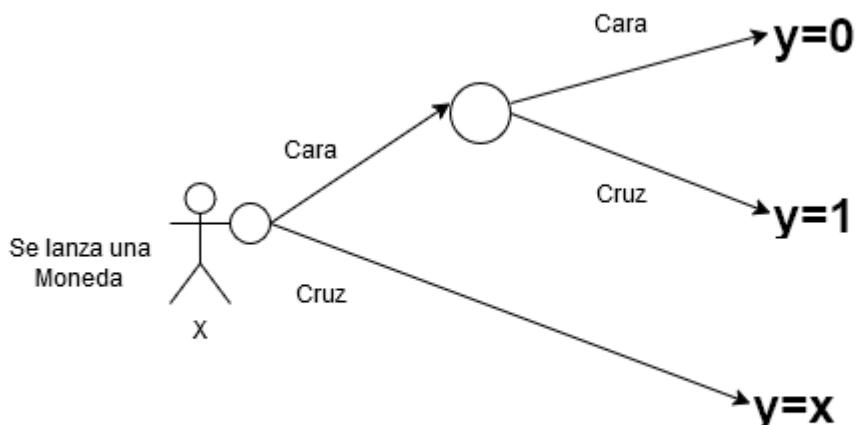
- Privacidad diferencial no acotada:  $D'$  se obtiene de  $D$  tras eliminar una entrada.

## Mecanismos de privacidad diferencial

Existen varios mecanismos que proveen privacidad diferencial y pueden ser aplicados a varios sistemas.

### Mecanismo Aleatorizado

Tenemos un mecanismo  $M: \{0,1\} \rightarrow \{0,1\}$ :



From:  
<http://www.knoppia.net/> - Knoppia



Permanent link:  
[http://www.knoppia.net/doku.php?id=pan:res\\_privacidad\\_diferencial&rev=1736275187](http://www.knoppia.net/doku.php?id=pan:res_privacidad_diferencial&rev=1736275187)

Last update: 2025/01/07 18:39