Contingut Personalitzat amb *Homomorphic Encryption*

Preservant la Privadesa dels Usuaris

Roger Rovira
Juny 2023

"A system that does not put users in control will – immediately or over time – be rejected by enough of them that it cannot become and remain a unifying technology."

~ Kim Cameron, 2005 ~

Abstract. [...]

Keywords: *[...]*

Taula de Continguts

Introducció

~ Objectius i Hipòtesis ~

Introducció.
[...]

Organization.
[...]

1

Centralització de dades personals

~ Meta, Google i Twitter ~

1.0 Secció.

Els models d'usuaris centralitzats continuen sent predominants al mercat, a causa de la prevalença dels serveis personalitzats a Internet [23]. El propòsit d'aquesta secció és informar al lector sobre la inquietant acumulació de poder que emergeix de la centralització de la informació.

1.1 Taula de subseccions.

1.2 Era digital.

En les últimes dues dècades, la humanitat ha estat testimoni del ràpid desenvolupament de tecnologies digitals capaces de millorar el benestar humà, així com ha presenciat una creixent concentració d'aquestes tecnologies en mans d'un reduït grup de potències tecnològiques, com Google, Apple, Meta, Amazon i Microsoft; sovint conegudes com les *Big Tech* [14] [36] [37]. Per consegüent, l'era digital, que es plantejava com una excel·lent oportunitat per democratitzar les institucions i millorar l'accés a la informació i al coneixement en tots els estrats de la societat [17] [14], va resultar ser una època contradictòria al que s'esperava. En lloc d'alliberar a la humanitat i fomentar la seva independència, l'era digital ha sotmès a les persones a la simple mercè de les grans companyies tecnològiques.

Fa vint-i-cinc anys, Meta no existia, Google era tan sols un projecte universitari i Amazon només s'enfocava en la venda de llibres [14]; pro, no obstant això, són actualment considerades com algunes de les empreses més valuoses i influents del món. Donada la seva immensa magnitud i abast, aquestes companyies tenen la capacitat d'influir en el comportament dels usuaris i l'economia internacional [37]; el que els atorga un poder sense precedents que es manifesta en diversos aspectes de la nostra vida quotidiana.

Un exemple rellevant és el cas de Cambridge Analytica, en el qual es va posar de manifest la funció que van exercir les dades acumulades de Facebook en les cinquantenes octaves eleccions presidencials dels Estats Units [8]. Demostrant així la gran repercussió que les dades de les grans empreses tecnològiques exerceixen sobre l'opinió pública i els processos democràtics.

De la mateixa manera que Meta s'ha apoderat de la informació dels seus usuaris, tant Google com Twitter han estat capaços de monopolitzar la informació de diverses maneres [14], col·locant-se en una posició privilegiada per capitalitzar la innovació futura i oferir serveis personalitzats.

1.3 Dades personals.

Les *Big Tech* han aconseguit establir-se entre les companyies més rellevants i poderoses del món gràcies a la seva innata habilitat per recopilar i utilitzar grans quantitats de dades en benefici del seu propi creixement. Google, per exemple, compta amb una àmplia base de dades que li permet identificar patrons i tendències per millorar els resultats de cerca que ofereix als seus usuaris [5] [14] [38]; i, de manera similar, Meta personalitza el contingut que mostra als seus usuaris basant-se en els interessos i preferències individuals [37].

Aquests escenaris comporten que els consumidors se sentin limitats a utilitzar un servei específic en comptes de la seva opció preferida, a causa de la comoditat i adaptabilitat de la plataforma en qüestió. Optant d'aquesta manera per no emprar un servei que els podria proporcionar una major privadesa [14] [36].

Per exemple, els usuaris poden preferir les polítiques de privacitat de DuckDuckGo, però romandre amb el motor de cerca dominant (Google), que, tot beneficiant-se de la seva àmplia xarxa d'usuaris i de la recopilació de dades que efectua, ofereix millors resultats de cerca [38] [36] [39]. De manera semblant, els conductors poden preferir un servei de navegació que prioritzi la confidencialitat, però quedar-se amb l'aplicació dominant de Google, coneguda com a Google Maps [36]; quina finalitat no és només presentar la millor informació de trànsit, sinó que també busca recopilar la quantitat més gran de dades possible.

	Google	Bing	Yahoo!	DuckDuckGo
2010	90.91	3.46	3.93	
2014	89.81	3.63	3.57	0.04
2018	91.4	2.82	2.15	0.29
2022	92.07	3.19	1.36	0.69

Taula 1.1: Participació percentual del mercat de motors de cerca a escala mundial. Font: Elaboració pròpia amb dades de StatCounter[1] i Photutorial[2].

Si ens enfoquem en el sector dels motors de cerca, és evident que Google ocupa una posició predominant en aquesta indústria (consulteu la Taula 1.1) [38] [41]. Aquesta posició es deu a la capacitat de Google d'analitzar les diverses dades que recopila amb els seus productes. Gràcies a aquesta habilitat, Google pot millorar els perfils d'usuari amb informació exclusiva que cap altre competidor posseeix. Aquest avantatge captiva els usuaris i els incentiva a utilitzar els serveis de Google, ja que no hi ha cap altra plataforma que ofereixi una personalització comparable o uns resultats de cerca tan

2. StatCounter. *Search Engine Market Share Worldwide.* Recuperado de https://gs.statcounter.com/search-engine-market-share (Consultado el 29 de Mayo de 2023).

3. Broz, M. (2022, Abril 23). *DuckDuckGo User Stats.* Photutorial. Recuperado de https://photutorial.com/duckduckgo-statistics (Consultado el 29 de Mayo de 2023).

precisos[3]. No obstant això, és important considerar si aquesta raó justifica plenament l'extensió en què Google recopila informació. Hem de comprometre les nostres dades personals per tal d'obtenir resultats més precisos? O és possible trobar un equilibri entre la comoditat i la protecció de dades?

1.4 El valor ocult de les dades.

Partint del fet descrit, s'infereix que la qualitat del producte està estretament relacionada amb la quantitat de dades recopilades; atès que aquestes dades permeten a la companyia oferir informació més precisa.

Tanmateix, a causa de les dades, les empreses són capaces d'identificar patrons, tendències i oportunitats de negoci [5] [14]; fet que els permet generar ingressos mitjançant models publicitaris o serveis addicionals. Per això diverses empreses tecnològiques ofereixin plataformes de comerç, xarxes socials (p. ex. Instagram o Twitter) o navegadors (p. ex. Google o Microsoft Bing) de manera gratuïta, ja que utilitzen les nostres dades com a forma de pagament [36].

> Si no ho estàs pagant,
> no ets el client;
> ets el producte que es ven[4].

Si les dades no tinguessin un valor elevat o no disposessin d'una importància transcendental, resultaria improbable que Google hagués pagat 82 milions de dòlars el 2009 a Apple, per assegurar la seva posició com a motor de cerca predeterminat en Safari [39]. De la mateixa manera, els pagaments de 1000 milions de dòlars que Google va realitzar el 2013 i el 2014 pel mateix propòsit [39] no trobarien cap justificació; i resultaria encara més difícil comprendre el desemborsament de 15000 milions de dòlars[5] que Google va realitzar el 2021 amb el mateix objectiu [40] [41].

Així mateix, en cas que les dades no posseïssin un alt valor, resultaria difícil explicar per què Meta va pagar 16000 milions de dòlars en l'adquisició de WhatsApp, una empresa de només 60 empleats i sense actius tangibles [14]. Encara més, hi hauria poca o cap justificació amb la qual explicar per què Meta va eliminar la petita tarifa de WhatsApp en alguns països[6] [36].

Malgrat que els usuaris (p. ex. de WhatsApp o Google) reben el benefici immediat d'utilitzar un servei gratuït, és important tenir en compte que el cost a llarg o curt termini de divulgar informació personal pot resultar fins i tot més gran que el de pagar una modesta tarifa anual. A causa de la manca de coneixement, per part dels usuaris, sobre com empraran aquestes dades en el futur i per quines entitats seran fetes servir [36].

1.5 Increment d'informació generada.

Segons un informe publicat per IBM el 2013, el 90% de les dades existents fins a aquell moment havien estat generades en els dos anys anteriors [29]; i, en una publicació més recent [11], realitzada el 2020, IBM ha assenyalat que la quantitat de dades generades cada dia se situa al voltant dels 2.5 trilions de bytes[7]. Aquesta gran quantitat de dades ja no es limita únicament a informació bàsica (p. ex., nom, gènere, edat i correu electrònic), sinó que ara també recull una àmplia varietat de dades no estructurades. Entre les quals s'inclouen registres de navegació, dades de transaccions, arxius de correu electrònic, missatges de text, informació *geoespacial*, imatges, contactes, relacions i fins i tot preferències i opinions personals [14]. A la Figura 1.1 es presenta un gràfic aproximatiu que il·lustra el creixement de la informació generada al llarg dels anys.

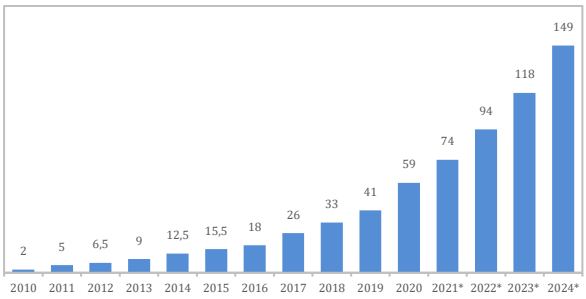


Figura 1.1: Volum de dades generades a escala mundial amb projeccions pel període 2021-2024. Font: Elaboració pròpia amb dades de Statista[8].

A partir d'aquest augment en la quantitat d'informació generada, va sorgir el concepte de *Big Data*. El qual es defineix com el conjunt de dades, tant estructurades com no estructurades, que es distingeixen per quatre característiques fonamentals conegudes com les *quatre Vs*: volum, velocitat, varietat i valor.

En el context del *Big Data*, l'èmfasi no recau tant en la qualitat de les dades, sinó en la quantitat [14]. De manera que totes les companyies que tinguin accés a aquesta quantitat de dades, així com la destresa per analitzar-les, posseiran l'habilitat d'indagar tant en els interessos i inclinacions individuals (de cada persona) com en les tendències i preferències col·lectives (p. ex. d'un país, una ciutat o una comunitat); mitjançant el reconeixement de patrons i formes de comportament.

El fet de conèixer les preferències de cada individu o col·lectiu permet a les companyies influir en les accions i la presa de decisions dels seus usuaris; la finalitat d'influir en el

1. En el sector de los motores de búsqueda, es necesario contar con información relativa a las búsquedas de cada usuario con el fin de mejorar los algoritmos. Cuantos más datos de búsqueda tenga un operador, mejores serán los refinamientos de sus algoritmos [38].

4. Lewis, A. (2010, Agosto 26). *User-driven discontent.* Recuperado de https://www.metafilter.com/95152/Userdriven-discontent#3256046 (Consultado el 8 de Abril de 2023). Traducción propia del inglés al español. Cita original: "If you are not paying for it, you're not the customer; you're the product being sold".

5. Los ingresos netos anuales de Alphabet (la empresa matriz de Google) para 2021 fueron de 76.033 mil millones de dólares [44]. Esto implica que Google pago aproximadamente a Apple una suma equivalente al 19.74% de los ingresos netos de su empresa matriz. Además, es probable que Google realice pagos aún mayores a Apple para asegurarse de que Microsoft no lo supere [40].

6. Antes de la adquisición, WhatsApp solía cobrar a los usuarios una tarifa anual de 0.99 dólares [36].

7. La escala se encuentra expresada en numeración española, lo que corresponde a un valor de 10^{18} bytes o 2.5 exabytes en términos internacionales.

8. Taylor, P. (2021, Junio). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025.* Recuperado de https://www.statista.com/statistics/871513/worldwide-data-created (Consultado el 29 de Mayo de 2023).

nostre comportament pot ser merament econòmica o senzillament democràtica. Conseqüentment, aquesta capacitat comporta una major concentració del poder i una disminució en la diversitat del mercat; fet que resulta en una progressiva limitació de la privacitat i llibertat individual.

1.6 Informació centralitzada.

Tota la informació que les empreses obtenen sobre nosaltres es troba completament centralitzada i resguardada als seus propis servidors. No obstant això, l'emmagatzematge centralitzat d'una quantitat tan gran de dades implica la irrupció de problemes derivats de la centralització, com ara l'eliminació intencionada de dades privades [23], l'ús indegut d'informació confidencial, l'accés no autoritzat a dades personals o la manca de control dels usuaris sobre aquestes.

En la següent secció, presentem el cas de Cambridge Analytica, els esdeveniments del qual ens permetran interioritzar les implicacions negatives d'emmagatzemar aquesta quantitat d'informació en servidors centralitzats, en servidors d'una sola companyia.

1.7 Facebook i Cambridge Analytica.

Escàndols, com el de Cambridge Analytica, han demostrat que hi ha deficiències significatives en els mecanismes de gestió i administració de dades personals [9]. No més lluny, però, aquests esdeveniments també han destacat la importància de revisar i reforçar la protecció de privacitat dels usuaris en l'àmbit digital.

D'acord amb la informació proporcionada per Meta Investor Relations [15], Facebook actualment consta amb més de 2000 milions d'usuaris actius diàriament, el que equival aproximadament a un quart de la població mundial. Quan es treballa amb xifres d'aquesta magnitud, inclús la modificació més petita o la menor transgressió dels drets del consumidor poden ocasionar conseqüències catastròfiques. La rellevància d'aquestes conseqüències augmenta quan es tracta d'informació personal amb caràcter polític, com succeeix amb les dades recopilades per la plataforma Facebook[9].

El 2018, es va formular una acusació contra Cambridge Analytica, una companyia privada especialitzada en anàlisi de dades, per haver utilitzat informació confidencial dels usuaris de Facebook per generar publicitat dirigida en suport a la campanya electoral de Donald Trump[10] [20].

Segons l'informe presentat per Mike Schroepfer [22], que en aquell moment ocupava el càrrec de *Chief Technology Officer* (CTO) de Meta Platforms, Inc., es va revelar que Cambridge Analytica havia obtingut dades potencials de més de 87 milions d'usuaris de Facebook[11]. L'obtenció d'aquestes dades es va realitzar mitjançant l'explotació d'una vulnerabilitat en les interfícies de programació d'aplicacions (APIs) de la plataforma[12]. A més, cal destacar que l'apropiació de les dades en qüestió es va exercir sense el degut consentiment dels usuaris afectats.

Posteriorment, les dades van ser analitzades amb la finalitat de crear perfils psicològics i mostrar anuncis específics per atreure votants. En termes generals, la campanya va llançar 4000 anuncis publicitaris que, en conjunt, van aconseguir més de 1400 milions d'impressions [14]. S'argumenta que els anuncis dirigits elaborats per Cambridge Analytica van ser fonamentals per l'èxit de la campanya electoral de Trump en el 2016, ja que va permetre dirigir-se als votants amb missatges específics que ressonaven amb les seves necessitats més intimes i els seus desitjos més preuats.

El cas exposat demostra que, tot i assumir que Facebook no té gens d'interès polític, la seva funció com a entitat centralitzada i la seva habilitat per recopilar i emmagatzemar grans quantitats de dades, va originar la conseqüència no desitjada de permetre que tercers utilitzessin aquesta informació amb fins maliciosos.

A la següent secció, s'exposarà a Google i la seva gran influència en la societat contemporània, així com el seu ampli abast en la recopilació de dades.

1.8 Google, *don't be evil*[13].

Google és una de les companyies amb més riquesa i influència del món que, a diferència de qualsevol altra, ha aconseguit assolir una posició única de poder [14] [38]. En efecte, Google és el mitjà amb el qual experimentem i comprenem el món que ens envolta. Amb la finalitat de:

> Organitzar la informació del món i
> fer que sigui útil i accessible
> per a tothom[14];

Google ofereix una àmplia gamma de serveis, majoritàriament gratuïts, amb la finalitat no només de presentar la informació d'una manera més organitzada, sinó també de recopilar la quantitat més gran d'informació possible [5]. De fet, el poder d'aquesta empresa es basa únicament en com utilitza la gran quantitat de dades que adquireix dels seus usuaris [4].

9. Si bien los datos que recopila Facebook no son estrictamente políticos, el análisis y el estudio de la información almacenada puede generar datos con relevancia política.

10. Cambridge Analytica generó escándalos en diversos ámbitos, destacando principalmente por su influencia en las elecciones de Estados Unidos. No obstante, es importante resaltar que este incidente no fue un caso aislado, ya que la empresa también participó en campañas electorales de distintos países.

11. A pesar de los informes de medios prominentes como Wired [34], The New York Times [20] y The Observer [33], que indicaban que el conjunto de datos obtenido por Cambridge Analytica contenía información de 50 millones de usuarios de Facebook, así como de las afirmaciones iniciales de Cambridge Analytica de haber recopilado sólo 30 millones de perfiles [1]; Meta Platforms, Inc., confirmó que en realidad habían obtenido datos potenciales de más de 87 millones de usuarios. Entre los cuales un 81.6% (70,6 millones) correspondían a la población de Estados Unidos [22].

12. Específicamente, la recopilación de datos se realizó a través de la aplicación de prueba de personalidad *This Is Your Digital Life*, desarrollada por Aleksandr Kogan en 2013. Alrededor de 300.000 personas instalaron esta aplicación y compartieron sus propios datos, así como algunos de los datos de sus amigos. Dada la forma en que Facebook operaba en ese momento, Kogan pudo acceder a decenas de millones de datos de los compañeros y amigos de los usuarios de *This Is Your Digital Life* [30].

13. Es relevante subrayar que, pese su célebre lema *Don't be evil*, Google se mantiene como una empresa comercial con responsabilidades ante sus accionistas y, por lo tanto, recopila información sobres sus consumidores y lo emplea con el fin de maximizar las ganancias [14]. Si bien su objetivo primordial es brindar servicios útiles y accesibles, resulta imprescindible considerar que su enfoque último se dirige hacia la generación de beneficios económicos.

14. Google. *Sobre Nosotros.* Recuperado de https://about.google/.

Tanmateix, hi ha nombrosos usuaris que neguen o mostren indiferència davant l'ús d'informació personal que duu a terme Google i, en conseqüència, desconeixen el poder d'aquesta empresa. (Si es reflexiona, és probable que ens vingui al cap algú amb aquestes actituds.) Així doncs, malgrat la imatge neutral que Google projecta[15], la seva capacitat per decidir qui accedeix a quina informació i oportunitats, la situa com una entitat de considerable poder i amenaça pel que fa a la privacitat i altres aspectes. Tot i que Google s'esforça per oferir resultats de cerca objectius i precisos, de vegades aquests resultats poden veure's esbiaixats o influïts pels interessos comercials de la companyia.

El 2015, la divisió Buró de Competència de la Comissió Federal de Comerç dels Estats Units (Federal Trade Commission, FTC), encarregada de l'eliminació i prevenció de pràctiques comercials anticompetitives, va publicar de manera accidental parts d'un informe sobre una investigació de Google al Wall Street Journal [4]. En l'informe, la FTC va assenyalar que Google estava afavorint els seus propis productes i adoptant una estratègia anticompetitiva atès que no mostrava certes pàgines web especialitzades en categories altament comercials [14] [38]. Tot seguit, en octubre de 2020, el Subcomitè del Poder Judicial de la Cambra de Representants dels Estats Units sobre Estat Administratiu, Reforma Reguladora i Antimonopoli va publicar un informe centrat en el domini de les quatre grans empreses de dades (Google, Apple, Meta i Amazon); en el qual s'assenyalaven novament les pràctiques anticompetitives de Google [4] [41].

Acciones	Información Compartida
El usuario se prepara por la mañana mientras la música de Google Play Music llena el ambiente.	1. Intereses musicales; gustos y preferencias. 2. Localización. 3. Información del dispositivo utilizado. 4. Patrones de uso: horas del día en que más se utiliza y la duración de las sesiones.
Luego, realiza unas búsquedas rápidas en Google para obtener cierta información.	1. Las consultas de búsqueda realizadas. 2. Localización aproximada durante la realización de las búsquedas. 3. Información del dispositivo utilizado. 4. Etc. (Apéndice A.2.)
Revisa su calendario y establece un evento en Google Calendar.	1. Detalles sobre los eventos y recordatorios. 2. Información sobre la organización de su agenda y programación diaria.
El usuario interactúa con Google Home y emplea un comando de voz para apagar la música.	1. Historial de comandos de voz utilizados y preguntas realizadas. 2. Información sobre el momento del día y el lugar en que la interacción fue realizada.
Por último, el usuario disfruta de algunos videos o shorts en YouTube antes de salir de casa y comenzar el día.	1. Historial de reproducción de videos. 2. Preferencias de contenido. 3. Duración de visualización de los videos y patrones de consumo. 4. Etc. (Apéndice A.3.)

Taula 1.2: Informació compartida amb Google en la matinada d'un *dia de cada dia*[16]. Font: Elaboració pròpia amb inspiració de [31].

La informació que proporcionem a companyies tecnològiques, com Google i Meta, és extremadament personal; arribant fins i tot a casos en què les companyies coneixen més atributs de la nostra pròpia personalitat que tots els nostres amics junts. En l'informe de Schmidt [31], es detallen diversos experiments que evidencien l'alarmant abast de la recopilació de dades que efectua Google. La companyia recopila informació cada vegada que els usuaris interactuen amb alguna de les seves plataformes (p. ex., Chrome i Android[17]), aplicacions (com Google Maps, YouTube i Gmail), eines per a editors (p. ex., Google Analytics i AdSense) i altres programes; cosa que resulta en un coneixement excessiu d'informació personal. (Consulteu l'Apèndix A per obtenir una perspectiva més àmplia de la *data-opoly*[18] de Google.)

Els aspectes més confidencials i íntims de la informació emmagatzemada per Google exemplifiquen la naturalesa única de les dades com a producte de comoditat diferent de qualsevol altra mercaderia controlada per entitats centralitzades [14]. Les dades, a excepció de les altres mercaderies, poden ser molt valuoses per a l'elaboració de perfils massius, utilitzats per comprendre les preferències i tendències col·lectives, així com per a la creació de perfils individuals, utilitzats per interioritzar les preferències i inclinacions personals [5]. (Vegeu la Taula 1.2.)

En acceptar els termes de servei, els usuaris accepten compartir la majoria de les seves dades amb Google, independentment de la privacitat o confidencialitat d'aquestes. (Vegeu l'Apèndix A per a una descripció detallada de la informació compartida.) Aquest conjunt d'informació es processa fent ús de diverses tècniques d'anàlisi de dades, amb la finalitat de generar perfils d'usuaris a escala individual i massiva [10].

Si bé la recopilació i el processament de dades personals són pràctiques necessàries per millorar la qualitat i personalització dels serveis, és essencial considerar els diversos riscos de seguretat associats amb l'intercanvi d'informació personal. Tanmateix, malgrat les nombroses implicacions negatives, la majoria d'usuaris no només estan d'acord, sinó que fins i tot desitjosos de compartir les seves dades confidencials amb la finalitat de rebre serveis més personalitzats i integrats [5]. De fet, els consumidors estan regalant la seva informació sota un fals ideal de comoditat i cedint els seus drets de privacitat en benefici d'un servei més personalitzat i individualitzat [5]. Això fa que els usuaris perdin la sobirania de les seves dades i quedin estrictament subjectes al poder d'aquells que les posseeixen; les companyies.

1.9 TikTok.

[...]

1.A Plataformes a les ombres.

[...]

15. La imagen neutral con la que Google es presentada, puede contribuir a la despreocupación que los usuarios muestran en relación al uso que dicha empresa da a sus datos.

16. En esta tabla se presentan unas pocas interacciones que un usuario cualquiera podría realizar, así como alguna de la información que Google podría recopilar.

17. Véase usted la Figura A.1.

18. El control prácticamente total que ejerce un pequeño grupo de empresas sobre la recopilación, el almacenamiento y el acceso de datos personales se denomina *data*-opoly [39] [36].

1.B *Data Network Effect.*

Les tecnologies digitals han avançat de manera contínua i progressiva cap a una arquitectura centralitzada que compromet els drets dels consumidors i posa en perill la confidencialitat de la informació [5]. Els casos de TikTok, Google i Facebook evidencien el control monopolitzat que les empreses tecnològiques exerceixen sobre la informació personal, el coneixement col·lectiu i els mitjans de comunicació. A més a més, aquests tres casos resulten ser exemples clars que il·lustren les possibles implicacions negatives de la centralització de les dades, o en altres paraules, del poder.

1. En el cas de Facebook, s'ha criticat la recopilació massiva de dades personals i el seu potencial impacte en les eleccions dels usuaris. A més a més, s'ha assenyalat que, tot i assumir que aquestes empreses no tenen altres motivacions més enllà de les econòmiques per analitzar i utilitzar les nostres dades, la seva estructura centralitzada i la seva capacitat per recopilar i emmagatzemar grans volums d'informació ha generat la possibilitat que tercers facin un ús indegut d'aquesta informació.
2. En referència a Google, s'ha observat amb preocupació l'extensa recopilació de dades que realitza l'empresa amb totes les seves plataformes; ja que aquestes no es limiten únicament a proporcionar informació, sinó que també tenen com a objectiu recopilar la quantitat més gran de dades possible.
3. [...].

L'aparició de grans potències digitals com Google i Meta, o l'establiment d'importants serveis en línia com TikTok, es deu principalment a un efecte de xarxa basat en les dades personals, conegut en anglès com a *data network effect* [14] [38] [39]. Els efectes de xarxa són tals que com més usuaris hi ha en una plataforma, més valuosa esdevé aquesta per a cada usuari [5] [35] [36].

Els telèfons constitueixen un exemple clàssic. A mesura que més persones adquireixen dispositius telefònics, més persones estaran disponibles per realitzar trucades. Així mateix, tenir un major nombre de persones a les quals trucar augmenta el valor inherent de posseir un telèfon [36].

En el context de les dades, en canvi, a mesura que la plataforma atreu un major nombre d'usuaris i acumula més informació, la seva habilitat per personalitzar el contingut s'intensifica i la comoditat de l'usuari augmenta. Aquest aspecte fomenta el retorn de l'usuari i estimula la voluntat de compartir informació personal (Figura 1.3). A més a més, ja que ningú pot oferir serveis tan individualitzats, el negoci de l'empresa es consolida en el mercat.

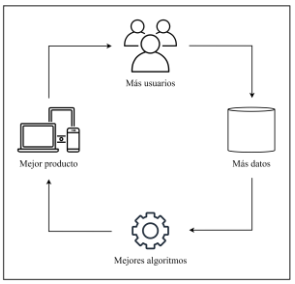


Figura 1.3: Efecte de xarxa basat en les dades.

Prenem com a exemple Facebook. Aquesta plataforma recull informació dels seus usuaris per generar béns i serveis personalitzats [36], amb l'objectiu últim d'augmentar la quantitat d'usuaris en la xarxa i la seva dependència d'ella, de manera que s'aconsegueixi bloquejar el nombre més gros d'usuaris en el sistema [5]. En efecte, a mesura que augmenta la quantitat d'usuaris en la xarxa i la plataforma obté més dades sobre nosaltres, es genera un major nivell de comoditat i una major dependència respecte al servei; la qual cosa contribueix al creixement i la consolidació de Facebook en el mercat de les xarxes socials. Aquest mateix fenomen també es pot aplicar als motors de cerca [36], com Google Search o Microsoft Edge, així com a la computació en el núvol [5], com AWS o IBM Cloud, entre altres exemples.

1.C Ingenuïtat humana.

La gran capacitat econòmica que posseeixen les dades és la que motiva les empreses a recopilar la quantitat més gran d'informació possible sobre els seus consumidors. Com a resultat, aquesta necessitat d'informació s'ha convertit en la principal lluita pels líders tecnològics [32].

Per culpa de l'objectiu de les empreses de saber tant com sigui possible dels seus usuaris, no importa quan esforç faci algú per mantenir la seva informació en privat, hi ha mecanismes i eines arreu que recopilen dades personals i les comuniquen a empreses terceres, ja sigui amb el consentiment de l'individu o sense [5]. Més sovint, però, són els mateixos usuaris els que desitgen compartir informació personal amb una àmplia varietat d'empreses interessades. La majoria dels usuaris proporcionen informació a empreses sense conèixer les implicacions de privacitat d'aquests serveis o valoren el servei i la comoditat per sobre del seu risc personal [5].

Per tant, és el desig de comoditat de l'ésser humà i la manca de coneixement i indiferència dels usuaris respecte a la recopilació d'informació el que dota a les empreses d'un poder inconcebible, que mai s'havia imaginat ni molt menys concentrat.

2

[...]

~ [...] ~

2.1 Secció.

En aquesta secció parlarem de les solucions tant tecnològiques com [...].

2.2 De camí a les solucions.

En utilitzar aplicacions en línia, els usuaris solen compartir una abundant quantitat d'informació personal, ja sigui en fer compres a través d'Internet o en utilitzar targetes de crèdit[19]; en accedir a enllaços i articles, o en publicar actualitzacions personals; en qualificar sèries i pel·lícules, o en comentar publicacions; la informació sempre es comparteix dins d'un àmbit particular. No obstant això, quan una part de la informació es mou més enllà del seu abast previst (ja sigui perquè es manté més temps de l'establert, es comparteix amb corporacions no autoritzades o s'abusa de la informació per a un propòsit diferent del pactat), es produeix una violació de la privacitat [43].

Actualment, resulta complicat conèixer amb certesa l'ús que se li està donant a les nostres dades. Però encara resulta més difícil preveure com i per qui seran utilitzades en el futur [36]. Arribant a un punt en el qual els mateixos usuaris, que són els legítims propietaris de la seva identitat[20], deixen de ser els veritables propietaris d'aquesta i cedeixen el poder a les empreses. L'apropiació del control de les nostres dades personals per part d'un reduït conjunt d'empreses ens priva i ens expropia de la facultat de gestionar la nostra informació confidencial; suspenent-nos del nostre legítim dret de propietat sobre les dades que conformen la nostra identitat. En canvi, les empreses adquireixen un domini absolut o parcial sobre les nostres dades.

Si bé l'Internet va ser dissenyat com un sistema descentralitzat per maximitzar la resistència i eliminar la possibilitat d'un únic punt de fallida [5], la proliferació de companyies (com Facebook o Google) que constantment desenvolupen tecnologies capaces d'emmagatzemar i recopilar grans quantitats d'informació personal ha transformat a Internet en un entorn centralitzat. On la nostra informació s'ha convertit en una mera moneda d'intercanvi.

> A mesura que l'Internet ha evolucionat, hem passat de l'era en què ningú sabia si eres un gos[21], a l'era en què les companyies no només saben que ets un gos, sinó que fins i tot coneixen els teus patrons de respiració i volen utilitzar aquesta informació per mostrar-te anuncis personalitzats.

Per tant, l'avanç de l'economia digital i el *Big Data* ha generat inquietuds entre certs usuaris que temen perdre el control sobre la manera en com es recopilen i utilitzen les seves dades. Tanmateix, altres, en prendre consciència que aquesta situació ja és una realitat, comencen a buscar alternatives per contrarestar aquestes preocupacions. En la literatura relacionada, tant empreses com investigadors han proposat diverses *solucions* per restituir la sobirania que es mereixen els usuaris. Entre aquestes *solucions*, les que destaquen són les jurídiques i les tecnològiques.

2.3 Solucions jurídiques per a la protecció de dades.

La manca d'un marc regulador sòlid que promogui la transparència i el control del consumidor sobre les seves pròpies dades pot generar series implicacions que afectin el bon funcionament dels mercats digitals [9] [14]. En resposta, s'han realitzat esforços a escala mundial per actualitzar i establir noves lleis que abordin els reptes relacionats amb la privacitat de la informació. Exemples destacats inclouen el Reglament General de Protecció de Dades (RGPD) [46] de la Unió Europea, que va entrar en vigor el 25 de maig de 2018, i la Llei de Privadesa del Consumidor de Califòrnia (CCPA), que es va implementar l'1 de gener de 2020.

La introducció del RGPD ha marcat l'inici d'una sèrie de canvis en la forma en què es gestionen les dades personals dels ciutadans europeus. Des d'una perspectiva jurídica, l'objectiu del RGPD és dotar el marc legal amb les garanties adequades que permetin el control individual sobre les dades personals [9]. Aquest principi de control es reflecteix en un conjunt de drets; els més destacats són:

1. Dret de Supressió o Dret a l'Oblit: Els usuaris tenen el dret d'obtenir la supressió de les dades que els afectin. Això significa que les empreses han d'eliminar o anonimitzar les dades d'un usuari quan ja no siguin necessàries en relació amb els fins per als quals van ser recollides o quan l'interessat retiri el seu consentiment [45] (Art. 17, [46]).
2. Requisit del Consentiment Informat: Cada client ha de ser informat en termes senzills sobre la finalitat de les dades que proporciona [45] (paràgraf 42-43, [46]). A més, es requereix que els clients autoritzin prèviament l'ús de les seves dades [14] [45] (Art. 6, paràgraf 1, [46]).
3. Dret a la Notificació Immediata de Violacions de Privadesa: En cas de violació de la seguretat, el responsable del tractament ha d'informar als usuaris afectats de manera immediata i sense demores indegudes, com a molt tardà 72 hores després que s'hagi tingut constància de l'esdeveniment [45] (Art. 33, [46]).
4. Dret a la Portabilitat de les Dades: Els clients tenen dret a rebre les dades conservades sobre ells en un format estructurat, d'ús comú i lectura mecànica; i tenen dret a transmetre-les a altres organitzacions i responsables [45] [14] (Art. 20, [46]).

17. Al comprar en Internet, es fácil de entender que recopilen información sobre nosotros, pero también es importante destacar que en cualquier transacción monetaria con dinero digital es posible analizar y rastrear la actividad.
19. En este contexto, la identidad se refiere al conjunto de datos que nos definen, incluyendo toda la información que proporciona una visión clara de quiénes somos, nuestra personalidad y nuestros gustos.

19. Steiner, P. (1993, Julio 5). *On the Internet, nobody knows you're a dog.* The New Yorker.

Els drets esmentats marquen fites significatives que han estat possibles gràcies als esforços encomiables de diverses entitats i individus compromesos en la creació i promoció del RGPD i altres regulacions similars. Resulta sorprenent constatar com aquestes legislacions representen un petit avanç per a la protecció de la privadesa i un gran pas cap a la sobirania de les dades.

No obstant això, també és important reconèixer els reptes que limiten la seva capacitat per resoldre de manera completa la manca de control que experimenten els usuaris. Tot i que aquestes regulacions ofereixen un marc regulador sòlid i estableixen drets fonamentals per a tot consumidor, encara hi ha certes limitacions que la llei no pot abordar per si mateixa.

2.4 Confiança en abundància.

El simple fet d'haver de confiar en els proveïdors de serveis o en el responsable de l'emmagatzematge planteja un desafiament en si mateix.

En l'actualitat, els usuaris dipositen molta confiança implícita en els proveïdors de serveis, esperant que gestionin la seva informació de manera justa i conscient; respectant els drets del consumidor i aplicant mesures tècniques de seguretat per garantir la confidencialitat i la privadesa de les dades; i que continuïn fent-ho en el futur [43]. En utilitzar el sistema, els usuaris estableixen una relació amb el proveïdor de serveis, qui (a causa de la seva forma de funcionar) pot veure tota la informació en el sistema, incloses les descàrregues privades, les sol·licituds d'accés a informació, el comportament de compra i navegació, etc.[22] Per tant, correspon al mateix proveïdor assegurar-se que no s'empra aquesta informació sense el consentiment corresponent. A més, és el proveïdor de serveis qui té l'autoritat final per decidir quina informació s'emmagatzema, quant temps es conserva i com s'usa o es distribueix [43]; no és l'usuari qui decideix de forma directa.

Si bé els reglaments com el RGPD o la CCPA estableixen mesures estrictes per a aquests aspectes, és responsabilitat del proveïdor decidir si compleix o no amb aquestes mateixes normes i acceptar les conseqüències associades a la seva decisió. A més a més, poden sorgir situacions en què la responsabilitat no recaigui directament en l'ètica de l'empresa, sinó que més aviat es deu a una mala gestió o manca de recursos tècnics per abordar adequadament aquests aspectes. Com ja s'ha presentat a la Secció 1.5, el cas de Facebook i Cambridge Analytica no va ser ocasionat per una mala ètica per part de Facebook, sinó que es va produir a causa d'una mala gestió momentània.

Generalment, les declaracions de privacitat es proporcionen per mostrar la postura adoptada pel proveïdor de serveis i obtenir el consentiment de l'usuari. No obstant això, molts usuaris es veuen forçats a acceptar els termes i condicions d'ús; perquè, en cas contrari, el servei els denegaria l'accés. Limitant el dret dels usuaris a seleccionar quina informació volen compartir i quines prefereixen mantenir privada [36].

> En la gran majoria, no hi ha terme mitjà; es tracta del "tómalo o déjalo" de les negociacions tradicionals: o acceptes l'ús que se'ls donarà a les teves dades o prescindeixes del servei.

L'equilibri de poder està clarament a favor del proveïdor de serveis [43], i no es pot simplement assumir que els usuaris consenten aquesta forma de recopilació i processament de dades [36]. Degut, en part, a la posició dominant del servei. El RGPD i altres regulacions similars obligaran les empreses a detallar millor la seva política de privadesa. No obstant això, per als qui acceptin la política proposada, aquestes empreses continuaran recopilant dades de manera extensa.

Tot i que les solucions jurídiques són efectives per establir un cert nivell de privacitat, encara pateixen la debilitat inherent dels models basats en la confiança. Per això, és necessari establir un sistema que garanteixi la confidencialitat de la informació a través d'un model basat en proves criptogràfiques en lloc de confiança. A més, a diferència de les lleis que generalment s'empren per resoldre problemes després que sorgeixin, un enfocament basat en la tecnologia serà capaç de prevenir les violacions de privacitat abans i tot de què succeeixin [43].

2.5 Solucions tecnològiques.

Les solucions tecnològiques ofereixen una major garantia de confidencialitat de la informació, ja que reemplacen la dependència de la confiança amb mecanismes matemàtics. En general, es reconeix que la tecnologia de xifratge és el mètode tècnic més simple i eficaç per protegir les dades dels usuaris [47] [48] [53].

La criptografia és un camp especialitzat que se centra a protegir la confidencialitat i seguretat de la informació [47], mitjançant l'ús de tècniques de codificació i descodificació. Aquestes tècniques asseguren que la informació roman inaccessible per a aquells sense la clau de xifratge adequada[23].

En criptografia, es distingeixen entre els esquemes de clau secreta, també coneguts com a criptografia simètrica, i els de criptografia de clau pública, també coneguts com a criptografia asimètrica. En els esquemes de criptografia simètrica, tant el xifratge com el desxifratge (d'un missatge específic, al qual anomenem m) es realitzen utilitzant la mateixa clau[24], coneguda com a k.

$$C = Enc(k, m) \tag{1}$$

$$m = Dec(k, C) \tag{2}$$

En l'Equació 1, la funció Enc representa la funció de xifratge, que pren una clau i un missatge sense xifrar com a paràmetres i retorna un text xifrat, conegut com a *ciphertext*. D'altra banda, en l'Equació 2, la funció Dec representa la funció de desxifratge, que utilitza la mateixa clau que s'ha utilitzat per

20. Aunque el proveedor del servicio asegure que no recopilará información al respecto, todavía tiene la capacidad de observar las interacciones que usted realiza con su servidor. Todos los datos que envíe para su procesamiento a través del servidor, así como todas las interacciones que realice con su base de datos, serán visibles para dicho proveedor. Este concepto se explica con mayor precisión en la Sección 2.6.

21. De acuerdo con el principio de Kerckoff, la seguridad del protocolo no debería depender de la opacidad del código, sino únicamente del nivel de secreto de la clave de descifrado [48].

22. Dado que se utiliza la misma llave tanto para cifrar como para descifrar, es necesario mantenerla en secreto o compartirla solo con aquellas personas con las que deseamos compartir información.

xifrar el missatge per tal de recuperar-ho. Com que es fa servir la mateixa clau, és necessari que l'emissor i el receptor acordin prèviament la clau que utilitzaran per establir qualsevol comunicació segura. Això implica que aquests esquemes no puguin ser emprats per dues persones que mai s'hagin conegut prèviament [48]. A més, en aquest sistema es requereix compartir una clau diferent amb cada individu amb qui volem comunicar-nos [48]. No obstant això, els esquemes simètrics presenten l'avantatge de ser realment ràpids i se'n fa us amb la freqüència més gran possible [48].

En contrast amb el model anterior, els esquemes asimètrics utilitzen un parell de claus; una de les quals, anomenada clau pública (representada com a k_p), s'utilitza per xifrar, mentre que l'altra, la clau privada (representada amb el subíndex s, k_s), es manté en secret i s'utilitza per desxifrar el *ciphertext*. Quan es desitja enviar un missatge xifrat, l'emissor fa ús de la clau pública del receptor per xifrar-lo. Després, el receptor farà servir la seva clau privada per desxifrar el *ciphertext* rebut i obtenir el missatge original.

$$C = Enc(k_p, m) \tag{3}$$

$$m = Dec(k_s, C) \tag{4}$$

L'Equació 3 xifra el missatge utilitzant la clau pública i l'Equació 4 el desxifra amb la clau secreta. Els esquemes asimètrics es consideren més flexibles que els simètrics, ja que no requereixen que l'emissor i el receptor acordin prèviament cap clau [48]. Tanmateix, aquests esquemes solen ser més lents que els simètrics [48]. Exemples destacats són el RSA i l'ElGamal.

2.6 Problemes relacionats amb la criptografia.

Els sistemes més coneguts d'encriptació depenen de compartir una clau, sigui pública o privada, entre els individus involucrats en l'intercanvi d'informació [49]. Ara bé, aquest enfocament planteja alguns problemes relacionats amb la privacitat. Els usuaris o proveïdors de serveis amb accés a la clau tenen drets exclusius sobre les dades [49]. Això implica que tenen el control i la capacitat d'accedir, utilitzar i gestionar les dades personals de manera exclusiva. Especialment en el cas dels serveis en línia, existeix el risc de perdre el control sobre la confidencialitat de la informació.

Així doncs, per tal de garantir que només els usuaris legítims puguin accedir de forma segura a les seves dades, és necessari xifrar la informació abans de transmetre-la i abstenir-se de compartir qualsevol clau secreta que permeti el seu desxifratge. No obstant això, apareix un inconvenient, perquè tot i que el xifratge convencional pot ser eficaç per evitar l'accés no autoritzat, també implica la destrucció de l'estructura semàntica subjacent de les dades; cosa que fa impossible efectuar operacions sobre elles [47]. En conseqüència, els resultats de les operacions sobre textos xifrats manquen de significat quan s'empren mètodes de xifratge tradicionals [47].

Per culpa de l'esmentat, es requereix que en qualsevol dels casos els serveis en línia hagin de desxifrar el missatge abans de fer operacions sobre ells, ja que si no, el resultat no tindrà cap significat; posant en compromís el nivell de privadesa que el servidor pot oferir als seus consumidors.

Quan es vol accedir a informació emmagatzemada en una base de dades, per exemple, l'usuari pot enviar de manera encriptada l'índex on es troba aquesta informació. No obstant això, una vegada arriba al servidor, aquest haurà de desxifrar-ho i llegir-ho per poder respondre amb la informació correcta, ja sigui amb el contingut d'una revista o amb la informació d'alguna medicina. (Li recomano que vegi la Figura 2.1.) En aquest cas, existeix la possibilitat que un operador curiós segueixi les consultes de l'usuari i recol·lecti les cerques que efectua [50]; independentment del nivell de confidencialitat de la consulta realitzada o de si s'està utilitzant una pestanya d'*incògnit* o no[25].

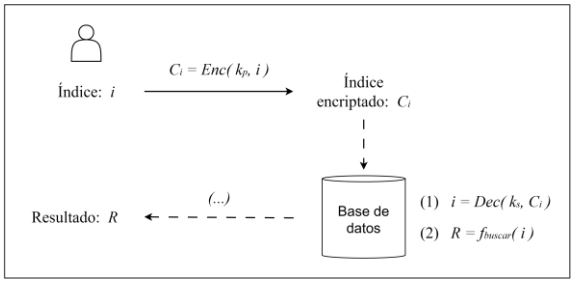


Figura 2.1: Sol·licitud d'informació a una base de dades utilitzant criptografia tradicional[26].

En aquest context i altres similars, els consumidors poden desitjar preservar la confidencialitat de les dades proporcionades [53]. Per exemple, un usuari pot desitjar mantenir la privacitat de la seva sol·licitud i encara així rebre el contingut sol·licitat. (Vegeu la Figura 2.2.) Això ens porta a la necessitat d'un sistema criptogràfic que respecti la confidencialitat de les dades, no només durant la comunicació i l'emmagatzematge, sinó també durant el seu processament [53]. Per descomptat, el resultat del processament ha de ser igual de bo que si les dades no estiguessin encriptades [53] [52] [48]. De fet, el resultat desxifrat hauria de ser equivalent al resultat calculat en text sense encriptar.

24. De hecho, en los casos en que las intenciones de los usuarios deben mantenerse en secreto, los usuarios a menudo son cautelosos al acceder a bases de datos [50]. Este es un concepto interesante que se tratara con en minucia en su propia sección, la Sección 4.

25. En esta la Figura 2.1, el usuario busca acceder a la información almacenada en el índice i. Antes de enviar i, sin embargo, cifra su valor utilizando la calve publica (k_p). Generando así el *ciphertext* de i, anotado como C_i. Una vez que el servidor reciba C_i, este lo descifrará utilizando la clave secreta (k_s) e implementara un algoritmo de búsqueda (f_{buscar}) para encontrar la información almacenada en ese índice (R). Una vez obtenida, se enviara al usuario. Dado que el servidor debe descifrar C_i para poder aplicar la función f_{buscar}, la confidencialidad de la solicitud, la privacidad del índice i, se pierde por completo.

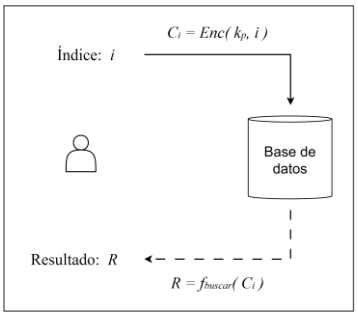


Figura 2.2: Sol·licitud d'informació a una base de dades sense desxifrar l'índex en qüestió[27].

Així doncs, aconseguir la coherència en els resultats de les operacions executades sobre textos xifrats representa un nou desafiament. Mentre que descobrir la solució es converteix en l'objectiu, i la resposta es transforma en el que molts criptògrafs consideren el *Sant Grial de la criptografia* [53] [52].

2.6 El Sant Grial de la criptografia.

L'any 1978, els criptògrafs Rivest, Adleman i Dertouzos van proposar una solució innovadora per abordar el desafiament de la privacitat durant el processament de dades [54]. La seva proposta, coneguda avui en dia com a Xifratge Homomòrfic (*Homomorphic Encryption* o HE en anglès), permetia realitzar càlculs sobre dades xifrades sense necessitat de desxifrar-les. El resultat de les operacions es retorna com un resultat xifrat, el qual, una vegada desxifrat, obté el mateix valor que si els càlculs s'haguessin efectuat sobre *plaintext*.

Des d'aquest article seminal, el disseny de sistemes eficients i esquemes segurs d'encriptació homomòrfica ha estat un dels sants grials de la comunitat criptogràfica.

27. En la Figura 2.2, el usuario pretende acceder al índice i de una base de datos. No obstante, antes de enviar el índice al servidor, el usuario lo encripta utilizando un método especial de criptografía. Una vez que el servidor obtiene el *ciphertext* de i (representado como C_i), se aplicará la función de búsqueda (f_{buscar}) sobre el mismo, sin necesidad de descifrarlo. Luego, f_{buscar} devuelve un resultado cifrado (C_R) que solo el usuario es capaz de descifrar, $Dec(k_s, C_R)$, y visualizar su contenido. Dado que el servidor ha podido buscar en la base de datos el índice i sin siquiera descifrarlo y, por tanto, sin necesidad de conocer la clave secreta del usuario; el servidor no ha adquirido información alguna sobre la solicitud ni la respuesta. Si esto parece contradictorio o un tanto inusual, le sugiero esperar a leer la Sección 4 para comprender el funcionamiento de este mecanismo.

3

Homomorphic Encryption

~ Operaciones aritméticas en *ciphertexts* ~

2.1 Sección.
Explicar Homomorphic Encryption (con base matemática)

2.6 Aplicaciones.
A partir de ahora nos dirigiremos a un caso específico: el acceso a información de manera privada, private information retrieval.

4

Private Information Retrieval

~ Privacidad por diseño ~

PIR:
At present, the significant ways to ensure database privacy data security include firewalls, identity verification, and auditing. However, leakage of sensitive database information still occurs frequently, showing that these security measures are minimal to solve the problem. Ensuring that legitimate users can safely access their data, the most direct way is to encrypt the data before it is stored in the database. It is generally acknowledged that data encryption technology is the simplest and most capable technical method to protect users' data from illegal access, used in many fields widely such as the Internet, electronic communications, online shopping, and online banking. Although data encryption can effectively avoid unlawful access to data, it also destroys the underlying semantic structure of data, making it impossible to perform operations such as calculation and retrieval of ciphertexts. Consequently, the results of ciphertexts operations make little sense by using traditional encryption methods, making ciphertexts retrieval become a new challenge

(Podríamos utilizar las leyes de Cameron para evaluar nuestro sistema)

Existe un riesgo de seguridad inherente en el uso de Internet para transferir información confidencial y datos personales. Por regla general, la información quiere ser compartida, y la mayor parte del valor que se puede extraer de ella surge del uso y comunicación de la misma. Sin embargo, cada vez que se publica en Internet, se pone necesariamente en riesgo la privacidad y confidencialidad de la información. [5]

One could further argue that the Sherman Act's term

Agraïments

Referencias

[1] Aiello, C. (2018, Abril 4). CNBC. *Cambridge Analytica says no more than 30 million people impacted by leak.* Recuperado de https://www.cnbc.com/2018/04/04/cambridge-analytica-says-no-more-than-30-million-people-impacted.html (Consultado el 20 de Mayo de 2023).

[2] Buterin, V. (2014). *Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform*. Recuperado de https://ethereum.org/669c9e2e2027310b6b3cdce6e1c52962/Ethereum_Whitepaper_-_Buterin_2014.pdf (Consultado el 23 de Marzo de 2023).

[3] Cameron, K. (2005, Mayo 11). *The Laws of Identity*. Recuperado de https://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf (Consultado el 26 de Marzo de 2023).

[4] Case, M. (2021, Agosto 15). *Google, Big Data, & Antitrust*. Recuperado de https://ssrn.com/abstract=3917218 (Consultado el 14 de Abril de 2023).

[5] De Filippi, P., & McCarthy, S. (2012, Octubre 26). *Cloud Computing: Centralization and Data Sovereignty.* Recuperado de https://ssrn.com/abstract=2167372 (Consultado el 16 de Abril de 2023).

[6] De Filippi, P. (2014, Agosto 25). *Ethereum: the decentralised platform that might displace today's institutions*. Recuperado de https://policyreview.info/comment/99 (Consultado el 5 de Abril de 2023).

[7] Dunphy, P., & Petitcolas, F. a. P. (2018, Agosto 6). *A First Look at Identity Management Schemes on the Blockchain*. Recuperado de https://doi.org/10.1109/msp.2018.3111247 (Consultado el 24 de Marzo de 2023).

[8] Folkenflik, D. (2017, Setiembre 26). *Facebook Scrutinized Over Its Role In 2016's Presidential Election*. Recuperado de https://www.npr.org/2017/09/26/553661942/facebook-scrutinized-over-its-role-in-2016s-presidential-election (Consultado el 7 de Abril de 2023).

[9] Giannopoulou, A. (2020, Septiembre 28). *Data Protection Compliance Challenges for Self-Sovereign Identity*. Recuperado de http://dx.doi.org/10.2139/ssrn.3671523

[10] Google – Privacy & Terms (2022, Diciembre 15). *Google Privacy Policy.* Recuperado de https://policies.google.com/privacy (Consultado el 29 de Abril de 2023).

[11] IBM Research (2020, Diciembre 16). *5 Things to Know About IBM's New Tape Storage World Record*. Recuperado de https://newsroom.ibm.com/IBM-research?item=32682# (Consultado el 8 de Abril de 2023).

[13] Lundkvist, C., & Heck, R., & Torstensson, J., & Mitton, Z., & Sena, M. (2017, Febrero 21). *uPort: A Platform for Self-Sovereign Identity*. Recuperado de https://whitepaper.uport.me/uPort_whitepaper_DRAFT20170221.pdf (Consultado el 4 de Abril de 2023).

[14] McIntosh, D. (2019, Enero). *We Need to Talk About Data: How Digital Monopolies Arise and Why They Have Power and Influence*. Recuperado de https://scholarship.law.ufl.edu/jtlp/vol23/iss2/2/

[15] Meta Investor Relations (2023, Febrero 2). *FORM 10-K*. Recuperado de https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/e574646c-c642-42d9-9229-3892b13aabfb.pdf (Consultado el 9 de Abril de 2023).

[16] Minkley, J. (2010, Marzo 31). *Nintendo's Shigeru Miyamoto In pursuit of happiness*. Recuperado de https://www.eurogamer.net/shigeru-miyamoto-interview?page=2 (Consultado el 5 de Abril de 2023).

[17] Mossberger, K., & Tolbert, C. J., & McNeal, R. S. (2007). *Digital Citizenship: The Internet, Society, and Participation*. Recuperado de https://doi.org/10.7551/mitpress/7428.001.0001

[18] Naik, N., & Jenkins, P. (2016, Octubre 13). *An Analysis of Open Standard Identity Protocols in Cloud Computing Security Paradigm*. Recuperado de https://doi.org/10.1109/dasc-picom-datacom-cyberscitec.20 16.85

[19] Nakamoto, S. (2008, Octubre 31). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Recuperado de https://bitcoin.org/bitcoin.pdf (Consultado el 21 de Marzo de 2023).

[32] Pariser, E. (2011, Mayo 12). *The Filter Bubble: What The Internet Is Hiding From You.* Recuperado de https://dl.acm.org/doi/book/10.5555/2029079

[20] Rosenberg, M., & Confessore, N., & Cadwalladr, C. (2018, Marzo 17). *How Trump Consultants Exploited the Facebook Data of Millions*. The New York Times. Recuperado de https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html (Consultado el 11 de Abril de 2023).

[21] Schardong, F., & Custódio, R. (2022, Julio 28). *Self-Sovereign Identity: A Systematic Review, Mapping and Taxonomy*. Recuperado de https://doi.org/10.3390/s22155641

[31] Schmidt, D. C. (2018, Agosto 15). *Google Data Collection.* Recuperado de https://static.poder360.com.br/2018/08/DCN-Google-Data-Collection-Paper.pdf (Consultado el 20 de Mayo de 2023).

[22] Schroepfer, M. (2018, Abril 4). *An Update on Our Plans to Restrict Data Access on Facebook*. Recuperado de https://about.fb.com/news/2018/04/restricting-data-access (Consultado el 11 de Abril de 2023).

[23] Shrestha, A. K., & Vassileva, J., & Deters, R. (2020, Octubre 22). *A Blockchain Platform for User Data Sharing Ensuring User Control and Incentives*. Recuperado de https://doi.org/10.3389/fbloc.2020.497985

[24] Turck, M. (2016, Enero 4). *The Power of Data Network Effects.* Recuperado de https://mattturck.com/the-power-of-data-network-effects (Consultado el 29 de Abril de 2023).

[25] Wang, F., & De Filippi, P. (2020, Enero 24). *Self-Sovereign Identity in a Globalized World: Credentials-Based Identity Systems as a Driver for Economic Inclusion*. Recuperado de https://doi.org/10.3389/fbloc.2019.00028

[26] Weyl, E. G., & Ohlhaver, P., & Buterin, V. (2022, Mayo 11). *Decentralized Society: Finding Web3's Soul*. Recuperado de https://dx.doi.org/10.2139/ssrn.4105763

[27] Windley, P. J. (2021, Julio 28). *Sovrin: An Identity Metasystem for Self-Sovereign Identity*. Recuperado de https://doi.org/10.3389/fbloc.2021.626726

[28] Wood, G. (2022, Octubre 24). *Ethereum: A Secure Decentralised Generalised Transaction Ledger*. BERLIN VERSION beacfbd. Recuperado de https://ethereum.github.io/yellowpaper/paper.pdf

[29] Zhan, Y., & Tan, K. H., & Li, Y., & Tse, Y. K. (2018, Noviembre). *Unlocking the power of big data in new product development*. Recuperado de https://doi.org/10.1007/s10479-016-2379-x

[30] Zuckerberg, M. (2018, Marzo 21). *I want to share an update on the Cambridge Analytica situation -- including the steps we've already taken and our (...).* Facebook. Recuperado de https://www.facebook.com/zuck/posts/10104712037900071

[34] Lapowsky, I. (2018, Marzo 17). *Cambridge Analytica Took 50M Facebook Users' Data—And Both Companies Owe Answers.* Wired. Recuperado de https://www.wired.com/story/cambridge-analytica-50m-facebook -users-data

[33] Cadwalladr, C., & Graham-Harrison E. (2018, Marzo 17). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.* The Guardian. Recuperado de https://www.theguardian.com/ news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[35] Yoo, C. S. (2012, Enero 7). *When Antitrust Met Facebook.* Recuperado de https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1421&context=faculty_scholarship

[36] Stucke, M. E. (2018, Marzo 19). *Should We Be Concerned About Data-opolies?* Recuperado de https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Stucke-pp-275-324.pdf (Consultado el 26 de Mayo de 2023).

[37] Birch, K., & Cochrane, D. T. (2021, Mayo 26). *Big Tech: Four Emerging Forms of Digital Rentiership.* Recuperado de https://doi.org/10.1080/09505431.2021.1932794 (Consultado el 27 de Mayo de 2023).

[38] Haucap, J., & Heimeshoff, U. (2013, Agosto 21). *Google, Facebook, Amazon, eBay: Is the Internet driving competition or market monopolization?* Recuperado de https://doi.org/10.1007/s10368-013-0247-6

[39] Stucke, M. E., & Grunes, A. P. (2017, Marzo 3). *Data-opolies.* En: http://dx.doi.org/10.2139/ssrn.2927018

[40] Moreno, J. (2021, Agosto 27). *Google Estimated To Be Paying $15 Billion To Remain Default Search Engine On Safari*. Forbes. Recuperado de https://www.forbes.com/sites/johanmoreno/2021/08/27/google-estimated-to-be-paying-15-billion-to-remain-default-search-engine-on-safari

[41] KShaikhutdinova (2020, Diciembre 15). *How Apple and Google Formed One of Tech's Most Powerful Partnerships*. Wall Street Journal. Recuperado de https://www.wsj.com/video/series/wsj-explains/how-apple-and-google-formed-one-of-techs-most-powerful-partnerships/ACED4938-1A00-4031-923F-390E6C6B0B6F

[42] Taibbi, M. [@mtaibbi]. (2022, Diciembre 3). *Twitter Files Part 1.* Hilo de Twitter. Recuperado de twitter.com/mtaibbi/status/1598822959866683394 (Consultado el 11 de Abril de 2023).

[43] Jeckmans, A., & Beye, M., & Erkin, Z., & Hartel, P., & Lagendijk, R., & Tang, Q. (2013). *Privacy in Recommender Systems.* Recuperado de https://ris.utwente.nl/ws/portalfiles/portal/5352108/Privacy_in_Recommender_Systems.pdf (Consultado el 24 de Junio de 2023).

[44] Alphabet Investor Relations (2022, Febrero 1). *FORM 10-K*. Recuperado de https://abc.xyz/assets/d9/85/b7649a9f48c4960adbce5bd9fb54/20220202-alphabet-10k.pdf (Consultado el 25 de Junio de 2023).

[45] G Data Software AG (2017, Septiembre). *El nuevo Reglamento de Protección de Datos de la UE (GDPR) – Lo que las empresas deben saber*. Recuperado de https://file.gdatasoftware.com/web/es/documents/whitepaper/G_DATA_El_nuevo_Reglamento_de_Proteccion_de_Datos_de_la_UE_GDPR.pdf (Consultado el 28 de Junio de 2023).

[46] *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. Recuperado de https://www.boe.es/doue/2016/119/L00001-00088.pdf (Consultado el 28 de Junio de 2023).

[47] Ma, Y., & Zhao, J., & Li, K., & Cao, Y., & Chen, H., & Zhang, Y. (2022, Octubre 4). *Research Review on the Application of Homomorphic Encryption in Database Privacy Protection*. Recuperado de https://www.researchgate.net/publication/355329464_Research_Review_on_the_Application_of_Homomorphic_Encryption_in_Database_Privacy_Protection (Consultado el 4 de Julio de 2023).

[48] Fontaine, C., & Galand, F. (2007, Octubre 24). *A Survey of Homomorphic Encryption for Nonspecialists*. Recuperado de https://jis-eurasipjournals.springeropen.com/articles/10.1155/2007/13801 (Consultado el 4 de Julio de 2023).

[49] Acar, A., & Aksu, H., & Selcuk, A., & Conti, M. (2018, Julio 25). *A Survey on Homomorphic Encryption Schemes: Theory and Implementation*. Recuperado de https://dl.acm.org/doi/10.1145/3214303 (Consultado el 4 de Julio de 2023).

[50] Chor, B., & Goldreich, O., & Eyal, K., & Sudan, M. (1998, Noviembre 1). *Private information retrieval*. Recuperado de https://dl.acm.org/doi/10.1145/293347.293350 (Consultado el 7 de Julio de 2023).

[51] Yang, P., & Gui, X., & An, J., & Tian, F. (2017, Mayo 26). *An Efficient Secret Key Homomorphic Encryption Used in Image Processing Service*. Recuperado de https://www.hindawi.com/journals/scn/2017/7695751/ (Consultado el 8 de Julio de 2023).

[52] Morris, L. (2013, Mayo 10). *Analysis of Partially and Fully Homomorphic Encryption*. Recuperado de http://gauss.ececs.uc.edu/Courses/c5156/pdf/homo-outline.pdf (Consultado el 9 de Julio de 2023).

[53] Aguilar-Melchor, C., & Fau, C., & Fontaine, C., & Gogniat, G., & Sirdey, R. (2013, Febrero 13). *Recent Advances in Homomorphic Encryption: A Possible Future for Signal Processing in the Encrypted Domain*. Recuperado de https://ieeexplore.ieee.org/document/6461628 (Consultado el 9 de Julio de 2023).

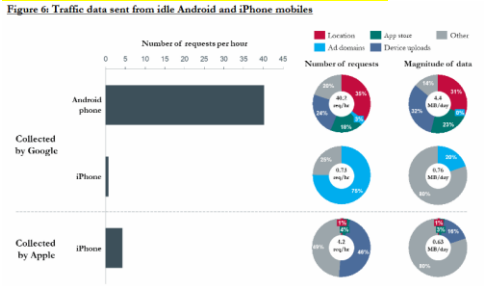
[54] Rivest, R., & Adleman, L., & Dertouzos, M. (1978). *On data banks and privacy homomorphisms*. Recuperado de https://cdn.sanity.io/files/r000fwn3/production/c365f01d330b2211e74069120e88cff37eacbcf5.pdf (Consultado el 9 de Julio de 2023).

Apéndices

a. La Supremacía de Google. Google conoce nuestros pensamientos e intereses debido a su buscador, así como nuestras preferencias de compra y hábitos de gasto a través de Google Pay y Google Wallet. También dispone de información acerca de los videos que visualizamos en YouTube, además de los comentarios y búsquedas que realizamos en la plataforma []; así como los sitios que frecuentamos y los recorridos que hacemos gracias a Google Maps. Adicionalmente, Google cuenta con información sobre nuestras actividades programadas y rutinas diarias debido, en parte, a Google Calendar, así como de los cursos a los que atendemos y las tareas que entregamos, incluyendo detalles sobre nuestra responsabilidad y puntualidad al entregarlas [Google Classroom]. De manera similar, Google tiene acceso a nuestras amistades y relaciones, así como a las conversaciones que sostenemos con ellas, debido a Gmail y otros productos relacionados. Asimismo, tiene acceso a nuestra apariencia física, así como a la de nuestros amigos, a través de herramientas como Google Photos. (https://policies.google.com/privacy)

Además de otros datos personales que puede obtener para establecer nuestra identidad, gracias a los múltiples productos y plataformas de las que es propietario

Figure 6: Traffic data sent from idle Android and iPhone mobiles



b. Conceptos Matemáticos.