Contenido Personalizado con *Homomorphic Encryption*

Preservando la Privacidad de los Usuarios

Roger Rovira
Junio 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. En la actualidad, la identidad digital está principalmente controlada por un pequeño grupo de empresas que gestionan y almacenan los datos personales de la mayoría de usuarios. Lo cual ha generado una disminución en la privacidad e independencia individual, al privar a los individuos del control de su propia identidad, así como un aumento en la centralización del poder, puesto que las empresas han obtenido suficiente información como para influir en la opinión pública y en los procesos democráticos. Por ende, en el presente escrito exponemos como la implementación de un sistema de identidad digital completamente descentralizado otorgaría nuevamente plena autonomía a los individuos sobre sus datos personales, permitiéndoles elegir qué información de su identidad desean compartir en función del contexto social en el que se encuentren. La criptografía de clave pública y las tecnologías descentralizadas como la *blockchain* sirven como marco para establecer dicho sistema, el cual no solo solucionaría la disminución de la libertad individual, sino que también reduciría la centralización del poder y proporcionaría un medio fiable con el que establecer relaciones sociales que trasciendan los intereses financieros en las tecnologías de registro distribuido. Asimismo, en las siguientes páginas se exploran otras posibles implementaciones del sistema planteado y se discutirán las posibles ventajas y desventajas que podrían surgir con su adopción generalizada. Se espera que este artículo demuestre la urgencia y la viabilidad de esta visión, presentando consigo un posible sistema de identidad digital con el que asentar las bases para un mundo más justo y equitativo.

Keywords: *[...]*

Tabla de Contenidos

Introducción

~ Objetivos e Hipótesis ~

Introducción.
El avenir de la era digital, del Internet y del *Big Data* ha estado marcada por el flujo constante de información y la omnipresencia de la tecnología, las aplicaciones en línea se han convertido en una parte importante de nuestra vida diaria. Pero así como todos y cada uno de nosotros obtiene información de dichas plataformas, ellas obtienen información de todos y cada uno de nosotros. Y cada vez más el precio que debemos pagar para estar conectados es nuestra propia privacidad. Y esto es un problema porque: Privacy is not an option, and it shouldn't be the price we accept for just getting on the Internet.

Motivación.
El mundo ha experimentado cambios constantes a lo largo de la historia, mucho antes de que yo naciera y antes de que todos los que ahora habitan la Tierra lo hicieran. Sin embargo, me aventuro a afirmar que nunca antes había ocurrido una aceleración tecnológica de tal magnitud en la historia. Es en tiempos de cambios como este cuando debemos asegurarnos de que los equilibrios adecuados queden establecidos. Por ello, he decidido enfocar este trabajo en la búsqueda de una solución al desafío más urgente e importante que acecha en los bosques digitales[1]: el problema de la privacidad.

Creo firmemente en la importancia de equilibrar el avance tecnológico con la protección de nuestros derechos y libertades. El trabajo realizado refleja esta convicción mía, en la cual deberíamos preservar un valor fundamental que nos permita retomar el control de nuestra información personal en la sociedad moderna.

Objetivos.

Metodología del trabajo.
El trabajo se constituye por dos partes interrelacionadas pero distintas en su enfoque y ejecución.

1. La dimensión escrita: Esta primera etapa ha sido elaborada mediante la revisión de diversos artículos, informes y publicaciones periodísticas. Obteniendo la información necesaria para exponer y analizar los desafíos asociados a la centralización de la información y la protección de datos. Contextualizando consigo la relevancia de la iniciativa práctica que complementa este trabajo.
2. La esfera práctica: La segunda fase del proyecto se centra en desarrollar una plataforma que garantice la confidencialidad del usuario en cualquier situación. Esta confidencialidad no es meramente una declaración de intenciones, sino una característica inherente a su diseño. Para lograr esta visión, se han utilizado distintos lenguajes de programación (como Rust, JavaScript, HTML, CSS, Python, etc.) junto con varias librerías, como Actix[2], y repositorios de código abierto, tales como Spiral[3] o SpiralWiki[4]. Estos diversos elementos se entrelazan para dar forma a una plataforma web que facilita el acceso a una variedad de artículos de manera totalmente confidencial, mediante el uso de *Private Information Retrieval.* (El código correspondiente está disponible en GitHub[5], bajo la licencia MIT.)

Organización.
En lo que respecta a la estructura del resto del escrito, se ha organizado de la siguiente manera. *[...]*

1. Existen diversos problemas en el ámbito digital que tienen bastante apoyo de la comunidad, pero la privacidad es uno de los problemas más importantes con dos tercios de los consumidores globales preocupándose por el control excesivo que ejercen las empresas sobre sus datos personales [].

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

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

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

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

1

Centralización de datos personales

~ Meta, Google y TikTok ~

1.0 Sección.

Los modelos de usuarios centralizados siguen siendo predominantes en el mercado, debido a la prevalencia de los servicios personalizados en Internet [23]. El propósito de esta sección es informar al lector acerca de la inquietante acumulación de poder que emerge de la centralización de la información.

1.2 Tabla de subsecciones.

1.2 Era digital.

En las últimas dos décadas, la humanidad ha sido testigo del rápido desarrollo de tecnologías digitales capaces de mejorar el bienestar humano, así como ha presenciado una creciente concentración de dichas tecnologías en manos de un reducido grupo de potencias tecnológicas, como Google, Apple, Meta, Amazon y Microsoft; comúnmente conocidas como las *Big Tech* [14] [36] [37]. Por consiguiente, la era digital, que se planteaba como una excelente oportunidad para democratizar las instituciones y mejorar el acceso a la información y el conocimiento en todos los estratos de la sociedad [17] [14], resultó ser una época contradictoria a lo esperado. En lugar de liberar a la humanidad y fomentar su independencia, la era digital ha sometido a las personas a la simple merced de las grandes compañías tecnológicas.

Hace veinticinco años Meta no existía, Google era apenas un proyecto universitario y Amazon solo se enfocaba en la venta de libros [14]; y, sin embargo, son actualmente consideradas como algunas de las empresas más valiosas e influyentes del mundo. Dada su inmensa magnitud y alcance, estas compañías tienen la capacidad de influir sobre el comportamiento de los usuarios y la economía internacional [37]; lo que les otorga un poder sin precedente que se manifiesta en varios aspectos de nuestra vida cotidiana.

Un ejemplo relevante es el caso de Cambridge Analytica, en el cual se puso de manifiesto la función que desempeñaron los datos acumulados de Facebook en las quincuagésimas octavas votaciones presidenciales de los Estados Unidos [8]. Demostrando así la gran repercusión que los datos de las grandes empresas tecnológicas pueden ejercer en la opinión pública y los procesos democráticos.

Del mismo modo que Meta se ha apoderado de la información de sus usuarios, tanto Google como Twitter han sido capaces de monopolizar la información de diversas maneras [14], colocándose en una posición privilegiada para capitalizar la innovación futura y ofrecer servicios personalizados.

1.3 Datos personales.

Las *Big Tech* han logrado establecerse entre las compañías más relevantes y poderosas del mundo debido a su innata habilidad para recopilar y utilizar grandes cantidades de datos en beneficio de su propio crecimiento. Google, por ejemplo, cuenta con una amplia base de datos que le permite identificar patrones y tendencias para mejorar los resultados de búsqueda que ofrece a sus usuarios [5] [14] [38]; y, de manera similar, Meta personaliza el contenido que muestra a sus usuarios basándose en sus intereses y preferencias [37].

Dichos escenarios comportan que los consumidores se sientan limitados a utilizar un servicio en específico en vez de su opción preferida, debido a la comodidad y adaptabilidad de la plataforma en cuestión. Optando de este modo por no utilizar un servicio que podría proporcionarles una mayor privacidad y soberanía [14] [36].

Por ejemplo, los usuarios pueden preferir las políticas de privacidad de DuckDuckGo pero permanecer con el motor de búsqueda dominante (Google), que, al beneficiarse de su amplia red de usuarios y de la recopilación de datos de los mismos, ofrece mejores resultados de búsqueda [38] [36] [39]. De manera análoga, los conductores pueden preferir un servicio de navegación que priorice la privacidad, pero quedarse con la aplicación dominante de Google, conocida como Google Maps [36]; cuya finalidad no es solo presentar la mejor información de tráfico posible, sino que también busca recopilar la mayor cantidad de datos.

	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

Tabla 1.1: Participación porcentual del mercado de motores de búsqueda a nivel mundial. Fuente: Elaboración propia con datos de StatCounter[6] y Photutorial[7].

Si nos enfocamos en el sector de los motores de búsqueda, resulta evidente que Google ocupa una posición predominante en dicha industria (véase la Tabla 1.1) [38] [41]. Posición que se atribuye a la habilidad de Google para analizar los diversos datos que recopila con sus productos. Debido a esta capacidad, Google puede mejorar los perfiles de usuario con información exclusiva que ningún otro competidor posee. Esta ventaja

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).

cautiva a los usuarios y los incentiva a utilizar los servicios de Google, ya que no hay ninguna otra plataforma que ofrezca una personalización comparable o unos resultados de búsqueda tan precisos[8]. Sin embargo, es importante considerar si esta razón justifica plenamente la extensión en la que Google recopila información. ¿Debemos comprometer nuestros datos personales con el fin de obtener resultados más precisos? ¿O es posible encontrar un equilibrio entre la comodidad y la protección de datos?

1.4 El valor oculto de los datos.

Partiendo del hecho descrito, se infiere que la calidad del producto está estrechamente relacionada con la cantidad de datos recopilados; ya que dichos datos permiten a la compañía ofrecer información más precisa.

Además, debido a los datos, las empresas son capaces de identificar patrones, tendencias y oportunidades de negocio [5] [14]; lo que a su vez les permite generar ingresos mediante modelos publicitarios o servicios adicionales. De ahí que varias empresas tecnológicas ofrezcan plataformas de comercio, redes sociales (p. ej. Instagram o Twitter) o navegadores (p. ej. Google o Microsoft Bing) de forma gratuita, puesto que utilizan nuestros datos como forma de compensación [36].

> Si no estás pagando por ello,
> no eres el cliente;
> eres el producto que se vende[9].

Si los datos carecieran de valor o no tuviesen una importancia elevada, resultaría improbable que Google hubiera pagado 82 millones de dólares en 2009 a Apple, para asegurar su posición como motor de búsqueda predeterminado en Safari [39]. De igual forma, los pagos de 1000 millones de dólares que Google realizó en 2013 y 2014 para el mismo propósito [39] no encontrarían justificación alguna; y resultaría aún más difícil comprender el desembolso de 15000 millones de dólares[10] que Google realizo en 2021 con el mismo fin [40] [41].

Asimismo, en caso de que los datos no tuvieran un alto valor, resultaría difícil explicar por qué Meta pagó 16000 millones de dólares en la adquisición de WhatsApp, una empresa compuesta únicamente por 60 empleados y sin activos tangibles [14]. Más aun, habría poca o ninguna justificación con la que explicar por qué Meta eliminó la pequeña tarifa de WhatsApp en algunos países[11] [36].

A pesar de que los usuarios (p. ej. de WhatsApp o Google) reciben el beneficio inmediato de utilizar un servicio gratuito, es importante tener en cuenta que el costo a largo o corto plazo de divulgar información personal puede resultar incluso mayor que el de pagar una modesta tarifa anual. Esto se debe principalmente a la falta de conocimiento, por parte de los usuarios, acerca de cómo se utilizarán dichos datos en el futuro y por qué entidades serán empleados [36].

1.5 Incremento de información generada.

Según un informe publicado por IBM en 2013, el 90% de los datos existentes hasta ese momento habían sido generados en los dos años anteriores [29]; y, en una publicación más reciente [11], realizada en 2020, IBM ha señalado que la cantidad de datos generados a diario se ubica en torno a los 2.5 trillones de bytes[12]. Esta gran cantidad de datos ya no se limita únicamente a información básica (p. ej., nombre, género, edad y correo electrónico), sino que ahora también abarca a una amplia variedad de datos no estructurados. Entre los que se incluyen registros de navegación, datos de transacciones, archivos de correo electrónico, mensajes de texto, información geoespacial, imágenes, contactos, relaciones, e incluso preferencias y opiniones personales [14]. (En la Figura 1.1 se presenta un gráfico aproximado que ilustra el crecimiento de la información generada a lo largo de los años.)

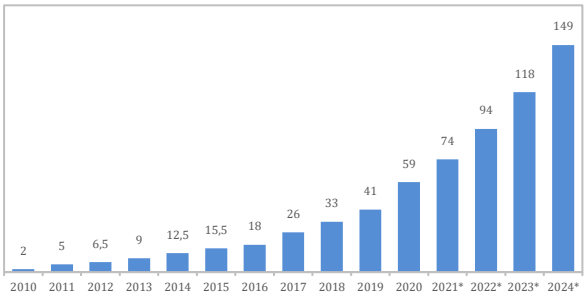


Figura 1.1: Volumen de datos generados a nivel mundial con proyecciones para el período de 2021-2024. Fuente: Elaboración propia con datos de Statista[13].

A partir de este aumento en la cantidad de información, surgió el concepto de *Big Data*. El cual se define como el conjunto de datos, tanto estructurados como no estructurados, que se distinguen por cuatro características fundamentales conocidas como las "cuatro Vs": volumen, velocidad, variedad y valor.

En el contexto del Big Data el énfasis no recae tanto en la calidad de los datos, sino en la cantidad [14]. De este modo, todas las compañías que tengan acceso a dicha cantidad de datos, así como la destreza para analizarlos, poseerán la habilidad de indagar tanto en los intereses e inclinaciones individuales (de cada persona) como en las tendencias y preferencias colectivas (p. ej. de un país, una ciudad o una

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).

comunidad); mediante el reconocimiento de patrones y formas de comportamiento.

El hecho de conocer las preferencias de cada individuo o colectivo, permite a las compañías influir en las acciones y la toma decisiones de dichos usuarios; la finalidad de influir en nuestro comportamiento puede ser meramente económica o sencillamente democrática. Dicha capacidad conlleva una mayor concentración del poder y una disminución en la diversidad del mercado; lo que, a su vez, resulta en una progresiva limitación de la privacidad y libertad individual.

1.6 Información centralizada.

Todo este nuevo conjunto de información que obtienen las empresas sobre nosotros se encuentra completamente centralizada y resguardada en sus propios servidores. No obstante, el almacenamiento centralizado de semejante cantidad de datos implica el surgimiento de problemas derivados de la centralización, tales como la eliminación intencionada de datos privados [23], el uso indebido de información confidencial, el acceso no autorizado a datos personales o la falta de control de los usuarios respecto a los mismos.

En el siguiente apartado, presentamos el caso de Cambridge Analytica, cuyos acontecimientos nos permitirán interiorizar las implicaciones negativas de almacenar dicha cantidad de información en servidores centralizados.

1.7 Facebook y Cambridge Analytica.

Escándalos, como el de Cambridge Analytica, han demostrado que existen deficiencias significativas en los mecanismos actuales de gestión y administración de datos personales [9]. No más lejos, sin embargo, dichos sucesos también han resaltado la importancia de revisar y fortalecer la protección de privacidad de los usuarios en el ámbito digital.

De acuerdo a la información provista por Meta Investor Relations [15], Facebook posee actualmente más de 2000 millones de usuarios activos a diario, lo que equivale aproximadamente a un cuarto de la población mundial. Cuando se trabaja con cifras de semejante magnitud, incluso la más ínfima modificación o la menor transgresión de los derechos del consumidor, puede ocasionar consecuencias catastróficas. La relevancia de las cuales aumenta cuando se trata de información personal con carácter político, tal como acontece con los datos recolectados por la plataforma Facebook[14].

En 2018, se formuló una acusación en contra de Cambridge Analytica, una compañía privada especializada en análisis de datos, por haber utilizado información confidencial de los usuarios de Facebook a fin de generar publicidad dirigida en apoyo a la campaña electoral de Donald Trump[15] [20].

Según el informe presentado por Mike Schroepfer [22], quien en ese entonces ocupaba el cargo de *Chief Technology Officer* (CTO) de Meta Platforms, Inc., se reveló que Cambridge Analytica había obtenido datos potenciales de más de 87 millones de usuarios de Facebook[16]. La obtención de dichos datos se realizó mediante la explotación de una vulnerabilidad en las interfaces de programación de aplicaciones (APIs) de la plataforma[17]. Asimismo, cabe destacar que la apropiación de los datos en cuestión se ejerció sin el debido consentimiento de los usuarios afectados.

Posteriormente, los datos fueron analizados con el fin de crear perfiles psicológicos y mostrar anuncios específicos para atraer votantes. En términos generales, la campaña lanzo 4 mil anuncios publicitarios que, en conjunto, lograron más de 1400 millones de impresiones [14]. Se argumenta que los anuncios dirigidos elaborados por Cambridge Analytica fueron fundamentales para el éxito de la campaña electoral de Trump en 2016, al dirigirse a los votantes con mensajes específicos que resonaron con sus necesidades y deseos.

De este modo, el caso expuesto demuestra que, aunque se asuma que Facebook no tiene ningún interés político, su función como entidad centralizada y su habilidad para recopilar y almacenar grandes cantidades de datos, originó la consecuencia no deseada de permitir que terceros utilizaran esa información para fines malintencionados.

En el siguiente apartado, se expondrá a Google y su gran influencia en la sociedad contemporánea, así como su amplio alcance en la recolección de datos.

1.8 Google, *don't be evil*[18].

Google es una de las compañías con más riqueza e influencia a nivel mundial que, a diferencia de cualquier otra, ha logrado alcanzar una posición única de poder [14] [38]. En efecto, Google es el medio a través del cual experimentamos y comprendemos el mundo que nos rodea. Con el propósito de:

> Organizar la información del mundo y
> hacer que sea útil y accesible
> para todos[19];

Google ofrece una amplia gama de servicios, en su mayoría gratuitos, cuyo fin último no es solo presentar la información

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/.

de una manera más organizada, sino también recopilar la mayor cantidad de información posible [5]. De hecho, el poder de esta empresa se basa únicamente en cómo utiliza la gran cantidad de datos que adquiere de sus usuarios [4].

No obstante, existen numerosos usuarios que niegan o muestran indiferencia ante el mencionado uso de datos que realiza Google y, en consecuencia, desconocen el poder de esta empresa. (Si se reflexiona sobre ello, es probable que se nos venga a la mente alguien con estas actitudes.) Sin embargo, a pesar de la imagen neutral que Google proyecta[20], su capacidad para decidir quién accede a qué información y oportunidades la sitúa como una entidad de considerable poder y amenaza en relación a la privacidad y otros aspectos. Si bien Google se esfuerza por ofrecer resultados de búsqueda objetivos y precisos, en ocasiones estos resultados pueden verse sesgados o influenciados por los intereses comerciales de la compañía.

En 2015 la división Buró de Competencia de la Comisión Federal de Comercio de los Estados Unidos (Federal Trade Commission, FTC), encargada de la eliminación y prevención de prácticas comerciales anticompetitivas, publico de manera accidental partes de un informe sobre una investigación de Google al Wall Street Journal [4]. En el informe, la FTC señalo que Google estaba favoreciendo a sus propios productos y adoptando una estrategia anticompetitiva al no mostrar ciertos sitios web especializados en categorías altamente comerciales [14] [38]. Asimismo, en octubre de 2020, el Subcomité del Poder Judicial de la Cámara de Representantes de los Estados Unidos sobre Estado Administrativo, Reforma Regulatoria y Antimonopolio publicó un informe enfocado en el dominio de las cuatro grandes empresas de datos (Google, Apple, Meta y Amazon); en el que se señalaban nuevamente las prácticas anticompetitivas 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.)

Tabla 1.2: Información compartida con Google en la mañana de un *día de cada día*[21]. Fuente: Elaboración propia con inspiración de [31].

La información que proporcionamos a compañías tecnológicas, como Google y Meta, es extremadamente personal; llegando incluso a casos en los que las compañías saben más atributos de nuestra propia personalidad que todos nuestros amigos juntos. En el informe de Schmidt [31], se detallan varios experimentos que evidencian el preocupante alcance de la recolección de datos por parte de Google. La compañía recopila información cada vez que los usurarios interactúan con alguna de sus plataformas (p. ej., Chrome y Android[22]), aplicaciones (como Google Maps, YouTube y Gmail), herramientas para editores (p. ej., Google Analytics y AdSense) y otros programas; lo que resulta en un conocimiento excesivo de información personal. (Consulte el Apéndice A ara obtener una perspectiva más amplia de la *data-opoly*[23] de Google.)

Los aspectos más confidenciales e íntimos de la información almacenada por Google, ejemplifican la naturaleza única de los datos como producto de comodidad distinto a cualquier otra mercancía controlada por entidades centralizadas [14]. Los datos, a excepción de las otras mercancías, pueden ser muy valiosos para la elaboración de perfiles masivos, utilizados para comprender las preferencias y tendencias colectivas, así como para la creación de perfiles individuales, utilizados para interiorizar las preferencias e inclinaciones personales [5]. (Véase usted la Tabla 1.2.)

Al aceptar los términos de servicio, los usuarios aceptan compartir la mayoría de sus datos con Google, independientemente de la privacidad o confidencialidad de los mismos. (Véase el Apéndice A para una descripción detallada de la información compartida.) Dicho conjunto de información se procesa utilizando diversas técnicas de análisis de datos, con la finalidad de generar perfiles de usuarios a nivel individual y masivo [10].

Si bien la recolección y el procesamiento de datos personales son prácticas necesarias para mejorar la calidad y personalización de los servicios, es esencial considerar los diversos riesgos de seguridad asociados con el intercambio de información personal. A pesar de las numerosas implicaciones negativas, sin embargo, la mayoría de usuarios no solo están de acuerdo, sino que incluso deseosos de compartir sus datos confidenciales con el fin de recibir servicios más personalizados e integrados [5]. De hecho, los consumidores están regalando su información bajo un falso ideal de comodidad y cediendo sus derechos de privacidad en pos de un servicio más personalizado e individualizado [5]; lo que hace

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].

que los usuarios pierdan la soberanía de sus datos y queden estrictamente sujetos al poder de aquellos que la poseen.

El siguiente apartado, constituye el último ejemplo de monopolización de información [...] En este se expondrá como

1.9 TikTok.

[...]

1.A Plataformas en las sombras.

[...]

1.B *Data Network Effect.*

Las tecnologías digitales han avanzado de manera continua y progresiva hacia una arquitectura centralizada que compromete los derechos de los consumidores y pone en peligro la confidencialidad de la información [5]. Los casos de TikTok, Google y Facebook evidencian el control monopólico que las empresas tecnológicas ejercen sobre la información personal, el conocimiento colectivo y los medios de comunicación. Asimismo, estos tres casos resultan ser ejemplos claros que ilustran las posibles implicaciones negativas de la centralización de los datos, o en otras palabras, del poder.

1. En el caso de Facebook, se ha criticado la recopilación masiva de datos personales y su potencial impacto en las elecciones de los usuarios. Además, se ha señalado que, aunque se asuma que estas empresas no posean motivaciones más allá de las económicas para analizar y utilizar dichos datos, su estructura centralizada y su capacidad para recopilar y almacenar grandes volúmenes de información, ha generado la posibilidad de que terceros realicen un uso indebido de esta.
2. En referencia a Google, se ha observado con preocupación la extensa recopilación de datos que realiza la compañía con todas sus plataformas; ya que estas, no se limitan únicamente a proporcionar información, sino que también tienen como objetivo recabar la mayor cantidad de datos posible.
3. Por último, en el caso de TikTok, [...] (~~El siguiente apartado, constituye el último ejemplo de monopolización de información [...] En este se expondrá como~~)

El surgimiento de grandes potencias digitales como Google y Meta, o el establecimiento de importantes servicios en línea como Twitter, se debe principalmente a un efecto de red basado en los datos personales, denominado en inglés como *data network effect* [14] [38] [39]. Los efectos de red son tales que cuantos más usuarios hay en una plataforma, más valiosa se vuelve ésta para cada usuario [5] [35] [36].

Los teléfonos constituyen un ejemplo clásico. A medida que más personas adquieren dispositivos telefónicos, más personas estarán disponibles para realizar llamadas. A su vez, tener un mayor número de personas a las que llamar aumenta el valor inherente de poseer un teléfono [36].

En el contexto de los datos, en cambio, a medida que la plataforma atrae a un mayor número de usuarios y acumula más información, su habilidad para personalizar el contenido se intensifica y la comodidad del usuario aumenta. Este aspecto fomenta el regreso del usuario y estimula la voluntad de compartir información personal (Figura 1.3). Asimismo, dado que nadie puede ofrecer servicios tan individualizados, el negocio de la compañía se consolida en el mercado.

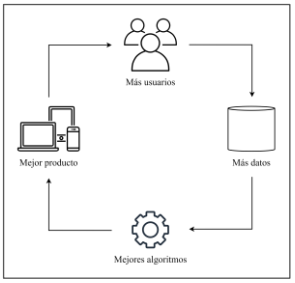


Figura 1.3: Efecto de red basado en los datos.

Tomemos como ejemplo a Facebook. Dicha plataforma recolecta información de sus usuarios para generar bienes y servicios personalizados [36], cuyo último fin es aumentar la cantidad máxima de usuarios en la red y su dependencia de ella, de manera que se logre bloquear la mayor cifra de usuarios en el sistema [5]. En efecto, a medida que aumenta la cantidad de usuarios en la red y la plataforma obtiene más datos sobre nosotros, se genera un mayor nivel de comodidad y una mayor dependencia respecto al servicio; lo que contribuye al crecimiento y la consolidación de Facebook en el mercado de las redes sociales. Este mismo fenómeno también puede aplicarse a los motores de búsqueda [36], como Google Search o Microsoft Edge, así como a la computación en la nube [5], como AWS o IBM Cloud, entre otros ejemplos.

1.C Ingenuidad humana.

La gran capacidad económica que poseen los datos es la que motiva a las empresas a recopilar la mayor cantidad de información posible sobre sus consumidores. Como resultado, dicha necesidad de información se ha convertido en la principal lucha para los lideres tecnológicos [32].

Debido al objetivo de las compañías de saber tanto como sea posible de sus usuarios, no importa cuánto uno trate de mantener su información en privado, existen mecanismos y herramientas por doquier que colectan datos personales y los comunican a empresas terciarias, ya sea con o sin el consentimiento del individuo [5]. Más a menudo, sin embargo, son los propios usuarios quienes desean compartir información personal con una amplia variedad de empresas interesadas. La mayoría de usuarios dan información a compañías sin conocer las implicaciones de privacidad de dichos servicios o valoran el servicio y la comodidad más allá de su riesgo personal [5].

Por ende, no es sino el afán de comodidad del ser humano y la falta de conocimiento e indiferencia de los usuarios, en relación a la recopilación de la información; lo que dota a las empresas de un poder inconmensurable, que jamás se había imaginado ni mucho menos concentrado.

2

Identidad y comodidad

~ Como los datos nos definen ~

2.1 Sección.

En esta sección hablaremos sobre las soluciones tanto tecnológicas como [...].

2.2 De camino a las soluciones.

Al utilizar aplicaciones en línea, los usuarios suelen compartir una abundante cantidad de información personal. Ya sea al realizar compras en Internet, o al utilizar tarjetas de crédito[24]; al acceder a enlaces y artículos, o al publicar actualizaciones personales; al calificar series y películas, o al comentar en publicaciones; la información siempre se comparte dentro de un ámbito particular. Sin embargo, cuando parte de la información se mueve más allá de su alcance previsto (bien sea porque se mantiene más tiempo del establecido, se comparte con corporaciones no autorizadas o se abusa de la información para un propósito diferente al acordado), se produce una violación de la privacidad [43].

Actualmente resulta complicado conocer con certeza el uso que le están dando a nuestros datos. Pero aún resulta más difícil prever cómo y por quien serán utilizados en el futuro [36]. Llegando a un punto en el cual los propios usuarios, quienes son los legítimos propietarios de su identidad[25], dejan de ser los verdaderos propietarios de esta y ceden el poder a las empresas. La apropiación del control de nuestros datos personales por parte de un reducido conjunto de empresas, nos priva y expropia de la facultad de manejar nuestra información confidencial; privándonos de nuestro legítimo derecho de propiedad sobre los datos que conforman nuestra identidad. En su lugar, estas empresas adquieren un dominio absoluto o parcial sobre nuestros datos.

Si bien el Internet fue diseñado como un sistema descentralizado para maximizar la resistencia y eliminar la posibilidad de un único punto de falla [5], la proliferación de compañías (como Facebook o Google) que constantemente desarrollan tecnologías capaces de almacenar y recolectar grandes cantidades de información personal, ha transformado a Internet en un entorno centralizado. Donde nuestra información se ha convertido en una mera moneda de intercambio.

> A medida que Internet ha evolucionado, pasamos de la era en la que nadie sabía si eras un perro[26], a la era en la que las compañías no solo saben que eres un perro, sino que incluso saben tus patrones de respiración y quieren utilizar esa información para mostrarte anuncios personalizados.

De esta manera, el avance de la economía digital y del *Big Data* ha generado inquietudes entre ciertos usuarios; quienes temen perder el control sobre la forma en que se recopilan y utilizan sus datos. Sin embargo, otros, al tomar conciencia de que esta situación ya es una realidad, comienzan a buscar alternativas para contrarrestar estas preocupaciones. En la literatura relacionada, tanto empresas como investigadores han propuesto diversas *soluciones* para devolver la soberanía de los datos a los usuarios. Entre dichas *soluciones*, las que destacan son las jurídicas y las tecnológicas.

2.3 Soluciones jurídicas para la protección de datos.

La ausencia de un marco regulatorio solido que promueva la transparencia y el control del consumidor sobre sus propios datos, puede generar serios riesgos que afecten el buen funcionamiento de los mercados digitales [9] [14]. En respuesta, se han realizado esfuerzos a nivel mundial para actualizar y establecer nuevas leyes que aborden los desafíos relacionados con la privacidad de la información. Ejemplos destacados incluyen el Reglamento General de Protección de Datos (RGPD) [46] de la Unión Europea, que entró en vigor el 25 de mayo de 2018; y la Ley de Privacidad del Consumidor de California (CCPA), que se implementó el 1 de enero de 2020.

La introducción del RGPD ha marcado el comienzo de un seguido de cambios en la forma en que se gestionan los datos personales de los ciudadanos europeos. Desde una perspectiva jurídica, el objetivo del RGPD es dotar al maco legal con las garantías adecuadas que permitan el control individual sobre los datos personales [9]. Este principio de control se refleja en un conjunto de derechos; los más destacados son:

1. El derecho de supresión o derecho al olvido: Los usuarios tienen el derecho a obtener la supresión de los datos que le conciernan. Esto implica que las empresas deben eliminar o anonimizar los datos de un usuario cuando ya no sean necesarios en relación a los fines para los que fueron recogidos, o cuando el interesado retire su consentimiento [45] (Art. 17, [46]).
2. El requisito del consentimiento informado: Cada cliente debe ser informado en términos sencillos sobre la finalidad de los datos que proporciona [45] (párr. 42-43, [46]). Además, se requiere que los clientes autoricen previamente el uso de sus datos [14] [45] (Art. 6, párr. 1, [46]).
3. El derecho a la notificación inmediata de violaciones de privacidad: En caso de violación de la seguridad, el responsable del tratamiento deberá informar a los usuarios afectados de manera inmediata y sin demoras indebidas; a más tardar 72 horas después de que se haya tenido constancia del suceso [45] (Art. 33, [46]).
4. El derecho a la portabilidad de los datos: Los clientes tienen derecho a recibir los datos conservados sobre ellos en un formato estructurado, de uso común y lectura

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.

mecánica; y tienen derecho a transmitirlos a otras organizaciones y responsables [45] [14] (Art. 20, [46]).

Los derechos mencionados marcan hitos significativos que han sido posibles gracias a los esfuerzos encomiables de diversas entidades e individuos comprometidos en la creación y promoción del RGPD y otras regulaciones similares. Resulta asombroso constatar cómo estas legislaciones representan un pequeño avance para la protección de la privacidad, y un gran paso hacia la soberanía de los datos.

No obstante, también es importante reconocer los desafíos que limitan su capacidad para resolver de manera completa la falta de control que experimentan los usuarios. Aunque estas regulaciones ofrecen un marco regulatorio sólido y establecen derechos fundamentales para todo consumidor, aún existen ciertas limitaciones que la ley no puede abordar por sí misma.

2.4 Confianza en abundancia.

El mero hecho de tener que confiar en los proveedores de servicios o en el responsable del almacenamiento plantea un desafío en sí mismo.

En la actualidad, los usuarios depositan mucha confianza implícita en los proveedores de servicios, esperando que manejen su información de manera justa y consciente; respetando los derechos del consumidor y aplicando mesuras técnicas de seguridad para garantizar la confidencialidad y privacidad de los datos; y que continúen haciéndolo en el futuro [43]. Al utilizar el sistema, los usuarios establecen una relación con el proveedor de servicios, quien (debido a su forma de funcionar) puede ver toda la información en el sistema, incluidas las descargas privadas, las solicitudes de acceso a información, la conducta de compra, el comportamiento de navegación, etc.[27] Por ende, le corresponde al propio proveedor asegurarse de que no se utiliza dicha información sin el consentimiento correspondiente. También, es el proveedor de servicios quien tiene la autoridad final para decidir qué información se almacena, cuánto tiempo se conserva y cómo se usa o distribuye [43]; no es el usuario quien decide de forma directa.

Si bien los reglamentos como RGPD o CCPA establecen estrictas medidas para estos aspectos, es responsabilidad del proveedor decidir si cumple o no con estas mismas normas y aceptar las consecuencias asociadas a su decisión. Asimismo, pueden surgir situaciones en las que la responsabilidad no recaiga directamente en la ética de la compañía, sino que más bien se deba a una mala gestión o falta de recursos técnicos para abordar de manera adecuada dichos aspectos. En la Sección 1.5 se presentó el caso de Facebook y Cambridge Analytica cuyos acontecimientos no fueron ocasionados por una mala ética por parte de Facebook, sino que se produjeron a causa de una mala gestión momentánea.

Por lo general, las declaraciones de privacidad se ofrecen para mostrar la postura adoptada por el proveedor de servicios y obtener el consentimiento del usuario. Sin embargo, muchos usuarios se ven forzados a aceptar los términos y condiciones de uso, ya que de lo contrario el servicio denegaría su acceso; limitando el derecho de los usuarios a seleccionar que información quieren compartir y cual prefieren mantener privada [36].

> En su mayoría, no existe término medio; se trata del "tómalo o déjalo" de las negociaciones tradicionales: o aceptas el uso que se le darán a tus datos o prescindes del servicio.

El equilibrio de poder está claramente a favor del proveedor de servicios [43], y no se puede simplemente asumir que los usuarios consienten efectivamente esta forma de recopilación y procesamiento de datos [36]; dado que los usuarios no tienen otra opción. Debido, en parte, a la posición dominante del servicio. El RGPD y otras regulaciones similares, obligaran a las compañías a delinear mejor su política de privacidad. Pero para aquellos que acepten la política propuesta, estas empresas seguirán recopilando datos de manera extensa.

Si bien las soluciones jurídicas son efectivas para establecer un cierto nivel de privacidad, aún sufren de la debilidad inherente de los modelos basados en la confianza. Por ende, es necesario establecer un sistema que garantice la confidencialidad de la información a través de un modelo basado en pruebas criptográficas en lugar de confianza. Asimismo, a diferencia de las leyes que generalmente se emplean para resolver problemas después de que surjan, un enfoque basado en la tecnología será capaz de prevenir las violaciones de privacidad antes de que sucedan [43].

2.5 Soluciones tecnológicas, introducción.

Las soluciones tecnológicas ofrecen una mayor garantía de confidencialidad de la información, al reemplazar la dependencia de la confianza con mecanismos matemáticos. En general, se reconoce que la tecnología de encriptación es el método técnico más simple y capaz para proteger los datos de los usuarios [47] [48] [53].

La criptografía es un campo especializado que se enfoca en proteger la confidencialidad y seguridad de la información [47], mediante el uso de técnicas de codificación y descodificación. Estas técnicas aseguran que la información permanezca inaccesible para aquellos sin la clave de cifrado adecuada[28].

En criptografía, se distingue entre los esquemas de clave secreta, también conocidos como criptografía simétrica, y los de clave pública, conocidos también como criptografía asimétrica. En los esquemas de criptografía simétrica, tanto el cifrado como el descifrado (de un mensaje específico, al que llamaremos m) se realizan utilizando la misma llave[29], denominada k.

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

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

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.

En la Ecuación 1, la función *Enc* representa a la función de cifrado, que toma una llave y un mensaje sin encriptar como parámetros, y devuelve un texto encriptado, conocido como *ciphertext*. Por otro lado, en la Ecuación 2, la función *Dec* representa a la función de descifrado, que utiliza la misma llave que se utilizó para encriptar el mensaje, y recupera el texto original, m. Dado que se utiliza la misma llave, es necesario que el remitente y el receptor acuerden previamente la clave que utilizarán para establecer cualquier comunicación segura. Esto implica que dichos esquemas no puedan ser utilizados por dos personas que nunca se conocieron [48]. Asimismo, en este sistema, se requiere compartir una clave diferente con cada individuo con el que queremos comunicarnos [48]. No obstante, a pesar de lo dicho, los esquemas simétricos presentan la ventaja de ser realmente rápidos y se utilizan con la mayor frecuencia posible [48].

En contraste con el modelo anterior, los esquemas asimétricos utilizan un par de llaves; una de las cuales, denominada clave pública (representada como k_p), se utiliza para encriptar, mientras que la otra, la clave privada (representada con el subíndice s, k_s), se mantiene en secreto y se utiliza para descifrar el *ciphertext*. Cuando se desea enviar un mensaje encriptado, el remitente utiliza la clave pública del receptor para cifrarlo. Luego, el receptor utilizará su clave privada para descifrar el *ciphertext* recibido y obtener consigo el mensaje original.

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

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

La Ecuación 3 encripta el mensaje utilizando la clave pública y la Ecuación 4 lo descifra con la clave secreta. Los esquemas asimétricos se consideran más flexibles que los simétricos, ya que no requieren que el remitente y el receptor acuerden previamente ninguna clave [48]. Estos esquemas, sin embargo, suelen ser más lentos que los simétricos [48]. Ejemplos destacados de este tipo de esquemas son el RSA y el ElGamal.

2.6 Problemas relacionados con la criptografía.

Los sistemas más conocidos de encriptación dependen de compartir una llave, bien sea pública o privada, entre los individuos involucrados en el intercambio de información [49]. Este enfoque, sin embargo, plantea algunos problemas relacionados con la privacidad. Los usuarios o proveedores de servicios con acceso a la llave tienen derechos exclusivos sobre los datos [49]. Esto implica que tienen el control y la capacidad de acceder, utilizar y administrar los datos personales de manera exclusiva. Especialmente en el caso de los servicios en la nube, existe el riesgo de perder el control sobre la privacidad de la información.

Por lo tanto, para garantizar que solo los legítimos usuarios puedan acceder de forma segura a sus datos, es necesario cifrar la información antes de transmitirla y abstenerse de compartir cualquier clave secreta que permita su descifrado. No obstante, surge un infortunio, y es que aunque el cifrado de datos puede ser eficaz para evitar el acceso no autorizado a los datos, también supone la destrucción de la estructura semántica subyacente de los datos, lo que imposibilita realizar operaciones sobre ellos [47]. En consecuencia, los resultados de las operaciones sobre textos cifrados carecen de significado cuando se emplean métodos de cifrado tradicionales [47]. Esto requiere que, en cualquiera de los casos, los servicios en línea deban descifrar el mensaje antes de operar en ellos, ya que sino el resultado no tendrá significado alguno; comprometiendo consigo el nivel de privacidad que el servidor puede ofrecer a sus consumidores.

Cuando se desea acceder a información almacenada en una base de datos, por ejemplo, el usuario puede enviar de forma encriptada el índice en donde se encuentra dicha información. No obstante, una vez llegue al servidor, este lo deberá descifrar y leer para poder responder con la información correcta; bien sea con el contenido de una revista o con información relacionada con alguna medicina. (Le recomiendo que vea la Figura2.1.) En este caso, existe la posibilidad de que un operador curioso siga las consultas del usuario e infiera en lo que busca [50]; independientemente del nivel de confidencialidad de la consulta realizada o de si se está utilizando una pestaña de *incógnito* o no[30].

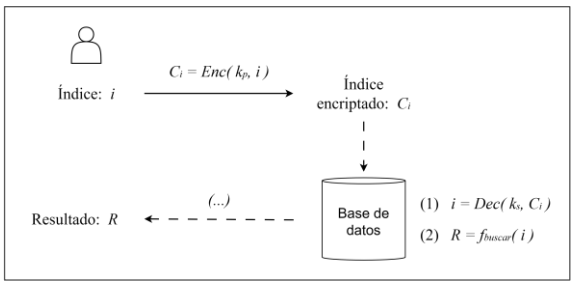


Figura 2.1: Solicitud de información a una base de datos utilizando criptografía tradicional[31].

En dicho contexto y otros similares, los consumidores pueden querer preservar la confidencialidad de los datos proporcionados [53]. Por ejemplo, un usuario puede desear mantener la privacidad de su solicitud y aun así recibir el contenido solicitado. (Véase la Figura 2.2.) Esto nos lleva a la necesidad de un sistema criptográfico que respete la confidencialidad de los datos, no solo durante la comunicación y el almacenamiento, sino también durante su procesado [53]. Por supuesto, el resultado del procesamiento debe ser igual de bueno que si los datos no estuvieran encriptados [53] [52] [48]. De hecho, el resultado descifrado debería ser equivalente al resultado computado en texto sin 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.

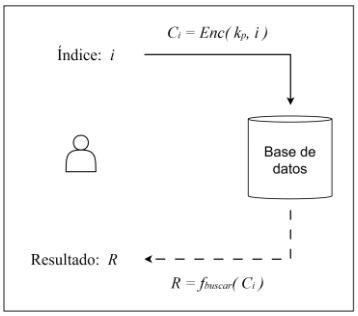


Figura 2.2: Solicitud de información a una base de datos sin descifrar el índice en cuestión[32].

Así pues, poder lograr la coherencia en los resultados de las operaciones ejecutadas sobre textos cifrados, representa un nuevo desafío. Mientras que descubrir la solución se convierte en el objetivo, y la respuesta se transforma en lo que muchos criptógrafos consideran *el Santo Grial de la criptografía* [53] [52].

2.6 El Santo Grial de la criptografía.

En el año 1978, los criptógrafos Rivest, Adleman y Dertouzos propusieron una solución innovadora para abordar el desafío de la privacidad durante el procesamiento de datos [54]. Su propuesta, conocida hoy en día como Cifrado Homomórfico (*Homomorphic Encryption* o HE en inglés), permitía realizar cálculos sobre datos cifrados sin necesidad de descifrarlos. El resultado de las operaciones es devuelto como un resultado encriptado, el cual una vez descifrado, obtiene el mismo valor que si los cálculos hubieran sido efectuados sobre *plaintext*.

Desde este artículo seminal, el diseño de sistemas eficientes y esquemas seguros de encriptación homomórfica ha sido uno de los santos griales de la criptocomunidad.

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

5

Incognito dB

~ Accede a artículos de manera privada ~

5.1 Sección.
[...]

5.2 Tabla de subsecciones.

5.2 Sobre Incognito dB.
[...]

Conclusiones

Agradecimientos

Autor

Des de bien pequeño que he vivido inmerso en un entorno tecnológico. un escenario que ha fomentado enormemente mi interés hacia el ámbito digital. Conocí el arte de programar a una edad temprana, y me interese en las redes descentralizadas cuando la burbuja de las criptomonedas aún no había empezado.

Con el avenir de una nueva era, marcada por el flujo constante de información y la omnipresencia de la tecnología, las aplicaciones en línea se han convertido en una parte importante de nuestra vida diaria. Las personas consumen medios, realizan compras, se informan e interactúan de manera online; con unos servicios que se adaptan totalmente a ellos. Nunca estamos aburridos. Nunca estamos enojados. Nuestros medios son un reflejo perfecto de nuestros intereses y deseos. [Pariser, 2011]

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).

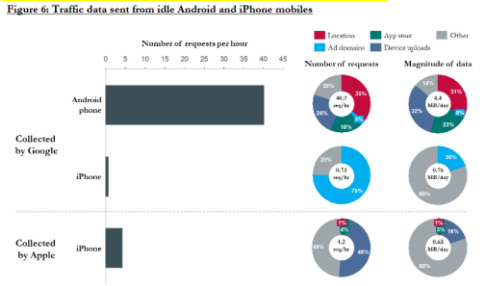
[54] Rivest, R., & Adleman, L., & Dertouzos, M. (1978). *On data banks and privacy homomorphisms*. Recuperado de https://cdn.sanity.io/files/r000fwn3/production/c365f01d330b2211e74069120e88cff37eacbcf5.pdf (Consultado el 9 de Julio de 2023).

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.