Barrido por el lado del **cliente**

una revisión sistemática

Comunicaciones privadas, investigaciones y **derechos**

iris
INSTITUTO DE REFERENCIA EN INTERNET Y SOCIEDAD

Barrido por el lado del **cliente**

una revisión sistemática

AUTORÍA

Gustavo Ramos Rodrigues

Paulo Rená da Silva Santarém

Victor Barbieri Rodrigues Vieira

Wilson Guilherme Dias Pereira

REVISIÓN

Lahis Pasquali Kurtz

Luiza Correa de Magalhães Dutra

TRADUCCIÓN

Zenaide Romanovsky

REVISIÓN EXTERNA

Camila Laranjeira da Silva

Roberta Battisti

PROYECTO GRÁFICO, TAPA, DIAGRAMACIÓN Y FINALIZACIÓN

Felipe Duarte

PRODUCCIÓN EDITORIAL

Instituto de Referência em Internet e Sociedade

CÓMO CITAR EN ABNT

PEREIRA, Wilson Guilherme Dias; RODRIGUES, Gustavo Ramos; SANTARÉM, Paulo Rená da Silva; VIEIRA, Victor Barbieri Rodrigues. **Barrido por el lado del cliente**: una revisión sistemática. Belo Horizonte: Instituto de Referencia en Internet y Sociedad, octubre 2022. Disponible en: <bit.ly/3EAhEDF>. Acceso el: dd mmm. aaaa.

INSTITUTO
DE REFERENCIA
EN INTERNET
Y SOCIEDAD

DIRECCIÓN

Gustavo Rodrigues

Paloma Rocillo

EQUIPO

Ana Bárbara Gomes | Coordinador de Políticas Públicas e Investigadora

Felipe Duarte | Coordinador de Comunicación

Fernanda Rodrigues | Coordinadora de Investigación e Investigadora

Juliana Roman | Investigadora

Júlia Caldeira | Investigadora

Lucas Samuel | Pasante de investigación

Luiza Dutra | Investigadora

Paulo Renã da Silva Santarém | Investigador

Rafaela Ferreira | Pasante de investigación

Thais Moreira | Analista de Comunicación

Victor Barbieri Rodrigues Vieira | Investigador y Responsable de la protección de datos

Wilson Guilherme | Investigadore

irisbh.com.br

SUMARIO

Resúmen ejecutivo

El proyecto "**Comunicações privadas, investigações e direitos"**, del Instituto de Referência em Internet e Sociedade – IRIS, busca ofrecer subsidios confiables para que el debate político y jurídico de investigaciones en comunicaciones privadas en Brasil pueda combinar seguridad de tecnologías de la información y comunicación con protección de derechos humanos y de garantías democráticas. Se pretende analizar impactos y riesgos ; sistematizar conocimiento científico; y, al final, producir recomendaciones para sectores público y privado. El objeto de análisis son tres mecanismos para investigaciones sobre comunicaciones privadas: rastreabilidad de mensajes instantáneos, *hacking* gubernamental, y barrido por el lado del cliente.

En este segundo informe,[1] se evaluó el escenario del **barrido por el lado del cliente** (en inglés *client-side scanning*, con las iniciales CSS, aquí abreviada con BPLC). El término se refiere a técnicas de escaneo realizado en el dispositivo de usuarios ("cliente") para identificación de instancias de reparto de materiales considerados ilícitos – especialmente involucrando contenido sexual de abuso de niños y adolescentes o CSAM (*child sexual abuse material*) – en ambientes protegidos por criptografía segura, en lugar de realizar ese escaneo a nivel de servidor. Por medio de una revisión sistemática de literatura, se investigó un total de 22 publicaciones seleccionadas. Se organizó el material encontrado en contexto, concepto, funcionamiento y problemas.

Primero, el **análisis contextual** de BPLC expone un conjunto de técnicas de detección de contenido indeseado, sus limitaciones en sistemas con criptografía, como también demuestra, de qué forma, la propuesta de 2021 de Apple para superar esa incompatibilidad, ha dado inicio a las controversias que son el objeto central de ese estudio. Segundo, se explica cómo se puede **definir conceptualmente** y clasificar tecnológicamente el barrido por el lado del cliente, considerando el conjunto de técnicas. Tercero, se detallan, en lenguaje accesible, las etapas y los procedimientos de **funcionamiento** del BPLC. Y, en el cuarto eje, el universo de **problemas apuntados** en la literatura se agrupa de acuerdo con la naturaliza de los desafíos a la implementación de esa técnica: a) en el **plan tecnológico**, se cuestiona el funcionamiento, la eficacia, la seguridad y el escopo del BPLC, en razón de los riesgos respectivos, aislados o cumulados, de que sea inutilizado, no alcance los resultados propuestos, se abran brechas de vulnerabilidad o sufra desvío de función; y b) en el **plan jurídico**, se discuten los posibles efectos negativos de la implementación del BPLC sobre la privacidad, el sigilo de las comunicaciones , la presunción de inocencia, y la seguridad pública, como también si la técnica atiende a los principios de proporcionalidad y necesidad.

Sumando análisis y problematizaciones, en la literatura revisada se verifica la prevalencia de las vulnerabilidades tecnológicas no resueltas y de las carencias de justificativas jurídicas compatibles con los riesgos generados por el BPLC. Las defensas de la implementación, mirando hacia el problema de la difícil detección de CSAM en sistemas

con criptografía, parecen despreciar la importancia de los debidos cuidados para una efectiva protección contra existentes fragilidades tecnológicas y probables violaciones de derechos de las personas en general, incluso niños y adolescentes.

Tales resultados confirman la percepción de que BPLC, como supuesta medida alternativa a la rotura de la criptografía, es una medida inadecuada incluso hasta el objetivo de enfrentar la pornografía infantil. Fuera de límites bien delineados, no hay evidencias suficientes de funcionamiento tecnológico efectivamente robusto o libre de una alta susceptibilidad a ataques, al paso en que no se resolvieron graves riesgos jurídicos generales y puntuales, además de cuestiones económicas y sociales.

Presentación

Los primeros debates sobre la criptografía fuerte involucraban la inserción de mecanismos para acceso excepcional de las agencias estatales de investigación y persecución penal a los algoritmos criptográficos. Sin embargo, la sociedad civil y la comunidad técnico-científica fueron bien sucedidas en la defensa de que políticas de seguridad pública deberían considerar riesgos tecnológicos, jurídicos y económicos.

Esos sectores demostraron que las herramientas de rotura de la criptografía – demandadas para investigaciones legítimas por agentes públicos – abrirían brechas para el acceso también por terceros mal intencionados, además de no impedir que personas interesadas en huir de las autoridades migren para otras plataformas que no tengan esas brechas. El resultado sería la población en general con menos seguridad y los sospechosos intocables.[2] Tales argumentos disminuyeron las demandas por soluciones como puertas clandestinas (*backdoors*).[3] Surgieron, aún, alternativas legislativas a la rotura de la criptografía, para dar a las autoridades acceso a datos e informaciones supuestamente necesarias para identificar y punir criminosos.

El proyecto “**Comunicações privadas, investigações e direitos”** busca sistematizar la literatura sobre métodos supuestamente alternativos a la rotura de la criptografía, para nutrir el debate científico, político y jurídico sobre el tema en Brasil. Se pretende ofrecer subsidios confiables para que decisiones políticas, regulatorias y judiciales combinen la seguridad de las tecnologías de información y comunicación con la protección de derechos humanos y garantías democráticas. En específico, se objetiva: 1) analizar impactos y riesgos a la seguridad de datos e informaciones digitales, y derechos involucrados; 2) sistematizar conocimiento sobre técnicas de investigación; y 3) producir recomendaciones para el Estado y empresas.

Los informes científicos analizarán tres métodos alternativos: a) rastreabilidad de mensajes instantáneos, en los cuales se guardan metadatos de la comunicación para futura identificación del camino o del origen de un eventual contenido ilícito; b) barrido por el lado del cliente, por el cual se analiza y compara un contenido en un dispositivo con

bases de datos previos, en busca de un patrón específico; y c) *hacking* gubernamental, por el cual se exploran vulnerabilidades ocultas y no-intencionales de un sistema.

A partir de los resultados, el Instituto de Referência em Internet e Sociedade – IRIS pretende dialogar con diversos sectores y construir posicionamientos sobre esos métodos, con base en evidencias científicas y en el respeto a los derechos humanos. El material se ofrecerá online, para consulta y uso general.

1. Introducción

En la segunda mitad del siglo XX, el debate sobre la disponibilidad pública de criptografía fuerte para la protección de comunicaciones privadas orbitó la inserción de mecanismos de excepción, que viabilizaran el acceso de agencias estatales de investigación y persecución penal al contenido protegido. El tema rindió grandes controversias públicas sobre efectos jurídicos, políticos y económicos, en conflictos en la gobernanza de la criptografía fuerte conocidos como guerras criptográficas (*crypto wars*), marcadas por la oposición de la comunidad técnico-científica, del sector privado y de activistas de derechos humanos en el área digital contra diversos arreglos de acceso excepcional.[4]

Aunque persista la presión pública de autoridades de diversos países por tales mecanismos,[5] la década de 2010 vio nuevos tipos de propuestas. Con la promesa de combinar la seguridad de los sistemas y medios para investigaciones de datos e informaciones exigidos para identificar y punir criminosos, ellas abarcan técnicas de *hacking* gubernamental, de rastreabilidad de mensajes instantáneos con criptografía y de barrido por el lado del cliente, objeto de este estudio.

La discusión sobre mecanismos de barrido por el lado del cliente (muchas veces citados en la literatura en inglés *client-side scanning* – CSS, aquí abreviada por las iniciales en español BPLC) se destacó en 2021, cuando Apple anunció nuevos recursos de seguridad para niños y adolescentes, a fin de combatir la pornografía infantil y la seducción. La propuesta involucraba la comparación de imágenes salvadas en iCloud Photos con un banco de materiales de abuso sexual infanto-juvenil. Lo que se identificara con el Contenido de Abuso Sexual Infantil[6] (*Child Sexual Abuse Material*) – CSAM sería computado y, ultrapasada una cantidad límite de correspondencias (30, inicialmente)[7], sometida a chequeo humano.[8] Ese monitoreo permanente fue criticado por entrar en conflicto con la protección de confidencialidad ofrecida por la criptografía. Si la propuesta de Apple, supuestamente, no reducía sus patrones de seguridad criptográfica fuerte, ¿serían los mecanismos de BPLC legales, eficientes y consistentes?

La pregunta central de la presente investigación es: ¿de qué forma la literatura académica pertinente ve la adecuación del BPLC con el medio investigativo en sistemas con criptografía fuerte sin mecanismos de acceso excepcional? El estudio explora riesgos y desafíos tecnológicos y jurídicos, organiza los principales puntos de defensa y crítica a

la propuesta, y evalúa su viabilidad, por la revisión sistemática de literatura de 22 textos seleccionados a la luz del escenario actual de debate, de los idiomas portugués e inglés, y de la técnica computacional.

Reflejando pros y contras, según el peso político y los parámetros legales del debate, los resultados componen cuatro secciones: contexto de la controversia, concepto de BPLC, funcionamiento de la tecnología, y problemas apuntados. Aun, el Apéndice 1 lista las obras analizadas y el Apéndice 2 replica el formulario de análisis (el mismo utilizado en el primero informe, sobre rastreabilidad de mensajes instantáneos).

2. Metodología

Diversamente del notable acúmulo de estudios detallando[9] riesgos e impactos del acceso excepcional a la criptografía, las supuestas alternativas carecen del mismo escrutinio. Específicamente, el barrido por el lado del cliente que ha alcanzado preeminencia pública internacional a partir del anuncio de que sería implementado por Apple.

El alegato escopo de enfrentamiento a la divulgación de material de exploración o abuso sexual infantil no prescinde de una discusión política y jurídica – sobre si las investigaciones de comunicaciones privadas podrían combinar seguridad de TICs y protección de derechos humanos – fundada en bases técnicas y académicas consistentes, sea para acciones por el sector priva do, sea para políticas públicas.

Para alcanzar ese nivel de densidad y confiabilidad, se realizó una revisión sistemática de literatura: se investigó el estado del arte sobre el tema, con recorte empírico en un grupo de obras seleccionadas y evaluadas mediante criterios y procedimientos explícitos y organizados. Ese método se propone a identificar eventuales lagunas en estudios académicos de cierto campo o temática,[10] como también , cuestiones y subtemas para nuevas investigaciones y proyectos. Pesquisas así:

> *[...] son particularmente útiles para integrar las informaciones de un conjunto de estudios realizados separadamente sobre determinada terapéutica/intervención, que pueden presentar resultados conflictitos y/o coincidentes, como también identificar temas que necesitan evidencia, auxiliando en la orientación para investigaciones futuras.*[11]

En este estudio, se analizó *corpus* documental de las siguientes fuentes: busca por palabra-clave y recogida de referencias bibliográficas de dos obras relevantes seleccionadas.

Primero, se operó una búsqueda por la **palabra-clave "*client-side scanning*"** en la plataforma Google Académico[12]. Aunque el informe adopte la traducción del término "barrido por el lado del cliente", la búsqueda por el término en inglés se justifica por la escasez de bibliografía amplia en portugués, como también por la preeminencia de los debates en inglés impulsados por las herramientas anunciadas por Apple. Se encontraron 38 referencias, siendo excluidas 3 entradas repetidas y 2 textos de acceso restricto.

El segundo y el tercer subconjuntos de obras vinieron de las referencias bibliográficas constantes de dos textos elegidos de modo discricionário: "*Bugs in our Pockets: The Risks of Client-Side Scanning*", de ABELSON y otros,[13] y "Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta", do Center For Democracy & Technology (con traducción en portugués). Esos dos textos, ya conocidos del equipo, fueron elegidos ´ por la relevancia en el debate sobre métodos alternativos a la rotura de la criptografía, y por enfocar controversias involucrando el barrido por el lado del cliente. Se seleccionaron todas las 68 referencias citadas en el primer texto; y, de los trechos del segundo texto que enfocaban específicamente el BPLC, se eliminaron 2 referencias, por repetición , sumándose 9 obras al *corpus*: un subtotal de 77.

La intención fue profundizar la discusión del barrido por el lado del cliente, considerando el bajo número de resultados de la primera fase, probablemente decurrente de la "novedad" del tema, que inviabiliza el enfoque académico de más fuerza, en monografías, disertaciones y tesis.

Así como en el informe anterior, se admite que la sistematicidad del estudio se reduce por ese paso, lo que afecta la representatividad de los resultados en función de la subjetividad de la selección por el equipo de pesquisa. Sin embargo, se recibió esa desventaja fue recibida como compensada por la ganancia de subsidios para reflexión sobre el tema, y atenuada por tratarse de un procedimiento en grupo, con revisión ciega por los integrantes del equipo.

Las 110 referencias fueron, entonces , sujetas a evaluación preliminar de pertinencia temática y formal. Título, resumen (si presente) y sección inicial fueron leídos por dos investigadores, que votaron por la inclusión o exclusión. En esa fase, para reducir sesgos, los investigadores no tenían acceso a los votos uno del otro. Existiendo disenso, la decisión era del tercer investigador, que tampoco sabía cómo habían votado los demás investigadores.. Además de las obras que no trataban propiamente del tema del BPLC (49) y de la barrera del idioma (1 texto en alemán), se deliberó por excluir del *corpus* obras no académicas: presentaciones de pantallas (4), artículos de opinión (3), código de programa (1), currículum (1), entrevista (1), lista de obras (2), noticias periodísticas (5), palestras (3), política interna (1), pronunciamiento oficial (1), propuestas normativas (2) y publicaciones de *blog* (14).

Las 22 obras restantes fueron entonces integralmente leídas, analizadas e inseridas en un formulario, con categorización (artículo, disertación, capítulo de libro etc); resumen; observaciones del investigador responsable; y citaciones en destaque. Se generó, entonces, una síntesis descriptiva, orientada a identificar, en cada obra: propuesta, metodología (o su ausencia), eventuales referencias relevantes (citadas como base para el concepto o posicionamiento del trabajo), y cuál el enfoque sobre el barrido por el lado del cliente.

Resáltese la limitación del impacto de las fases 2 y 3 sobre el resultado final: la mayoría de las obras de hecho analizadas (68,18%) advino de la búsqueda por palabras-llave, método no-discrecional. Esa particularidad – montante de textos seleccionados según un dato método, o inseridos por vía discrecional – mitiga el daño a la sistematicidad y preserva la replicabilidad: se puede reproducir el estudio sin esas obras, o en la íntegra.

Así, el *corpus* documental final abarca 22 publicaciones, listadas en el Apéndice 1: 3 obras publicadas en anales de eventos, 7 artículos científicos, 3 notas técnicas y 12 informes.

Adicionalmente, el informe cuenta con referencias bibliográficas que no compusieron el *corpus de pesquisa*, por su pertinencia en la contextualización del momento de emergencia del debate sobre el BPLC en torno del anuncio de Apple, en 2020, o en la explicación del funcionamiento de las técnicas de detección de contenido, o, aún, en la fundamentación de opciones metodológicas. Sin embargo, esas referencias adicionales no han compuesto el objeto del análisis que buscó responder a la pregunta central de este estudio.

3. Resultados

Los resultados de la revisión sistemática de la literatura seleccionada se dispusieron en cuatro ejes: contexto; concepto; funcionamiento y problemas.

3.1. Contexto

La expansión del acceso a Internet por el mundo está acompañada de un aumento en los índices de prácticas ilegales online. Entre ellas, el compartir de material ilícito, a ejemplo de contenido de abuso sexual infantil – CSAM, es tópico de constante preocupación para autoridades públicas. En ese sentido, varias técnicas y herramientas investigativas categorizan materiales ilícitos online, hacen moderación, identifican remitentes, y providencian la persecución de involucrados.

En ese contexto, se presentan a seguir algunas técnicas para detección de CSAM, a limitación de su aplicación en sistemas con criptografía y como Apple ha pretendido presentar soluciones que impulsaron el debate sobre el BPLC.

3.1.1. Técnicas de detección de CSAM

Una vía simple es el análisis de texto, para reconocer palabras-llave (por la mera forma) o, con más complejidad, reconocer patrones (por el significado), en el procesamiento de lenguaje natural, cuyo tenor indique seducción o extorsión . Hay *aprendizaje de máquina* para detectar atributos indicativos de CSAM en texto, imagen y video, buscando patrones y con atribución de valor semántico a los datos. Y, aún, hay la clasificación de CSAM por medio de *modelos de correspondencia*, que atribuye identificadores (*hashes*) al material analizado y permite a las plataformas evaluar imágenes y videos compartidos, mediante el cotejo – por verificación criptográfica de identidad o por verificación perceptiva de semejanza (ver 5.3) – del *hash* atribuido al contenido en análisis con la base de *hashes* previamente atribuidos a contenidos ilegales ya conocidos.

Las herramientas API Content Safety de Google,[14] PhotoDNA de Microsoft,[15] Safer de Thorn,[16] y PDQ Hash de Meta[17] son soluciones de detección automatizada usados en redes sociales, foros, bancos de imágenes y videos muy populares:[18] *hash* perceptivo impide el retorno de perfiles banidos del servicio de relación social OkCupid; SimSearchNet/ SimSearchNet++ verifica correspondencia con material de desinformación, incluso sobre COVID-19, en imágenes en Instagram y en Facebook; e identificadores *hash* de audio evitan el accionamiento involuntario de “Alexa” por anuncios.

Tales técnicas de detección actúan en contenido almacenado en los bancos de datos de las plataformas, o por ellas transmitido, siendo denominadas de *server-side scanning* – SSS (en español, “barrido por el lado del servidor” – BPLS).[19] Son soluciones prontamente aplicables a sistemas sin criptografía de punta a punta, con contenido accesible y comprobable por plataformas y autoridades estatales.

3.1.2. Detección en sistemas con criptografía

Pero la aplicación de esas técnicas BPLS no es trivial en ambientes digitales con criptografía asimétrica, en los cuales los participantes de la comunicación usan protocolos de llaves públicas de criptografía y llaves privadas de decodificación. Bajo tal seguridad criptográfica, por definición, solo se puede conocer el contenido por los legítimos remitentes y destinatarios, salvo por la inserción de eventual vulnerabilidad en los algoritmos criptográficos, medida que el Alto Comisariado de las Naciones Unidas apunta, desde hace años, como fuente de graves riesgos a garantías fundamentales, como la privacidad, la libertad de expresión y de reunión pacífica, la protección de datos personales, entre otras.[20]

En septiembre de 2020, se divulgó sin autorización una minuta interna de la Comisión Europea[21] sobre las dificultades de detectarse CSAM en comunicaciones criptografadas de punta a punta, manteniendo privacidad y seguridad. Sin hablar de los aspectos políticos decurrentes, el borrador describió medios tecnológicos para esa identificación

(puertas clandestinas, barrido por el lado del cliente, enclave seguro, clasificación, computación homomórfica, etc.) y los analizó en cinco criterios: eficacia, viabilidad, privacidad, seguridad y transparencia.

Tales medios de detección de CSAM se clasificaron según el elemento básico de las comunicaciones codificadas que afectan: la técnica, el servidor, o el dispositivo del usuario.[22] La detección en la *técnica criptográfica* interfiere en el protocolo, por medio de la llamada "criptografía" homomórfica, que todavía se puede combinar con aprendizaje de máquina y clasificadores, aplicado integralmente en los servidores, creando *hashes* para contenido codificado antes del envío, de acuerdo con las técnicas del cuadro a seguir:

MEDIOS DE DETECCIÓN DE CSAM EN SISTEMAS CON CRIPTOGRAFÍA

ELEMENTO AFECTADO	TÉCNICA	SERVIDOR	DISPOSITIVO
MODALIDAD UTILIZADA	"criptografía" homomórfica	a) enclave en el servidor de la plataforma b) enclave en el servidor de un tercero	a) todo en el dispositivo; b) *hash* en el dispositivo y verificación en el servidor c) parte de *hash* en el dispositivo y parte en el servidor, y verificación en el servidor d) clasificadores en el dispositivo.

Según el cuadro, las técnicas de detección *en el servidor* pueden realizar algunas o todas las operaciones sin criptografía de punta a punta de BPLS en un reducto, o enclave seguro a) en el servidor de la plataforma, o b) en el servidor de un tercero; c) en el servidor de múltiples terceros. Ya las técnicas de detección *en el dispositivo* podrían realizar algunas o todas las operaciones sin criptografía de punta a punta del BPLS, siendo categorizadas en cuatro modalidades: a) integralmente realizadas en el dispositivo; b)

identificación integral en el dispositivo y correspondencia en el servidor; c) identificación parcial en el dispositivo y identificación parcial y correspondencia en el servidor; d) uso de clasificadores en el dispositivo. Ese último conjunto recibe el nombre de *client-side scanning* – CSS, o "barrido por el lado del cliente" – BPLC.

La comunidad técnica había criticado los presupuestos y las consideraciones de la minuta europea filtrada en 2020,[23] que se proponía a un análisis técnico de las posibilidades de detección, sin problemas y sin pasar por las cuestiones políticas y jurídicas involucradas.[24] En especial, se acordó de que en julio de 2020 había sido aprobada por el Parlamento Europeo una norma provisoria para el tema, con vigencia de 3 años, autorizando el escaneo voluntario del contenido de comunicaciones como esfuerzo proactivo de las plataformas digitales – regla que ya había sido criticada por contrariar previsiones normativas de la misma Unión Europea.[25]

Hasta entonces, varias modalidades alternativas a la rotura de la codificación se venían cuestionando en sus impactos sobre el sigilo y la seguridad de las comunicaciones. Con el anuncio de Apple, el BPLC ha tomado el centro del debate:[26]

> *En agosto de 2021, Apple anunció planes para introducir tal sistema para sus servicios iMessage e iCloud, pero suspendió la implementación del cambio propuesto después de fuertes críticas de una amplia gama de especialistas en seguridad de tecnología de la información, criptógrafos y grupos de derechos humanos.*

Aunque la propuesta de Apple no sea el objeto central de esta pesquisa, se hace oportuno narrar en detalles los acontecimientos entre la presentación y la suspensión de la propuesta que ha proporcionado tanta controversia en torno al BPLC.

3.1.3. Apple y la protección de la infancia

El 5 de agosto de 2021, Apple anunció que en aquel año todavía año adoptaría tres cambios [27] en sus sistemas operacionales (iOS 15, watchOS 8, iPadOS 15 y macOS Monterey), a fin de optimizar el enfrentamiento a materiales de abuso sexual infantil, o CSAM (sigla para el término en inglés *Child Sexual Abuse Material*).[28]

Primero, la "**Seguridad de las comunicaciones en mensajes**"[29] interferiría en el dispositivo de niños y adolescentes, "empañando" las imágenes con tenor sexual explícito recibidas o enviadas vía iMessage. La app entonces pediría una confirmación para permitir la visualización. Específicamente para niños, con menos de 13 años, si la familia optase por ser notificada, iMessage exhibiría en el celular un segundo pedido de confirmación, con el aviso al niño de que se comunicaría a la familia, pero sin exponer la imagen.

Con el segundo cambio, "**Detección de CSAM**", antes de que se enviaran a iCloud Photos, las imágenes serían convertidas en *hashes* y, en el dispositivo, comparadas con los *hashes* de materiales de abuso sexual infantil de un banco de dados suministrado por la ONG estadounidense NCMEC – *National Center for Missing and Exploited Children* ("Centro Nacional para Niños Desaparecidos y Explotados").

Por fin, con la "**Orientación ampliada en Siri y en la B**úsqueda", la empresa añadiría comandos que exhibirían orientaciones de seguridad e informaciones de contactos para denuncia, cuando hubiera pesquisas, por texto o voz, involucrando materiales de abuso sexual infantil.

Las dos primeras medidas actuarían tanto en el dispositivo como en el servidor: la base de datos de los *hashes* ilícitos sería embarcada en el sistema operacional de los aparatos de la fabricante, lo que posibilitaría la identificación; y posteriormente se haría la verificación de correspondencia en los servidores de NCMEC.

Hubo muchas críticas,[30] además de algunas cartas abiertas solicitando que no se adoptaran los cambios.[31] Apuntadas como "puertas clandestinas" (*backdoors*), las dos primeras herramientas, para iMessage y iCloud Photos, fueron acusadas de fragilizar la seguridad prometida por la adopción de criptografía segura. La empresa reaccionó que los riesgos de los nuevos productos serían mínimos en términos de privacidad, por no exponer, ni a Apple, ni a terceros, las comunicaciones de los usuarios: los responsables solo sabrían de la existencia de riesgo, pero no del contenido, que no se archivaría en ningún otro local, además del propio dispositivo que envió o recibió la imagen.

Las críticas siguieron y, el 3 de septiembre , Apple comunicó que adiaría la adopción de la herramienta de detección de CSAM. El anuncio fue celebrado, pero no encerró pedidos de que se abandonaran completamente los "planes de vigilancia".[32] Un año después, en octubre de 2022, la sesión "Child Safety" en el site oficial[33] informa que la herramienta de seguridad en iMessage es opcional y no comunica a la familia ni cualquier tercero, ofreciendo al niño o adolescente la posibilidad de buscar ayuda de alguien de su confianza; y que, para la tercera herramienta las orientaciones e informaciones están también en Spotlight (sistema de busca en el desktop). No hay ninguna mención más al mecanismo de detección de CSAM, cuya adopción, aparentemente, sigue suspensa sin plazo previsto para retornar la agenda de la empresa.

El caso de Apple ilustra la cuestión central de este estudio. A pesar de críticas o elogios a la propuesta específica o a la postura de la empresa, merece comprensión la idea de detectarse contenidos indeseados a partir de los dispositivos de los usuarios, a fin de viabilizar medios de investigación en ambientes codificados. ¿Pero cómo se puede definir el concepto BPLC? ¿Cómo funcionaría esa tecnología? ¿Y cuáles los riesgos?

3.2. Concepto de BPLC

En su universo, las herramientas de análisis automatizado de contenido se pueden clasificar en dos tipos de modelos: de correspondencia o de predicción.[34] El primer modelo se puede realizar por verificación criptográfica o preceptiva, con el fin de reconocer un contenido como idéntico o semejante; mientras tanto, en el segundo modelo, se pretende reconocer las características de un contenido aprendido previamente por la máquina. Ambos modelos pueden operar tanto en barrido por el lado del servidor como por el del cliente.

<table>
<tr><th colspan="3">MECANISMOS PARA ANÁLISIS AUTOMATIZADO DE CONTENIDO</th></tr>
<tr><td colspan="2">modelos de correspondencia

reconocer el contenido como idéntico o suficientemente semejante al contenido visto anteriormente</td><td rowspan="2">modelos de predicción
reconocer características del contenido con base en aprendizaje de máquina previo</td></tr>
<tr><td>Verificación criptográfica

Identificación de correspondencia a partir de hash altamente sensible a alteraciones</td><td>Verificación perceptiva

Identificación de correspondencia a partir de hash por determinado grado de semejanza</td></tr>
<tr><td colspan="2">barrido por el lado del servidor: contenido se convierte en hash y es enviado a la plataforma para evaluación

barrido por el lado del cliente: contenido se convierte en hash y evaluado en el propio dispositivo</td><td>clasificación por algoritmos por el lado del servidor

clasificación por algoritmos por el lado del cliente</td></tr>
</table>

El **barrido por el lado del cliente**, aquí abreviado como BPLC (muchas veces citada en el término en inglés *client-side scanning*, con las iniciales CSS) – también citado como "correspondencia de *hash* perceptivo" (*perceptual hash matching*)[35] "escaneo del cliente" o "filtraje en la punta"[36] – se presenta como una opción alternativa al acceso excepcional, que depende de la cooperación con la plataforma.

Se puede definir el BPLC así:[37]

> *[...] el concepto de que, a través de ciertas formas de implementación tecnológica, un sistema se podría desarrollar para digitalizar fotografías y mensajes antes de que se las envíen desde un usuario (o después de recibidas por otro usuario) a fin de determinar si las imágenes o mensajes en cuestión violan prohibiciones legales.*

De esa forma, el cerne del BPLC es el filtraje o la moderación de contenido en el ámbito del dispositivo tecnológico particular, en un aparente carácter menos invasivo, que se pretende compatible con la criptografía de punta a punta y la protección de datos personales.

Así, el BPLC consiste en el análisis de correspondencia de datos digitales, tanto en flujo (saliendo o llegando) o en almacenamiento , realizada en el dispositivo, como parte de un sistema de comunicación, codificado o no. En ese sentido, el análisis de correspondencia se realiza con una lista de contenidos "indeseados", por efectiva ilicitud o mera incompatibilidad con la política del intermediario.

Es importante destacar que el barrido por el lado del cliente consiste en una propuesta tecnológica de reconocimiento de contenido que se pretende adoptar por las plataformas, y no en una funcionalidad ya existente en los dispositivos de comunicación. Los efectos de la adopción de esa propuesta se tornaron una parte sensible de la controversia. Se debate intensamente sobre los métodos disponibles para su concretización, como también su eficacia para la finalidad pretendida, los riesgos de seguridad decurrentes y la compatibilidad de la medida con los atributos de la criptografía de punta a punta.

El BPLC se distingue del Barrido Por el Lado del Servidor (BPLS), en la medida en la que esta última realiza el análisis de contenido durante la transmisión o a partir del almacenamiento por el intermediario del sistema de comunicación, de modo que se amplía el riesgo a la privacidad del usuario.[38] Por no almacenar, ni intervenir en la transmisión del contenido, el BPLC fue visualizado por Apple como una forma de combate a contenidos ilegales, en específico CSAM, sin la rotura de la privacidad del usuario y de la criptografía de punta a punta.[39]

Se pueden identificar seis fases del BPLC aplicado a la moderación de contenido, que pueden ocurrir sucesiva o simultáneamente:[40] 1) definición de lo que es o no permitido en el servicio; 2) detección del contenido generado por el usuario potencialmente irregular a la luz de políticas internas o de la legislación; 3) evaluación de la irregularidad del contenido; 4) intervención contra el contenido identificado como irregular; 5) recurso contra la decisión de intervención; y 6) educación sobre la política de moderación de contenido.

El diferencial optimista de BPLC sería auxiliar empresas de tecnología a actuar contra la pornografía infantil y otros tipos de comunicaciones ilícitas, pero sin el costo de degradar los sistemas de criptografía fuerte.[41] BPLC supuestamente podría viabilizar el combate al contenido indeseado, "*pero también mantener la privacidad del tel*éfono *de los usuarios antes de ser codificado*",[42] incluso con la premisa de que el análisis en el dispositivo sería sistemáticamente más eficiente "*de lo que proveer acceso excepcional para aplicación de la ley*"[43]

Aún, en el BPLC parecería ventajosa la "*oportunidad para activistas y otros detectar alteraciones en el software del lado del cliente y entender sus efectos, añadiendo alguna transparencia y responsabilidad*".[44] Sería posible capacitar usuarios para entender y controlar su experiencia online, sin apenas abrir las puertas para la censura automatizada.

3.3. Funcionamiento

El ejemplo más difundido actualmente tuvo como justificativa la búsqueda por contenido de exploración sexual infantil. En estos casos, el BPLC compara el material con la lista de *hashes* y, si hay correspondencia, el sistema puede no enviar el mensaje, relatar la tentativa (a las autoridades públicas o a una organización de la sociedad civil), o combinar esas acciones.[45] El primer paso para el funcionamiento de tal recurso es la comparación de un contenido "contra una base de datos predefinidas de contenidos dañosos, los cuales estarían señalizados con identificadores únicos",[46] los llamados *hashes*.

Las funciones de *hash* se proyectan para convertir un conjunto de datos (o entrada), como una imagen o un archivo de texto, en un conjunto corto de caracteres de tamaño estándar (o salida), chamado *hash*, algo como una "impresión digital" de un determinado contenido.[47] Si implementadas a través de criptografía, tales funciones agregan atributos, a ejemplo de la inviabilidad de dos entradas distintas convertirse en una misma salida.[48]

La finalidad es detectar correspondencia no solo entre copias exactas, sino también entre medios semejantes (por ejemplo, una imagen en versión redimensionada):[49]

> *Los hashes perceptivos son diferentes de los hashes criptográficos, pues el primero cambia gradualmente a medida que la imagen cambia, mientras el último cambia significativamente cuando un único pixel cambia. Es importante resaltar que los hashes perceptivos se proyectan para detectar instancias de medios visuales que son visualmente semejantes (por ejemplo, una versión redimensionada) sin ser copias exactas.*

A lo largo de los últimos años inúmeras herramientas se desarrollaron a partir del filtraje del tráfico y del contenido de la web, con fundamento en la existencia de determinadas palabras-clave, metadatos o patrones pre-establecidos.[50] Por ejemplo, para detectar imágenes asociadas a material protegido por derechos autorales o pornografía infantil, como los algoritmos de correspondencia de *hash*, "ellos identifican las imágenes por un código único – una especie de "huella dáctil" para una determinada imagen – llamado hash, y las comparan con el hash de imágenes conocidas con derechos autorales o pornografía infantil".

Por lo tanto, en el caso de Apple, las imágenes de los usuarios serían convertidas en *hashes* por el sistema NeuralHash y comparadas con los *hashes* de contenido del banco de datos de materiales de abuso sexual infantil, mantenido por la ONG Centro Nacional para Niños Desaparecidos y Explotados de los EEUU y por otras organizaciones de seguridad infanto-juvenil:[51]

> *Un sistema de detección llamado NeuralHash crea identificadores que pueden ser comparados con IDs del Centro Nacional para Niños Desaparecidos y Explotados y otras entidades para detectar contenido conocido de CSAM en las bibliotecas de fotos de iCloud. La mayoría de los proveedores de nube ya verifica las bibliotecas de usuarios para esas informaciones – el sistema de Apple es diferente porque hace la correspondencia en el dispositivo y no en la nube. (traducción libre)*

Se ejemplifica como la tecnología funcionaría en el caso de la verificación de un archivo codificado conteniendo imágenes de abuso sexual infantil por una aplicación de comunicación: "Se compararía la imagen a una lista de imágenes ilegales conocidas y vetadas antes que se enviara la imagen".[52] En el caso de WeChat, por ejemplo, la censura de contenidos en general ocurre en tiempo real.[53]

En el caso de Apple,[54] para protección de la privacidad del cliente, el sistema trabajaría con una correspondencia de *hashes*, en lugar de digitalizar las imágenes en la nube. El mero hecho de encontrar correspondencias entre los materiales, no generaría – a priori – ninguna información a ser prestada a la empresa, ni al usuario y tampoco a las instituciones de responsabilización. Sin embargo, se generaría un registro para acompañamiento de otros materiales en caso de reincidencia de CSAM, a partir de los denominados *vouchers* informacionales, o "recibos". De este modo, se adoptó la técnica *threshold secret sharing*, que solo permite tratar los datos después de ser excedido, por parte de un mismo usuario, un límite de correspondencias previamente establecidas.

Se almacenaría este recibo con la imagen en la cuenta de iCloud Photos. En el caso de las reincidencias exceder un "límite de correspondencias" (en cantidad no divulgada por la empresa), el usuario y el material pasarían por una verificación humana, que

descodificaría solo las correspondencias de CSAM, sin dar acceso a las otras imágenes. Así, personas designadas apenas accederían a "derivados visuales" de baja resolución de las imágenes (como una imagen en baja resolución , en negro y blanco en miniatura). Si confirmado el CSAM, se suspendería la cuenta y se comunicaría a la ONG pertinente responsable, con un informe del procedimiento.

La etapa de verificación humana de los contenidos detectados previne el riesgo del falso positivo, pero no neutraliza los riesgos graves de la exposición de datos que se consideren lícitos. Como apunta el Center For Democracy And Technology,[55] el chequeo del contenido por seres humanos representa en igual medida un riesgo a la privacidad, pues todo acceso por puerta clandestina fragiliza la seguridad del usuario, que puede pasar a ser víctima no solo de ataques terroristas, sino hasta de ataques políticos institucionales. De esta manera, se verifica que un medio pensado para enfrentar violencias sexuales contra niños y adolescentes puede servir para perjudicar otros derechos humanos.

Con la finalidad de mitigar los riesgos de ataques virtuales que buscan acceder a las imágenes almacenadas con recibos, como también la alteración de los datos de un recibo almacenado, Apple ha informado que la base de datos obtenida por el cruce de las imágenes con CSAM se inseriría de forma codificada dentro del propio sistema operacional de los aparatos.[56] De esa forma, no se podría actualizar la lista de manera independiente por la internet, lo que aumentaría la seguridad del sistema.

Es importante destacar que para generar la base de los h*ashes* sería todavía necesario el cuidado con los materiales utilizados para evitar contenidos infiltrados y/o maliciosos. Para tanto, la propuesta de Apple exigía la comparación de los materiales con una lista de hashes apuntados como indeseados por dos instituciones no gubernamentales que operasen en jurisdicciones diferentes,[57] no valiéndose apenas de la base de NCMEC. Esa medida buscaba impedir el uso del sistema por autoridades estatales para fines de censura o vigilantismo por algún gobierno.

El BPLC del caso Apple se mostraba aún como una tecnología híbrida, de escaneo de cliente y servidor. Así, se puede especular que si el usuario no usara iCloud Photos estaría exento del cruce de los datos para detección de CSAM.

Para Erik Neuenschwander,[58] las herramientas pretendían responder a presiones de diversas instituciones cuanto al acceso a informaciones codificadas, bajo alegaciones de combate a actividades de terrorismo o, como es el caso, de CSAM, al mismo paso en que se protegiera la privacidad de los usuarios.

En la propuesta formulada por REIS y otros,[59] por ejemplo, el BPLC se podría utilizar para detectar, a partir de *hashes*, contenidos rotulados previamente como desinformación por instituciones de chequeo de hechos. Se podría implementar la solución en los dispositivos de quien envía, con la ventaja de detectar y limitar la distribución de contenidos indeseados; en los de quien recibe, auxiliando contra la seducción; o en ambos, viabilizando un enfoque de careo, que permitiría identificar interferencias en el sistema.[60]

3.4. Problemas

En la contramano de lo que alegó Apple, las críticas en la literatura académica acusan la defensa de la adopción del BPLC de un optimismo ingenuo y acrítico, y presentan preocupaciones de varios órdenes – ningún de ellos solucionado por las propuestas de BPLC.

Desde el punto de vista **sociológico**, se encontró el hecho de que la automatización de la búsqueda por cualquier tipo de contenido indeseado afecta solo la difusión del material, pero no elimina su fuente:[61]

> *Las soluciones tecnológicas para detectar contenido problemático por si solo no abarcarán las cuestiones más amplias de, por ejemplo, la distribución de desinformación o CSAM, que necesitan identificar y enfocar en su núcleo los problemas sociales y políticos causados por detrás de esos fenómenos.*

Se puntuó, todavía, que intereses domésticos de cada nación ocasionarían **perjuicios económicos,** en ese sentido la medida:[62] "podría minar la competitividad de servicios nacionales",[63] al reducir el interés general en los productos y servicios ofrecidos por las empresas de países cuya legislación exigiera la adopción obligatoria de BPLC.

En enfoques más profundos, las preocupaciones en las obras analizadas se pueden agrupar en dos categorías: tecnológicas y jurídicas.

3.4.1. Aspectos tecnológicos

En la visión tecnológica, se puede afirmar que "los sistemas de correspondencia de hash CSS no son técnicamente robustos".[64] Las inconsistencias no respondidas por los proponentes del BPLC se pueden ver en función del objeto de la vulnerabilidad, es decir, de cuál el atributo del propio sistema se afecta por un efecto negativo: el funcionamiento, la eficacia, la seguridad, el escopo.

Cuanto al **funcionamiento**, diversamente de los servidores de los intermediarios, la falta de estandarización del "lado del cliente" implicaría gran diversidad de posibles restricciones para el mejor empleo del BPLC. Es decir, el mecanismo de barrido en si no podría siquiera operar por una baja capacidad de almacenamiento o procesamiento en el dispositivo cliente sea por falta de actualización de los softwares; por una mala conexión con Internet; o por falta de carga en la batería.[65]

Específicamente, la construcción y actualización del banco de datos de contenidos indeseados exigiría una comunicación entre servidores y dispositivos potencialmente

falla. Según Hua y otros,[66] "la escala de los bancos de datos torna prohibitivo tanto enviar los hashes conocidos como malos para el cliente como, si los hashes son sensibles, aplicar técnicas 2PC para garantizar lo mínimo posible sobre las filtraciones de banco de datos para los clientes".[67]

Aunque si el sistema opere correctamente, puede fallar todavía en la **eficacia**: contenidos nuevos o comunicados por primera vez no se detectarían como irregulares, por requerir previa colección y rotulación[68] para el reconocimiento por correspondencia. En ese sentido, el contenido indeseado necesitaría circular más de una vez para que el BPLC sea eficaz – y esa no es la regla. Por ejemplo, de todas las imágenes de material de abuso sexual de niños delatadas en los EEUU, el 84% fueron denunciadas solo una vez[69] – agregándose todavía la capa de sub-notificación decurrente de la limitación de recursos humanos y de la capacidad manual para revisar todas las denuncias, cuyo número creció de diez mil al año en 1998 para casi un millón al mes en 2017.[70]

En sus limitaciones, el BPLC no propone analizar el significado o el contexto, siendo indiferente a situaciones de uso legítimo de obras protegidas por derechos autorales, o eventual ausencia de intención criminosa[71]. Hasta por eso, se notó fácil, para cualquier usuario del sistema de comunicación, inutilizar el sistema automatizado de BPLC, al neutralizar o contaminar contenidos por la manipulación de datos digitales con técnicas de gradiente y transformación, pero sin afectar de modo relevante su percepción por seres humanos.[72]

Es decir, cambios de luminosidad, brillo o inserción de leves ruidos generarían falsos negativos, al molestar la correspondencia al *hash* indexado; o se podrían modularlos para generar falsos positivos, forzando disparo de alarmes falsos con la correspondencia de *hashes* en materiales inocuos. Se apunta como especialmente preocupante que, a pesar del grado de transparencia sobre el funcionamiento del algoritmo de generación del *hash* (tratándose de una "caja -negra"), la necesidad de ampliación de los límites de detección (para contornar tentativas deliberadas de evitar el reconocimiento de contenidos indeseados) permitiría la generación de muchos falsos positivos.[73] Se dice que reconocer contenidos por semejanza abre el riesgo de falsos positivos, con la detección equivocada de una correspondencia en el sistema de *hashes* perceptivos.[74]

Así, como menos el sistema sea sensible a cambios sutiles (luminosidad, color, sombras, inversión etc.) – es decir, como menos considere ruidos, a fin de detectar la correspondencia entre imágenes semejantes con pocas diferencias –, más el sistema se sujeta a dos problemas. Primero, esa insensibilidad permite más falsos positivos, es decir, identificar como correspondiente un contenido que en verdad no corresponde; y, segundo, permite todavía a adversarios provocar una elevación artificial de la demanda por el análisis de imágenes inofensivas, alteradas intencionadamente apenas en la medida en la que se produzca un *hash* semejante al de imágenes previamente reconocidas como indeseadas.

También el propio volumen creciente de compartir material de exploración sexual infantil puede representar una dificultad para el procesamiento tempestivo de las denuncias.[75] No sería una tarea trivial definir el nivel aceptable de falsos positivos adecuado para la implementación del BPLC en un sistema de comunicación con criptografía de punta a punta.[76]

E incluso si funciona el sistema y es efectivo, en lo que se refiere a la **seguridad**, en tesis, el banco de datos de *hashes*, eje central de operación del sistema, puede ser fácilmente manipulado por agentes mal-intencionados.[77]

Aunque una postura más transparente por parte de Apple pudiera generar más confianza al usuario,[78] de manera general, en cualquier sistema, la suma de componentes agrega más potencial de fallas de seguridad.[79] En el caso del BPLC, aunque haya o no interferencia directa en la criptografía, "hay necesaria reducción de la seguridad del sistema a causa de la ampliación de la superficie de ataque",[80] de tal modo que especialistas temen "riesgos similares a los del acceso excepcional vía backdoor",[81] a saber, "abuso, el efecto inhibitorio y los daños a la confianza en el ecosistema digital".[82]

Incluso la defensa del BPLC contra asedio o desinformación[83] advierte que el *hash* perceptivo "puede no ser adecuado para todas las clases de contenido abusivo, como CSAM, en el que el destinatario puede ser un adversario".[84] Esos modelos – al operar por semejanza– fueron diagnosticados, en el análisis de la herramienta Neural Hash, por ejemplo, como "altamente susceptibles a varios ataques, algunos de ellos triviales, que rompen el sistema".[85] Grover y otros destacan que el BPLC "confía en el dispositivo del usuario final para calcular con veracidad el hash del mensaje".[86] Por ese modo de construcción, se afirma[87] que, en ambiente adversos, los medios de BPLC son tan inseguros y vulnerables que ni Apple, con todo su esfuerzo de ingeniería , consiguió ofrecer un proyecto técnico confiable[88] o a prueba de brechas en sus sistemas.[89] Es sintomático que un año después el auge de la controversia, hasta noviembre de 2022, la gigante de tecnología mantenga todavía la decisión de suspender la adopción de la herramienta de detección de CSAM.

A propósito, si para el iOS, Apple controla la tecnología desde el *hardware* hasta el *software*, en el ecosistema Android numerosas empresas diferentes producen los aparatos celulares y varían mucho las versiones activas del sistema operacional móvil. Aún así, se verificó no haber diferencias relevantes entre Android e iOS en varias dimensiones relacionadas a la privacidad del usuario, incluso por la ausencia de una reglamentación estatal: la nivelación de la seguridad ocurre por bajo.[90] Por lo tanto, delante de la enorme cantidad de dispositivos en uso, de muchas marcas y modelos, se pueden esperar futuras fallas en la actualización de los programas de detección, perjudicando la seguridad del funcionamiento del mismo mecanismo.[91]

Contra el optimismo de que sería posible "encontrar una solución simple para el problema"[92] de la difusión de contenidos nocivos como material de abuso sexual

infantil, la reticencia general de especialistas a las supuestas propuestas alternativas[93] se extiende al barrido por el lado del cliente: “una percepción frecuente fue de que habría un comprometimiento principiológico de la criptografía incluso sin una interferencia directa en el algoritmo criptográfico o en el sistema de gestión de llaves”.[94] La promesa de vigilancia limitada se tiene por ilusoria: titulares de los datos personales analizados no podrían prever ni auditar la acción de autoridades.[95]

Finalmente, aunque que funcione bien, con eficacia y seguridad, se apunta el riesgo de que el BPLC tenga su **escopo** ampliado para otros propósitos, yendo más allá del enfrentamiento material de exploración sexual infantil,[96] para abarcar otros tipos de contenidos, tales como desinformación,[97] terrorismo o evasión de documentos estatales,[98] y cuya ilegalidad puede no ser tan evidente o libre de discusión.

Se viabilizaría , por ejemplo, eventual censura de mensajes políticas legítimas.[99] En el peor de los escenarios, por la ampliación de la lista de contenidos indeseados, “*la totalidad del diccionario podría incorporarse a esa base, efectivamente posibilitando el desciframiento total de los mensajes y nulificando el propósito de la criptografía*”.[100] Y cabría a los intermediarios la tarea de resistir a la presión por expansión o abusos.[101]

No se debe descuidar de la real posibilidad del objeto de la moderación ser cuestionable, pues la naturaleza del BPLC no garantiza ni exige la legitimidad en el objetivo de la aplicación, como se afirma consensualmente en relación a la exploración infantil: “Hashes para otro contenido sensible pero legal (como político o sexual) se pueden añadir al banco de datos y sin el conocimiento del usuario”.[102]

3.4.2. Aspectos jurídicos

Jurídicamente, hay una forma de pensar la cuestión “que reconoce las estructuras técnicas de los sistemas criptográficos como indisociables de las connotaciones políticas que adquirieron a lo largo de los años en lo que se refiere a la defensa de los derechos humanos”,[103] y serían sometidos a altos riesgos de graves amenazas puntuales o incluso a restricciones generales y sistematizadas.

A respeto del **sigilo de las comunicaciones y de la privacidad**,[104] se afirma que, “si los resultados de la comparación de hash se comparten con el servidor, las garantías de privacidad de la criptografía de punta a punta se violarán”.[105]

Así, en afrenta al principio de la **presunción de inocencia**, el BPLC “deteriora la finalidad de la criptografía de punta a punta relacionada a la libertad de información y expresión, pues el contenido de la comunicación se filtra por patrón”.[106] Profundizando la cuestión tecnológica del escopo, desde el punto de vista político, se abre una preocupación con otras modalidades de moderación, incluso las que revelan menos cuidado con excesos o abusos. Al debilitar los argumentos en favor de servicios con criptografía de punta a punta,[107] la adopción del BPLC “acaba siendo un trampolín para mecanismos de backdoor/

censura más arriesgados , porque una vez implantado el primero, será más fácil implantar o justificar la implantación del último".[108] E incluso si fuera efectiva la contención de riesgos en una democracia, su implementación podría legitimar herramientas semejantes en países habituados a censurar y vigilar.[109]

En la **seguridad pública** de la colectividad, si "*es posible para un actor ingenioso adivinar el contenido de un mensaje a partir de su hash*",[110] organizaciones criminosas podrían usar el sistema para ilícitos. Delante de la viabilidad de prácticas de ingeniería reversa, la existencia de un sistema con garantías de protección no pueden depender de la confianza en la ética de quien opera el sistema de detección de contenido, pues no cabe exigir ética en ambientes adversariales, tales como en el contexto del enfrentamiento a la CSAM.

Así, el BPLC serviría como "un **modelo para vigilancia masiva**, pues puede no ser posible para el usuario o la sociedad civil monitorear la lista de hash usada por su teléfono para garantizar que esté apenas relatando o impidiendo la transmisión de imágenes de abuso sexual infantil"[111] . Y, más grave, delante de la amplitud, vigilando datos privados "de todo el mundo, todo el tiempo, sin orden judicial o sospecha",[112] no sería posible asegurar la aplicación estrictamente regular y confiable, incluso en relación a la infancia y juventud, pues "esos riesgos afectarán todos los usuarios de plataformas de comunicación digital; todos los niños y todos los adultos del mundo, ahora y posiblemente en el futuro, y las consecuencias son difíciles de prever".[113]

Así, considerando los derechos humanos como un todo, "esa vigilancia en masa puede resultar en un significativo efecto atemorizante en la libertad de expresión y, de hecho, en la propia democracia".[114] Y para la sociedad civil organizada, también "la posibilidad de desvirtuación de función del sistema representa un riesgo democrático".[115] Luego, a partir de una visión de pros y contras en confronto,[116] la gama de riesgos sería **desproporcional** a los potenciales beneficios, como también **innecesaria** para los objetivos pretendidos.[117]

4. Conclusión

En este segundo informe,[118] se evaluó el escenario del **barrido por el lado del cliente** (muchas veces referenciado por el término en inglés *client-side scanning*, o las iniciales CSS, aquí abreviada como BPLC). El término se refiere al uso de técnicas de escaneo de los dispositivos de usuarios (clientes) para identificación de instancias de compartimiento de materiales considerados ilícitos en ambientes protegidos por criptografía segura, en lugar de realizarse ese escaneo a nivel de servidor. Por medio de una revisión sistemática, se investigó un total de 22 publicaciones seleccionadas.

El primer aspecto notorio sobre las controversias y críticas relacionadas a técnicas de BPLC se refiere a **aspectos tecnológicos**. Las preocupaciones de esa naturaleza apuntadas a lo largo del estudio se pueden resumir en cuatro elementos analizados a partir de las soluciones presentadas y propuestas, hasta el momento de la redacción de este informe: funcionamiento, eficacia, seguridad y escopo.

Primeramente, cuanto al **funcionamiento** de esas soluciones, se constata que la primera barrera para la viabilidad del BPLC se refiere a la imposibilidad de emplearse esa tecnología en cualquier hardware al nivel de cliente. Por desplazar el procesamiento de la comparación de *hashes* de los servidores de las empresas proveedoras para los aparatos de los usuarios finales, cuestiones como la capacidad de almacenamiento y procesamiento de esos dispositivos, de la opción del usuario por no actualizar su sistema operacional o incluso de la obsolescencia del hardware imposibilitar esa actualización, entre otras, se tornan obstáculos para el uso del BPLC de manera verificablemente amplia por autoridades públicas.

Cuanto a la **eficacia**, se observa que el uso de técnicas de *hash* perceptivo – propuestas en la gran mayoría de las soluciones de BPLC hasta el momento – necesitan una base de *hashes* correspondientes al contenido ilegal que se quiere identificar para que puedan funcionar, visto que dependen de una comparación entre el material compartido por los usuarios y esa base original de *hashes* ilícitos. La composición de esas bases de *hashes* depende de denuncias iniciales y de la subsecuente constatación de la ilicitud del contenido vehiculado. Esto, a su vez, genera preocupaciones como consecuencia de la constatación de que contenido de abuso sexual infantil (CSAM) – blanco de parcela significativa de las soluciones propuestas de BPLC – es, en la gran mayoría de las veces (84%), denunciado una única vez.

Aún en la cuestión de la eficacia, se apunta que la definición de la sensibilidad de las técnicas de BPLC a alteraciones en el contenido analizado generan preocupaciones relevantes independientemente del grado de sensibilidad atribuido al algoritmo. Niveles más altos de sensibilidad hacen con que sencillas alteraciones en el contenido vehiculado resulten en la atribución de *hashes* diferentes a contenidos esencialmente idénticos – pero que fueron alterados a través de recortes, ajustes de saturación , color, entre otros. Concomitantemente, niveles más bajos de sensibilidad del algoritmo posibilitan la adulteración de contenidos inofensivos para atribuir a ellos *hashes* idénticos a los de contenidos marcados como ilícitos, posibilitando la activación de falsos positivos en los sistemas de comparación de esos *hashes*, en especial por agentes mal-intencionados.

Ya en el aspecto **seguridad**, se constata que las bases de *hashes* ilícitos pueden ser fácilmente adulteradas por agentes mal-intencionados, representando así una ampliación en la superficie de ataque de sistemas criptográficos. Eso, a su vez, añade al algoritmo criptográfico riesgos similares a los relacionados a la inserción de mecanismos de acceso excepcional (*backdoors*) en esos sistemas, que pueden ser usurpados por terceros no autorizados para obtener acceso a todo el sistema.

Por fin, cuanto al **escopo** de las técnicas de BPLC, se evidencia la posibilidad de abuso de esas herramientas por parte de autoridades públicas o incluso de las mismas plataformas que las administran, a fin de identificar y reprender instancias de compartidos de contenido por motivos ideológicos, políticos, socioculturales, entre otros. La posibilidad de ampliación del escopo del contenido rastreado por el BPLC representa

un riesgo significativo – en especial para comunidades y populaciones marginalizadas y perseguidas –, lo que se opone diametralmente a las expectativas de seguridad de la información y libertad de expresión que se busca proteger a través del uso de algoritmos criptográficos en un primer momento.

La segunda dimensión notoria analizada se refiere a los **aspectos jurídicos** en los cuales las técnicas de BPLC implican. Los apuntes aquí traídos sobre el tema se resumen en repercusiones para: la privacidad y el sigilo de las comunicaciones, la presunción de inocencia, la seguridad pública y, por fin, la proporcionalidad y la necesidad.

Cuanto a la **privacidad** y el **sigilo de las comunicaciones**, se constata que las garantías de privacidad y sigilo de la criptografía de punta a punta se violan en casos en los que los resultados de la comparación de *hashes* se compartan con el servidor. Ese compartido, sin embargo, es necesario para que sea posible una verificación humana del material apuntado como ilícito por el algoritmo, para evitar la penalización de falsos positivos.

Em lo que se refiere a la **presunción de inocencia**, se resalta que o el uso de técnicas de BPLC representa una rotura con esa prerrogativa constitucional y procesual. Esto porque se aplica la herramienta a todos los usuarios de una determinada plataforma, mal-intencionados o no, lo que resulta en el filtraje de todo el contenido compartido por patrón.

Por fin, a lo que se refiere a la **proporcionalidad** y a la **necesidad**, se observa , por todo lo expuesto, que técnicas de BPLC representan un riesgo desproporcional en comparación con los beneficios obtenidos. Adicionalmente, ese riesgo se revela innecesario en relación al objetivo, teniendo en vista todas las barreras tecnológicas apuntadas a lo largo de este trabajo, que tornan el BPLC una herramienta poco eficaz para el combate a los ilícitos que se pretende reprender a través de esa técnica.

Finalmente, cabe evidenciar preocupaciones adicionales que transcurren del uso del BPLC. Una de ellas se refiere a la dimensión **sociológica**: las técnicas aquí analizadas representan mecanismos de combate a la diseminación de contenido ilícito, pero no eliminan la fuente creadora de materiales de esa naturaleza. La otra se refiere a los **perjuicios econ**ó**micos** que se pueden causar a causa de la obligación legal de filtraje en masiva de contenidos por parte de las plataformas, lo que podría resultar en la imposibilidad de plataformas de pequeño porte actuar en ese mercado y, así, resultar en una concentración de mercado todavía más intensa por grandes proveedores de aplicación.

La conclusión de este análisis apunta que el uso de técnicas de BPLC para combate a la diseminación de contenidos ilícitos en ambientes codificados – como CSAM, material relativo a terrorismo, entre otros – se muestra una medida inadecuada por diversos motivos. No solo las soluciones de BPLC descriptas hasta el momento presentan diversos fallos y brechas desde un punto de vista tecnológico, como también representan

un debilitamiento de diversas garantías jurídicas consagradas como derechos fundamentales – tales como el derecho a la privacidad, a la libertad de expresión, a la presunción de inocencia, entre otros. En ese sentido, el BPLC se configura como una herramienta potencialmente tan dañosa como la misma rotura de la criptografía segura a través de la inserción de mecanismos de acceso excepcional (*backdoors*) en algoritmos criptográficos.

Aunque la metodología adoptada ofrezca sistematización y consistencia a los argumentos analizados, se notó que la limitación a obras de carácter académico no consigue abarcar los debates más intensos, en el calor del momento, que ocurren inevitablemente fuera del universo de artículos científicos y estudios investigativos. Ese punto ciego inherente acaba exigiendo fuentes adicionales sobre la cronología de los acontecimientos, reacciones por la prensa y eventuales pronunciamientos oficiales.

A lo largo del presente estudio, se realizó una revisión bibliográfica extensiva sobre los debates que permean la propuesta de barrido por el lado del cliente, como también sus repercusiones sociales, jurídicas y políticas. Se espera que el análisis realizado pueda ser utilizado como base para la profundización de las discusiones en trabajos futuros. Se puede indagar sobre la implementación exclusiva en favor del interés del cliente, considerando la distinción a ambientes adversariales, cuando la herramienta debería operar contra el interés del propietario del dispositivo. Se pueden considerar otras propuestas concretas de BPLC, incluyendo algún eventual nuevo anuncio de Apple, siempre y cuando haya compromiso con medidas que cuiden las cuestiones apuntadas en enfoque amplio, por ejemplo, por medio de sistemas con código fuente integralmente abierto.

NOTAS

1 El primer informe, sobre el panorama de la rastreabilidad de mensajes instantáneos, se publicó el 18 de mayo de este año (RODRIGUES, Gustavo Ramos; SANTARÉM, Paulo Renã da Silva; VIEIRA, Victor Barbieri Rodrigues. **Comunicações privadas, investigações e direitos: rastreabilidade de mensagens instantâneas.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, mayo de 2022. Disponible en: https://bit.ly/3yLlb0P. Acceso el: 30 ago 2022).

2 Una discusión en detalles sobre los diversos aspectos pertinentes y las percepciones cuanto a las propuestas de inserción de mecanismos de acceso excepcional en ambientes con criptografía fue objeto de un estudio previo conducido por IRIS (PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en: https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022).

3 SANTARÉM, Paulo Renã da Silva. **"Portas clandestinas": uma tradução mais precisa para debatermos backdoors em criptografia.** Blog: Instituto de Referência em Internet e Sociedade. 17 ene. 2022. Disponible en: https://irisbh.com.br/portas-clandestinas-uma-traducao-mais-precisa- para-debatermos-backdoors-em-criptografia/. Acceso el: 10 oct. 2022.

4 SCHULZE, M. Clipper meets Apple vs. FBI – a comparison of the cryptography discourses from 1993 and 2016. **Media and Communication,** v. 5, n. 1, p. 54-62, 22 mar. 2017.

5 REUTERS. 'Five Eyes' security alliance calls for access to encrypted material. **Reuters,** 30 jul. 2019. Disponível em: https://www.reuters.com/article/us-security-fiveeyes-britain-idUSKCN1UP199. Acceso el: 28 abr. 2022.

6 El término infantil aquí se refiere a personas con menos de 18 años de edad. Desde 1988 Brasil abandonó el "sistema menorista" de la doctrina de la situación irregular – expresada en el Código Mello Matos (Decreto nº 17.943-A/1927) y en el Código de Menores (Ley nº 6.697/1979) – y adoptó el paradigma de la protección integral de niños y adolescentes, por el cual se reconocen integrantes de ese segmento social como sujetos de derecho en situación peculiar de desarrollo, destinatarios de protección y asistencia por el Estado, por la familia y por la sociedad, con prioridad absoluta a sus garantías y derechos básicos (artigo 227 de la Constitución de la República). ECA (Ley nº 8.069/1989) nunca usa el término "menor", preconizando la expresión "niños y adolescentes", a fin de no reforzar la lógica perniciosa y excluyente de la perspectiva anterior (BRANCHER, 2000: 126).

7 Después de las primeras reacciones negativas al anuncio, el número fue informado por el Vice-Presidente de ingeniería de software Craig Federighi en entrevista al Wall Street Journal (STERN, Joanna; HIGGINS, Tim. Apple Executive Defends Tools to Fight Child Porn, Acknowledges Privacy Backlash. The Wall Street Journal. 13. aug. 2021. Disponible en https://www.wsj.com/articles/apple-executive-defends-tools-to-fight-child-porn-acknowledges-privacy-backlash-11628859600. Acceso el 13 oct. 2022) y solo entonces constó de un documento oficial (APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acceso 13 oct. 2022. p. 10).

8 "The second protection is human review: there is no automated reporting in Apple's system. All positive matches must be visually confirmed by Apple as containing CSAM before Apple will disable the account and file a report with the child safety organization" ("La segunda protección es la revisión humana: no hay denuncia automática en el sistema de Apple. Todas las correspondencias positivas se deben confirmar visualmente por Appla como conteniendo CSAM antes de Apple desactivar la cuenta y enviar una denuncia a la organización de protección infantil", en traducción literal) (APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acceso el 13 oct. 2022. p. 8.).

9 Ver DONEDA, Danilo; MACHADO, Diego (orgs.). **A criptografia no direito brasileiro.** São Paulo: Thomson Reuters Brasil, 2020.

10 GALVÃO, Maria C.; RICARTE, Ivan L. M. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion: Filosofia da Informação,** [S.l.] , v. 6, n. 1, p. 57 - 73, set. 2019. P. 58. 7 - 73. Disponible en: http://revista.ibict.br/fiinf/article/view/4835. Acceso el: 10 jun. 2021.

11 SAMPAIO, R. F.; MANCINI, M. C. Estudos de Revisão Sistemática: um guia para síntese criteriosa da evidência científica. **Revista Brasileira de Fisioterapia,** São Carlos, v. 11, n. 1., p. 83-89, 2007. p. 84.

12 Google Académico, o Google Scholar, está accesible en la dirección: https://scholar.google.com.br/.

13 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022.

14 Por medio de API Content Safety, Google identifica nuevas imágenes de CSAM por medio de clasificadores con inteligencia artificial (Anexo 8 de EU - EUROPEAN UNION. European Commission. Commission Staff Working Document. Impact Assessment Report. Accompanying the document. Proposal For a Regulation Of The European Parliament and Of The Council. **Laying down rules to prevent and combat child sexual abuse. {SWD(2022) 209 final}**. Bruselas, 11 may. 2022. Disponible en https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acceso el 13 oct. 2022).

15 Herramienta más difundida de su tipo, PhotoDNA tiene dos etapas: detección y creación de hash. Primero, identifica imágenes superiores a determinado tamaño e, ignorando el texto, analiza si se la conoce. Segundo convierte la imagen original para una versión en escala de gris de baja resolución, aplica un filtro de alta-frecuencia y divide en cuadrantes de cual se extraen medidas estadísticas que generan el hash, una "firma" que permite reconocer imágenes similares sometidas al mismo proceso, pero no permite que se obtenga la imagen original en regreso a partir del hash. (Anexo 8 de EU - EUROPEAN UNION. European Commission. Commission Staff Working Document. Impact Assessment Report. Accompanying the document. Proposal For a Regulation Of The European Parliament and Of The Council. **Laying down rules to prevent and combat child sexual abuse. {SWD(2022) 209 final}**. Bruselas, 11 may. 2022. Disponible en https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acceso el 13 oct. 2022).

16 Safer es una solución modular que identifica, remueve y denuncia imágenes de CSAM. Opera contra CSAM conocido, por modelo de correspondencia, o CSAM potencialmente nuevo y no denunciado, por modelo predictivo de clasificador – una tecnología de aprendizaje de máquina entrenada por la empresa Thorn con centenas de miles de imágenes (Anexo 8 de EU - EUROPEAN UNION. European Commission. Commission Staff Working Document. Impact Assessment Report. Accompanying the document. Proposal For a Regulation Of The European Parliament and Of The Council. **Laying down rules to prevent and combat child sexual abuse. {SWD(2022) 209 final}**. Bruselas, 11 may. 2022. Disponible en:https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acceso el 13 oct. 2022).

17 PDQ de Meta opera por un algoritmo de hash perceptivo de funcionamiento semejante al PhotoDNA, también para detectar CSAM (Anexo 8 de EU - EUROPEAN UNION. European Commission. Commission Staff Working Document. Impact Assessment Report. Accompanying the document. Proposal For a Regulation Of The European Parliament and Of The Council. **Laying down rules to prevent and combat child sexual abuse. {SWD(2022) 209 final}**. Bruselas, 11 may. 2022. Disponible en: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acceso el 13 oct. 2022).

18 MAYER, Jonathan. **Content moderation for end-to-end encrypted messaging.** Princeton University, 2019. P. 42. Disponible en http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging. Acceso el 13 out. 2022.

19 Puede operar por a) enclaves seguros en el servidor de la plataforma; b) correspondencia única de terceros; o c. Correspondência de varios terceros. Las explicaciones de cada técnica rehuyen al objetivo de este estudio, y se pueden encontrar en EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document,** September 2020. Disponible en https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acceso el 13 out. 2022.

20 UN. Human Rights Council. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; Secretariat. **Encryption and anonymity follow-up report: note / by the Secretariat (A/HRC/38/35/Add.5)**. Genebra: UN, 13 July 2018. 18 p. Disponible en https://digitallibrary.un.org/record/1638475. Acceso el 13 oct. 2022.

21 El borrador "Technical solutions to detect child sexual abuse in end-to-end encrypted communications" ("Soluciones Técnicas para Detectar Abuso Sexual de Niños en Comunicaciones Criptografadas de Punta a Punta", en traducción literal) (EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document,** September 2020. Disponible en https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acceso el 13 oct. 2022) fue posteriormente incorporado como parte del "Annex 9: Encryption and the fight against child sexual abuse" (Anexo 9: Criptografía y lucha contra el abuso sexual infantil) en una propuesta pública de la Comisión Europea para la regulación , en el ámbito de la UE, de la prevención y combate al abuso sexual infantil (EU - EUROPEAN UNION. European Commission. Commission Staff Working Document. Impact Assessment Report. Accompanying the document. Proposal For a Regulation Of The European Parliament and Of The Council. **Laying down rules to prevent and combat child sexual abuse. {SWD(2022) 209 final}**. Bruxelas, 11 mai. 2022. Disponible en https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acceso el 13 out. 2022).

22 EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document,** September 2020. Disponible en https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acceso el 13 oct. 2022.

23 GLOBAL ENCRYPTION COALITION. **Breaking Encryption Myths: What the European Commission's leaked report got wrong about online security.** 19 nov. 2020. Disponible en https://www.globalencryption.org/2020/11/breaking-encryption-myths/.

Acceso el 13 oct. 2022; y EDRI - EUROPEAN DIGITAL RIGHTS. **Is surveilling children really protecting them? Our concerns on the interim CSAM regulation.** 24 set. 2020. Disponible en https://edri.org/our-work/is-surveilling-children-really-protecting-them-our-concerns-on-the-interim-csam-regulation/. Acceso el 13 oct. 2022.

24 Constó en la minuta una excepción: "Este artículo visa ofrecer una primera evaluación técnica para auxiliar a identificar posibles soluciones. Trabajo adicional sustantivo, para además del escopo de este artículo, probablemente sería necesario para evaluación, desarrollo y aplicación futuras de soluciones técnicas a través de la infraestructura de las empresas" (EU, 2020: 2). La línea se mantuvo en la versión pública da propuesta: "Este documento visa mapear posibles soluciones que puedan garantizar la privacidad de las comunicaciones electrónicas (incluyendo la privacidad de los niños) y la protección de los niños contra el abuso y la exploración sexual. Las soluciones exploradas son de naturaleza puramente técnica, y ***este artículo no toma ninguna posición sobre el aspecto político relacionado"*** (Anexo 9 de EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document,** September 2020. Disponible en https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acceso el 13 oct. 2022. – grifo nuestro).

25 POHL, Hartmut. Against surveillance of digital communications in Europe. **Gesellschaft Für Informatik,** 8 nov. 2021. Disponible en https://gi.de/meldung/against-surveillance-of-digital-communications-in-europe. Acceso el 13 oct. 2022.

26 UN. Human Rights Council. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; Secretariat. **Encryption and anonymity follow-up report: note / by the Secretariat (A/HRC/38/35/Add.5)**. Ginebra: UN, 13 July 2018. 18 p. Disponible en https://digitallibrary.un.org/record/1638475. Acceso el 13 oct. 2022. P. 8.

27 La presentación simultánea de las tres herramientas, dificultando la distinción entre los nuevos recursos, fue criticada como una posible causa de problemas de comprensión por los usuarios, cuanto a su seguridad y privacidad. El jefe de privacidad de Apple, Erik Neuenschwander, ha respondido tratarse de un conjunto de recursos que actuarían simultáneamente. Ver PANZARINO, Mateus. Interview: Apple's head of Privacy details child abuse detection and Messages safety features. **Tech Crunch+,** [S. l.], p. -, 10 ago. 2021. Disponible en: https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acceso el: 20 sept. 2022.

28 El documento en PDF "**Expanded Protections for Children — Technology Summary**", actualmente disponible en el site de la empresa (APPLE. **Expanded Protections for Children — Technology Summary**. Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf.

Acceso el 13 oct. 2022), es una versión levemente ampliada del texto originalmente publicado en .html en la dirección https://www.apple.com/child-safety/, según registro en el Internet Archive (https://web.archive.org/web/20210805191220/https://www.apple.com/child-safety/). La empresa todavia ha detallado la herramientas "Seguridad de las comunicaciones en mensajes" y "Detección de CSAM" en el documento "Revisão do Modelo de Ameaça de Segurança dos Recursos para Proteção de Crianças" (APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acceso el 13 oct. 2022.).

29 APPLE. **Expanded Protections for Children — Technology Summary**. Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf. Acceso el 13 oct. 2022. p. 4. APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acceso el 13 oct. 2022. p. 2-4.

30 Luiza Brandão ha listado, por ejemplo, críticas de International Association of Privacy Professionals – IAPP, de Global Encryption Coalition, de Internet Society – ISOC y de la ong Electronic Frontier Foundation – EFF (BRANDÃO, Luiza. **Apple e o Mito Privacidade x Segurança.** Blog: Instituto de Referência em Internet e Sociedade – IRIS. 9 ago. 2021. Disponible en: https://irisbh.com.br/apple-e-o-mito-privacidade-x-seguranca/. Acceso el : 19 sept. 2022). También merece destaque la crítica de la empresa WhatsApp (PROVENZANO, Brianna. WhatsApp acusa Apple de crear sistema de vigilancia al verificar abuso infantil en fotos. Uol - Giz_br, [S. l.], p. -, 8 ago. 2021. Disponible en: https://gizmodo.uol.com.br/whatsapp-apple-sistema-abuso-infantil-vigilancia/. Acceso el: 20 sept. 2022).

31 El día siguiente, una carta abierta copiló críticas y pidió la suspensión inmediata de los cambios propuestos y la renovación del compromiso de Apple con la criptografía de punta a punta y la privacidad: hasta meados de septiembre de 2022, la carta fue firmada por más de 30 organizaciones y 8.700 individuos **(AN OPEN LETTER AGAINST APPLE'S PRIVACY-INVASIVE CONTENT SCANNING TECHNOLOGY.** 6 ago. 2021. Disponible en: https://appleprivacyletter.com/. Acceso el: 20 sept. 2022).

La ONG estadounidense Electronic Frontier Foundation (EFF) organizó una campaña denominada "Dígale a Apple: No Escanee Nuestros Teléfonos" ("Tell Apple: Don't Scan Our Phones"). Ver ELECTRONIC FRONTIER FOUNDATION – EFF. **Tell Apple: Don't Scan Our Phones. Action Center.** 1 sept. 2021. Disponible en: https://act.eff.org/action/tell-apple-don-t-scan-our-phones. Acceso el 19/09/2022.

También en los EEUU, la ong CDT clasificó el cambio como una menaza a la seguridad y a la privacidad. Ver CENTER FOR DEMOCRACY AND TECHNOLOGY – CDT. **CDT: Apple's Changes to Messaging and Photo Services Threaten Users' Security and Privacy.** 5 ago. 2021. Disponible en: https://cdt.org/press/cdt-apples-changes-to-messaging-and-photo-services-threaten-users-security-and-privacy/. Acceso el 19/09/2022.

Y el 19 de agosto, noventa organizaciones de la sociedad civil de varios países de todos los continentes (incluyendo EFF, CDT y este Instituto de Referência em Internet e Sociedade) firmaron una carta abierta instando Apple a "*abandonar los planes (...) de construir recursos de vigilancia en iPhones, iPads y otros productos*". Ver FRANKLIN, Sharon Bradford. Greg Nojeim. **International Coalition Calls on Apple to Abandon Plan to Build Surveillance Capabilities into iPhones, iPads, and other Products.** Center for Democracy and Technology – CDT. 19 ago. 2021. https://cdt.org/insights/international-coalition-calls-on-apple-to-abandon-plan-to-build-surveillance-capabilities-into-iphones-ipads-and-other-products/. Acceso el 13 oct. 2022.

32 COHN, Cindy. **Delays Aren't Good Enough — Apple Must Abandon Its Surveillance Plans.** 3 set. 2021. Electronic Frontier Foundation. Disponível em https://www.eff.org/pt-br/deeplinks/2021/09/delays-arent-good-enough-apple-must-abandon-its-surveillance-plans. Acesso em 19/09/2022.

33 La página https://www.apple.com/child-safety tiene el subtítulo "Expanded Protections for Children".

34 JAIN, Shubham; CRETU, Ana-Maria; DE MONTJOYE, Yves-Alexandre. Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning. In: **NeurIPS 2021 Workshop Privacy in Machine Learning.** 2021. Disponible en: https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acceso el 13 oct. 2022. CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponible el https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022. SHENKMAN, C., THAKUR, D., & LLANSÓ, E. (2021). Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis. Center for Democracy & Technology. Disponible en:https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acceso el 13 oct. 2022.; MAYER, Jonathan. **Content moderation for end-to-end encrypted messaging.** Princeton University, 2019. P. 42. Disponible en: http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging. Acceso el 13 oct. 2022.

35 KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In:

Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021. p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022.

36 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en: https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 54.

37 ROSENZWEIG, Paul. **The Law and Policy of Client-Side Scanning** (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https://digitalcommons.wcl.american.edu/research/58. Acceso el 13 oct. 2022. p. 2.

38 HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 out. 2022. p. 1.

39 APPLE. Expanded Protections for Children: frequently Asked Questions. v1.1. Agosto de 2021. Disponible en: https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Frequently_Asked_Questions.pdf. Acceso el 13 oct. 2022. P. 3.

40 CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponible en https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. p. 10-13.

41 ROSENZWEIG, Paul. **The Law and Policy of Client-Side Scanning** (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https:/ a/digitalcommons.wcl.american.edu/research/58. Acceso el 13 oct. 2022. p. 2.

42 KNOCKEL, Jeffrey; PARSONS, Christopher; RUAN, Lotus; XIONG, Ruohan; CRANDALL, Jedidiah; DEIBERT, Ron. **We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus.** Citizen Lab Research Report Nº 127. University of Toronto, May 2020. Disponible en https://citizenlab.ca/2020/05/we-chat-they-watch/. Acceso el 03 jul. 2022. p. 10.

43 KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). **Encryption, Privacy and Children's Right to Protection from Harm**, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. p. 3.

44 HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 oct. 2022. p. 3.

45 KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). **Encryption, Privacy and Children's Right to Protection from Harm**, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. p. 3.

46 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en: https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p.35.

47 DUARTE, Natasha; LLANSÓ, Emma; LOUP, Anna. **Mixed Messages? The Limits of Automated Social Media Content Analysis.** Center for Democracy & Technology. 28 nov. 2017. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/. Acceso el 05 jul. 2022. p. 9.

48 KNOCKEL, Jeffrey; PARSONS, Christopher; RUAN, Lotus; XIONG, Ruohan; CRANDALL, Jedidiah; DEIBERT, Ron. **We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus.** Citizen Lab Research Report Nº 127. University of Toronto, May 2020. Disponible en https://citizenlab.ca/2020/05/we-chat-they-watch/. Acceso el 03 jul. 2022. p. 10.

49 JAIN, Shubham; CRETU, Ana-Maria; DE MONTJOYE, Yves-Alexandre. Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning. In: **NeurIPS 2021 Workshop Privacy in Machine Learning.** 2021. Disponible en https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acceso el 13 oct. 2022. p. 2318.

50 DUARTE, Natasha; LLANSÓ, Emma; LOUP, Anna. **Mixed Messages? The Limits of Automated Social Media Content Analysis.** Center for Democracy & Technology. 28 nov. 2017. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/. Acceso el 05 jul. 2022. p. 9.

51 PANZARINO, Mateus. Interview: Apple's head of Privacy details child abuse detection and Messages safety features. **Tech Crunch+,** [S. l.], p. -, 10 ago. 2021. Disponible en: https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acceso el: 20 sept. 2022.

52 ROSENZWEIG, Paul. **The Law and Policy of Client-Side Scanning** (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https://digitalcommons.wcl.american.edu/research/58. Acceso el 13 oct. 2022. p. 2.

53 KNOCKEL, Jeffrey; PARSONS, Christopher; RUAN, Lotus; XIONG, Ruohan; CRANDALL, Jedidiah; DEIBERT, Ron. **We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus.** Citizen Lab Research Report Nº 127. University of Toronto, May 2020. Disponible en https://citizenlab.ca/2020/05/we-chat-they-watch/. Acceso el 03 jul. 2022. p. 10.

54 APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf.

55 CENTER FOR DEMOCRACY AND TECHNOLOGY – CDT. **CDT: Breaking encryption myths What the European Commission's leaked report got wrong about online security**. 19 nov. 2020. Disponible en https://cdt.org/insights/breaking-encryption-myths-what-the-european-commissions-leaked-report-got-wrong-about-online-security/. Acceso el 26/09/2022.

56 APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en: https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf.

57 APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. p. 6.

58 PANZARINO, Mateus. Interview: Apple's head of Privacy details child abuse detection and Messages safety features. **Tech Crunch+,** [S. l.], p. -, 10 ago. 2021. Disponível em: https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acceso en: 20 sept. 2022.

59 REIS, Julio C. S., MELO, Philipe, GARIMELLA, Kiran, & BENEVENUTO, Fabrício. Can WhatsApp benefit from debunked fact checked stories to reduce misinformation? **Harvard Kennedy School Misinformation Review**. 20. ago 2020. https://doi.org/10.37016/mr-2020-035. Disponible en https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acceso el 13 oct. 2022. p. 2.

60 EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document,** September 2020. Disponible en https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acceso el 13 oct. 2022. p. 7.

61 LUMNIOTIS, 2021: 27)

62 Sobre posibles impactos económicos negativos de la previsión legal de mecanismos de debilitamiento de la criptografía, ver BARKER, George. LEHR, William. LONEY, Mark. SICKER, Douglas. **O Impacto Econômico das Leis que Enfraquecem a Criptografia.** Law & Economics Consulting Associates (LECA). Traducción de Paulo Rená da Silva Santarém. Reston, VA: Internet Society, 2021. Disponible en https://www.isoc.org.br/files/The-Economic-Impact-of-Laws-the-Weaken-Encryption-PT.pdf. Acceso el 13 oct. 2022.

63 KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022. p. 10.

64 ROSENZWEIG, Paul. **The Law and Policy of Client-Side Scanning** (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https://digitalcommons.wcl.american.edu/research/58. Aceso el 13 oct. 2022. p. 15.

65 CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Rená da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponible en https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022. p. 27.

66 HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 out. 2022. p. 1.

67 En sistemas de procesamiento de bancos de datos, un protocolo de 2PC o confirmación de dos-fases (two-phase commit) opera un conjunto de cambios en un sistema distribuido: los resultados pretendidos, en bloque, se efectivan o abortan en la segunda etapa de acuerdo con las respuestas, respectivamente, positivas o negativas de los participantes en la primera etapa de verificación preparatoria.

68 DUARTE, Natasha; LLANSÓ, Emma; LOUP, Anna. **Mixed Messages? The Limits of Automated Social Media Content Analysis.** Center for Democracy & Technology. 28 nov. 2017. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-

media-content-analysis/. Acceso el 05 jul. 2022. p. 9. REIS, Julio C. S., MELO, Philipe, GARIMELLA, Kiran, & BENEVENUTO, Fabrício. Can WhatsApp benefit from debunked fact checked stories to reduce misinformation? **Harvard Kennedy School Misinformation Review**. 20. ago 2020. https://doi.org/10.37016/mr-2020-035. Disponible en https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acceso el 13 oct. 2022. p. 5. NEGREIRO ACHIAGA, Maria Del Mar. Curbing the surge in online child abuse: The dual role of digital technology in fighting and facilitating its proliferation. European Parliament, Think Tank, 23 nov. 2020. Disponible en: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360. Acceso el 13 oct. 2022. p. 9.

69 CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 feb. 2022. Disponible en: https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022. p. 25.

70 BURSZTEIN, Elie. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In **Enigma**. Burlingame, CA: USENIX Association, January 2019. Disponible en https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf. Acceso el 13 oct. 2022.

71 DUARTE, Natasha; LLANSÓ, Emma; LOUP, Anna. **Mixed Messages? The Limits of Automated Social Media Content Analysis.** Center for Democracy & Technology. 28 nov. 2017. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/. Acceso el 05 jul. 2022. p. 9.

72 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 26. GROVER, Gurshabad; RAJWADE, Tanaya; KATIRA, Divyank. The Ministry and the Trace: Subverting End-to-End Encryption. **NUJS L. Rev.,** v. 14, 2021. http://nujslawreview.org/2021/07/09/the-ministry-and-the-trace-subverting-end-to-end-encryption/. Acceso el 13 oct. 2022. p. 11. STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, **Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In 2022 ACM Conference on Fairness, Accountability, and Transparency.** 2022. p. 58-69. Disponible en https://doi.org/10.48550/arXiv.2111.06628. Acceso el 13 oct. 2022. p. 12.

73 JAIN, Shubham; CRETU, Ana-Maria; DE MONTJOYE, Yves-Alexandre. Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning. In: **NeurIPS 2021 Workshop Privacy in Machine Learning.** 2021. Disponible en https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acceso el 13 oct. 2022.

74 LIMNIOTIS, Konstantinos. Cryptography as the Means to Protect Fundamental Human Rights. **Cryptography,** v. 5, n. 4, p. 34, 2021. https://doi.org/10.3390/cryptography5040034. Acceso el 13 oct. 2022. p. 26.

75 BURSZTEIN, Elie. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In **Enigma**. Burlingame, CA: USENIX Association, January 2019. Disponible en https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf. Acceso el 13 oct. 2022.

76 KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022. p. 9-10

77 GROVER, Gurshabad; RAJWADE, Tanaya; KATIRA, Divyank. The Ministry and the Trace: Subverting End-to-End Encryption. **NUJS L. Rev.,** v. 14, p. 11, 2021. http://nujslawreview.org/2021/07/09/the-ministry-and-the-trace-subverting-end-to-end-encryption/. p. 11. CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 feb. 2022. Disponible en: https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022. p. 27; STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, **Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In 2022 ACM Conference on Fairness, Accountability, and Transparency.** 2022. p. 58-69. Disponible en https://doi.org/10.48550/arXiv.2111.06628. Acceso el 13 oct. 2022. p. 12. PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 54.

78 KOLLNIG, K., SHUBA, A., BINNS, R., VAN KLEEK, M., & SHADBOLT, N. Are IPhones Really Better for Privacy? A Comparative Study of IOS and Android Apps. **Proceedings on Privacy Enhancing Technologies,** v. 2022, n. 2, 2022. https://ora.ox.ac.uk/objects/uuid:f29c7413-222e-45bf-ac0c-de927df105ab. Acceso el 13 oct. 2022. p. 21.

79 REIS, Julio C. S., MELO, Philipe, GARIMELLA, Kiran, & BENEVENUTO, Fabrício. Can WhatsApp benefit from debunked fact checked stories to reduce misinformation? **Harvard Kennedy School Misinformation Review**. 20. ago 2020. https://doi.org/10.37016/mr-2020-035. Disponible en: https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acceso el 13 oct. 2022. p. 4. KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media

in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022. p. 10.

80 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 9.

81 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 37.

82 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible em:https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 37.

83 "Aplicaciones de mensajes criptografados pueden usarlo para ayudar a alertar a los usuarios sobre contenido malicioso, con privacidad significativamente mejor de lo que los enfoques que envían datos en texto simple para servidores de terceros. En otra configuración, plataformas de medios sociales que actualmente consultan datos de texto simple de sus usuarios a servicios de terceros para ayudar a identificar abusos pueden usar nuestras técnicas para mejorar la privacidad de sus usuarios" (HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 oct. 2022. p. 1. p. 2.)

84 HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 oct. 2022. p. 1.

85 STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, **Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In 2022 ACM Conference on Fairness, Accountability, and Transparency.** 2022. p. 58-69. Disponible en https://doi.org/10.48550/arXiv.2111.06628. Acceso el 13 oct. 2022. p. 12.

86 GROVER, Gurshabad; RAJWADE, Tanaya; KATIRA, Divyank. The Ministry and the Trace: Subverting End-to-End Encryption. **NUJS L. Rev.,** v. 14, 2021. Disponible en http://nujslawreview.org/2021/07/09/the-ministry-and-the-trace-subverting-end-to-end-encryption/. Acceso el 13 oct. 2022. p. 11.

87 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 34.

88 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 37.

89 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 27.

90 KOLLNIG, K., SHUBA, A., BINNS, R., VAN KLEEK, M., & SHADBOLT, N. Are IPhones Really Better for Privacy? A Comparative Study of IOS and Android Apps. **Proceedings on Privacy Enhancing Technologies,** v. 2022, n. 2, 2022. https://ora.ox.ac.uk/objects/uuid:f29c7413-222e-45bf-ac0c-de927df105ab. Acceso el 13 oct. 2022.

91 CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponible en https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022. p. 27

92 ROSENZWEIG, Paul. **The Law and Policy of Client-Side Scanning** (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https://digitalcommons.wcl.american.edu/research/58. Acceso el 13 oct. 2022. p. 15.

93 NEGREIRO ACHIAGA, Maria Del Mar. Curbing the surge in online child abuse: The dual role of digital technology in fighting and facilitating its proliferation. European Parliament, Think Tank, 23 nov. 2020. Disponível em https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360. Acceso en 13 oct. 2022. p. 7.

94 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 7.

95 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 3.

96 STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, **Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In 2022 ACM**

Conference on Fairness, Accountability, and Transparency. 2022. p. 58-69. Disponible en https://doi.org/10.48550/arXiv.2111.06628. Acceso el 13 oct. 2022. p. 12.

97 REIS, Julio C. S., MELO, Philipe, GARIMELLA, Kiran, & BENEVENUTO, Fabrício. Can WhatsApp benefit from debunked fact checked stories to reduce misinformation? **Harvard Kennedy School Misinformation Review**. 20. ago 2020. https://doi.org/10.37016/mr-2020-035. Disponible en https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acceso el 13 oct. 2022.

98 KNOCKEL, Jeffrey; PARSONS, Christopher; RUAN, Lotus; XIONG, Ruohan; CRANDALL, Jedidiah; DEIBERT, Ron. **We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus.** Citizen Lab Research Report Nº 127. University of Toronto, May 2020. Disponible en https://citizenlab.ca/2020/05/we-chat-they-watch/. Acceso el 03 jul. 2022. p. 47-48.

99 LIMNIOTIS, Konstantinos. Cryptography as the Means to Protect Fundamental Human Rights. **Cryptography,** v. 5, n. 4, p. 34, 2021. https://doi.org/10.3390/cryptography5040034. Acceso el 13 oct. 2022. p. 26. PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponíible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 54. KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022.

100 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 54-55.

101 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 2.

102 KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). **Encryption, Privacy and Children's Right to Protection from Harm**, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. p. 11.

103 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en: https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 55.

104 STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, **Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In 2022 ACM Conference on Fairness, Accountability, and Transparency.** 2022. p. 58-69. Disponible en https://doi.org/10.48550/arXiv.2111.06628. Acceso el 13 oct. 2022. p. 12.

105 CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Rená da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 feb. 2022. Disponible en https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022. p 25.

106 KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). **Encryption, Privacy and Children's Right to Protection from Harm**, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. p. 11.

107 KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en: https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022. p. 10.

108 HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 oct. 2022. p. 3.

109 KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022. p. 10.

110 (GOVER e outros, 2021: 10)

111 KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING,

Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). **Encryption, Privacy and Children's Right to Protection from Harm**, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. p. 11.

112 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 38-39.

113 KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). **Encryption, Privacy and Children's Right to Protection from Harm**, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. p. 10.

114 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 2.

115 PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en https://bit.ly/3kGTde3. Acceso el: 19 abr. 2022. p. 9.

116 KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022.

117 ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022. p. 35

118 El primer informe sobre el panorama de la rastreabilidad de mensajes instantáneos, fue publicado el 18 de mayo de este año (RODRIGUES, Gustavo Ramos; SANTARÉM, Paulo Renã da Silva; VIEIRA, Victor Barbieri Rodrigues. **Comunicações privadas, investigações e direitos: rastreabilidade de mensagens instantâneas.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, mayo de 2022. Disponible en: : https://bit.ly/3yLlb0P. Acceso el: 30 ago 2022).

Referencias

ABELSON, Hal e outros. **Bugs in our Pockets: The Risks of Client-Side Scanning.** arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022.

AN OPEN LETTER AGAINST APPLE'S PRIVACY-INVASIVE CONTENT SCANNING TECHNOLOGY. 6 ago. 2021. Disponible en: https://appleprivacyletter.com/. Acceso el: 20 sept. 2022.

BRANDÃO, Luiza. **Apple e o Mito Privacidade x Segurança.** Blog: Instituto de Referência em Internet e Sociedade. 9 ago. 2021. Disponible en:: https://irisbh.com.br/apple-e-o-mito-privacidade-x-seguranca/. Acceso el: 19 sept. 2022.

APPLE. **Expanded Protections for Children — Technology Summary.** Agosto de 2021. Disponible en: https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf. Acceso el 13 oct. 2022.

APPLE. **Expanded Protections for Children: frequently Asked Questions. v1.1.** Agosto de 2021. Disponible en: https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Frequently_Asked_Questions.pdf. Acceso el 13 oct. 2022.

APPLE. **Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features.** Agosto de 2021. Disponible en:https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acceso el 13 oct. 2022.

BARKER, George. LEHR, William. LONEY, Mark. SICKER, Douglas. **O Impacto Econômico das Leis que Enfraquecem a Criptografia.** Law & Economics Consulting Associates (LECA). Traducción de Paulo Renà da Silva Santarém. Reston, VA: Internet Society, 2021. Disponible en:https://www.isoc.org.br/files/The-Economic-Impact-of-Laws-the-Weaken-Encryption-PT.pdf. Acceso el 13 out. 2022.

BRANCHER, Narciso. Organização e gestão do sistema de garantias de direitos da infância e da juventude. In **Encontros Pela Justiça na Educação**. Brasília: Fundescola/MEC, 2000, p. 126.

BURSZTEIN, Elie. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In **Enigma**. Burlingame, CA: USENIX Association, January 2019. Disponible en:https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf. Acceso el 13 oct. 2022.

CENTER FOR DEMOCRACY AND TECHNOLOGY – CDT. **CDT: Apple's Changes to**

Messaging and Photo Services Threaten Users' Security and Privacy. 5 ago. 2021. Disponible en:https://cdt.org/press/cdt-apples-changes-to-messaging-and-photo-services-threaten-users-security-and-privacy/. Acceso el 19/09/2022.

CENTER FOR DEMOCRACY AND TECHNOLOGY – CDT. **CDT: Breaking encryption myths What the European Commission's leaked report got wrong about online security.** 19 nov. 2020. Disponible en:https://cdt.org/insights/breaking-encryption-myths-what-the-european-commissions-leaked-report-got-wrong-about-online-security/. Acceso el 26/09/2022.

CENTER FOR DEMOCRACY & TECHNOLOGY – CDT. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta.** Traducción: SANTARÉM, Paulo Rená da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponible en:https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acceso el 13 oct. 2022.

COHN, Cindy. **Delays Aren't Good Enough—Apple Must Abandon Its Surveillance Plans.** Electronic Frontier Foundation. 3 sept. 2021. Disponible en:https://www.eff.org/pt-br/deeplinks/2021/09/delays-arent-good-enough-apple-must-abandon-its-surveillance-plans. Acceso el 19/09/2022.

DONEDA, Danilo; MACHADO, Diego (orgs.). **A criptografia no direito brasileiro.** São Paulo: Thomson Reuters Brasil, 2020.

DUARTE, Natasha; LLANSÓ, Emma; LOUP, Anna. **Mixed Messages? The Limits of Automated Social Media Content Analysis.** Center for Democracy & Technology. 28 nov. 2017. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/. Acceso el 05 jul. 2022.

EDRI - EUROPEAN DIGITAL RIGHTS. **Is surveilling children really protecting them? Our concerns on the interim CSAM regulation.** 24 set. 2020. Disponible en:https://edri.org/our-work/is-surveilling-children-really-protecting-them-our-concerns-on-the-interim-csam-regulation/. Acceso el 13 oct. 2022.

ELECTRONIC FRONTIER FOUNDATION - EFF. **Tell Apple: Don't Scan Our Phones. Action Center.** 1 set. 2021. Disponible en:https://act.eff.org/action/tell-apple-don-t-scan-our-phones. Acceso el 19/09/2022.

EU - EUROPEAN UNION. European Commission. Commission Staff Working Document. Impact Assessment Report. Accompanying the document. Proposal For a Regulation Of The European Parliament and Of The Council. **Laying down rules to prevent and combat child sexual abuse. {SWD(2022) 209 final}**. Bruselas, 11 may. 2022. Disponible en:https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acceso el 13 oct. 2022.

EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document**, September 2020. Disponible en:https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acceso el 13 oct. 2022.

FRANKLIN, Sharon Bradford. Greg Nojeim. **International Coalition Calls on Apple to Abandon Plan to Build Surveillance Capabilities into iPhones, iPads, and other Products.** Center for Democracya and Technology – CDT. 19 ago. 2021. https://cdt.org/insights/international-coalition-calls-on-apple-to-abandon-plan-to-build-surveillance-capabilities-into-iphones-ipads-and-other-products/. Acceso el 13 oct. 2022.

GALVÃO, Maria C.; RICARTE, Ivan L. M. **Revisão sistemática da literatura: conceituação, produção e publicação.** Logeion: Filosofia da Informação, [S.l.] , v. 6, n. 1, p. 57 - 73, sept. 2019. P. 58. 7 - 73. Disponible en: http://revista.ibict.br/fiinf/article/view/4835. Acceso el: 10 jun. 2021.

GLOBAL ENCRYPTION COALITION. **Breaking Encryption Myths: What the European Commission's leaked report got wrong about online security.** 19 nov. 2020. Disponible en:https://www.globalencryption.org/2020/11/breaking-encryption-myths/. Acceso el 13 oct. 2022.

GROVER, Gurshabad; RAJWADE, Tanaya; KATIRA, Divyank. The Ministry and the Trace: Subverting End-to-End Encryption. **NUJS L. Rev.**, v. 14, p. 11, 2021. http://nujslawreview.org/2021/07/09/the-ministry-and-the-trace-subverting-end-to-end-encryption/. Acceso el 13 oct. 2022.

HUA, Yiqing e outros. **Increasing Adversarial Uncertainty to Scale Private Similarity Testing.** arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua. Acceso el 13 oct. 2022.

JAIN, Shubham; CRETU, Ana-Maria; DE MONTJOYE, Yves-Alexandre. Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning. In: **NeurIPS 2021 Workshop Privacy in Machine Learning.** 2021. Disponible en:https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acceso el 13 oct. 2022.

KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection. **Encryption, Privacy and Children's Right to Protection from Harm**. Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en:https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. Acceso el 13 oct. 2022.

KNOCKEL, Jeffrey; PARSONS, Christopher; RUAN, Lotus; XIONG, Ruohan; CRANDALL, Jedidiah; DEIBERT, Ron. **We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus.** Citizen Lab Research Report Nº 127. University of Toronto, May 2020. Disponible en:https://citizenlab.ca/2020/05/we-chat-they-watch/. Acceso el 03 jul. 2022.

KOLLNIG, K., SHUBA, A., BINNS, R., VAN KLEEK, M., & SHADBOLT, N. Are IPhones Really Better for Privacy? A Comparative Study of IOS and Android Apps. **Proceedings on Privacy Enhancing Technologies**, v. 2022, n. 2, 2022. https://ora.ox.ac.uk/objects/uuid:f29c7413-222e-45bf-ac0c-de927df105ab. Acceso el 13 oct. 2022.

KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: **Proceeding of the 30th USENIX Security Symposium, August 11-13, 2021.** p. 893-910. Disponible en:https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acceso el 13 oct. 2022.

LIMNIOTIS, Konstantinos. Cryptography as the Means to Protect Fundamental Human Rights. **Cryptography**, v. 5, n. 4, p. 34, 2021. https://doi.org/10.3390/cryptography5040034. Acceso el 13 oct. 2022.

MAYER, Jonathan. **Content moderation for end-to-end encrypted messaging.** Princeton University, 2019. Disponible en:http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging. Acceso el 13 oct. 2022.

NEGREIRO ACHIAGA, Maria Del Mar. **Curbing the surge in online child abuse: The dual role of digital technology in fighting and facilitating its proliferation.** European Parliament, Think Tank, 23 nov. 2020. Disponible en:https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360. Acceso el 13 oct. 2022.

ONU - ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Escritório do Alto Comissariado para Direitos Humanos. **The right to privacy in the digital age: report of the Office of the United Nations High Commissioner for Human Rights. A/HRC/51/17.** Ginebra: ONU, 4 ago. 2022. Disponible en:https://digitallibrary.un.org/record/3985679?ln=en. Acceso el 13 oct. 2022.

PANZARINO, Mateus. Interview: Apple's head of Privacy details child abuse detection and Messages safety features. **Tech Crunch+**, [S. l.], p. -, 10 ago. 2021. Disponible en:https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acceso el: 20 sept. 2022.

PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. **Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise.** Belo Horizonte: Instituto de Referência em Internet e Sociedade,

2021. Disponible en:: https://bit.ly/3kGTde3. Acceso el 13 oct. 2022.

POHL, Hartmut. Against surveillance of digital communications in Europe. **Gesellschaft** Für **Informatik**, 8 nov. 2021. Disponible en:https://gi.de/meldung/against-surveillance-of-digital-communications-in-europe. Acceso el 13 oct. 2022.

REIS, Julio C. S., MELO, Philipe, GARIMELLA, Kiran, & BENEVENUTO, Fabrício. Can WhatsApp benefit from debunked fact checked stories to reduce misinformation? **Harvard Kennedy School Misinformation Review.** 20 ago. 2020. https://doi.org/10.37016/mr-2020-035. Disponible en: https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acceso el 13 oct. 2022.

REUTERS. 'Five Eyes' security alliance calls for access to encrypted material. **Reuters**, 30 jul. 2019. Disponível em: https://www.reuters.com/article/us-security-fiveeyes-britain-idUSKCN1UP199. Acceso el: 28 abr. 2022.

RODRIGUES, Gustavo Ramos; SANTARÉM, Paulo Renã da Silva; VIEIRA, Victor Barbieri Rodrigues. **Comunicações privadas, investigações e direitos: rastreabilidade de mensagens instantâneas.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, mayo de 2022. Disponible en: https://bit.ly/3yLlb0P. Acceso el: 30 ago 2022.

ROSENZWEIG, Paul. **The Law and Policy of Client-Side Scanning** (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https://digitalcommons.wcl.american.edu/research/58. Acceso el 13 oct. 2022.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de Revisão Sistemática: um guia para síntese criteriosa da evidência científica. **Revista Brasileira de Fisioterapia**, São Carlos, v. 11, n. 1., p. 83-89, 2007.

SANTARÉM, Paulo Renã da Silva. **"Portas clandestinas": uma tradução mais precisa para debatermos backdoors em criptografia.** Blog: Instituto de Referência em Internet e Sociedade. 17 ene. 2022. Disponible en:: https://irisbh.com.br/portas-clandestinas-uma-traducao-mais-precisa-para-debatermos-backdoors-em-criptografia/. Acceso el 13 oct. 2022.

SCHULZE, M. Clipper meets Apple vs. FBI – a comparison of the cryptography discourses from 1993 and 2016. **Media and Communication**, v. 5, n. 1, p. 54-62, 22 mar. 2017.

SHENKMAN, C., THAKUR, D., & LLANSÓ, E. (2021). **Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis.** Center for Democracy & Technology. Disponible en https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acceso el 13 oct. 2022.

STERN, Joanna; HIGGINS, Tim. Apple Executive Defends Tools to Fight Child Porn, Acknowledges Privacy Backlash. **The Wall Street Journal.** 13. aug. 2021. Disponible en:https://www.wsj.com/articles/apple-executive-defends-tools-to-fight-child-porn-acknowledges-privacy-backlash-11628859600. Acceso el 13 oct. 2022.

STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In: **2022 ACM Conference on Fairness, Accountability, and Transparency.** 2022. p. 58-69. Disponible en: https://doi.org/10.48550/arXiv.2111.06628. Acceso el 13 oct. 2022.

UN. Human Rights Council. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; Secretariat. **Encryption and anonymity follow-up report: note / by the Secretariat (A/HRC/38/35/Add.5)**. Ginebra: UN, 13 Jul. 2018. 18 p. Disponible en: https://digitallibrary.un.org/record/1638475. Acceso el 13 oct. 2022.

Apéndice 1 - Corpus total de textos analisados

REFERENCIA	CATEGORIA	FUENTE
ABELSON, Hal e outros. Bugs in our Pockets: The Risks of Client-Side Scanning. arXiv preprint arXiv:2110.07450, 15/10/2021. https://arxiv.org/abs/2110.07450. Acceso el 19/05/2022.	Informe	Google Académico
APPLE. Expanded Protections for Children, August 2021. Disponible en:https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf	Informe	Referencia de ABELSON et al., 2021.
APPLE. Security Threat Model Review of Apple's Child Safety Features: Protections Against Attacks and Misuse of Apple's Child Safety Features, 2021. Disponible en:https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf.	Informe	Referencia de ABELSON et al., 2021.
BURSZTEIN, Elie. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In Enigma. Burlingame, CA: USENIX Association, January 2019. Disponible en: https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf.	Trabajo publicado en anales de evento académico	Referencia de ABELSON et al., 2021.
CENTER FOR DEMOCRACY & TECHNOLOGY. Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta. Traducción: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 de Febrero de 2022. Disponible en: https://bit.ly/3GKVY7e	Informe	Inserción discricionaria

REFERENCIA	CATEGORIA	FUENTE
DUARTE, Natasha; LLANSÓ, Emma; LOUP, Anna. “Mixed Messages? The Limits of Automated Social Media Content Analysis.” Center for Democracy & Technology. 28 nov. 2017. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/. Acceso más reciente el 05 jul. 2022.	Informe	Referencia de CENTER FOR DEMOCRACY & TECHNOLOGY, 2022.
EU - EUROPEAN UNION. European Commission. Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document, September 2020. Disponible en:https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf.	Informe	Referencia de ABELSON et al., 2021.
GROVER, Gurshabad; RAJWADE, Tanaya; KATIRA, Divyank. The Ministry and the Trace: Subverting End-to-End Encryption. NUJS L. Rev., v. 14, p. 11, 2021. http://nujslawreview.org/2021/07/09/the-ministry-and-the-trace-subverting-end-to-end-encryption/	Artículo científico	Google Académico
HUA, Yiqing e outros. Increasing Adversarial Uncertainty to Scale Private Similarity Testing. arXiv preprint arXiv:2109.01727, 2021. https://www.usenix.org/conference/usenixsecurity22/presentation/hua	Artiículo científico	Google Académico
JAIN, Shubham; CRETU, Ana-Maria; DE MONTJOYE, Yves-Alexandre. Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning. In: NeurIPS 2021 Workshop Privacy in Machine Learning. 2021. Disponible en:https://www.usenix.org/conference/usenixsecurity22/presentation/jain.	Trabajo publicado en anales de evento académico	Google Académico

REFERENCIA	CATEGORIA	FUENTE
KARDEFELT-WINTHER, Daniel; DAY, Emma; BERMAN, Gabrielle; WITTING, Sabine K.; BOSE, Anjan, on behalf of UNICEF's cross-divisional task force on child online protection (2020). Encryption, Privacy and Children's Right to Protection from Harm, Innocenti Working Papers, nº 2020-14, UNICEF Office of Research - Innocenti, Florence: UNICEF Office of Research – Innocenti. Disponible en: https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html.	Informe	Google Académico
KNOCKEL, Jeffrey; PARSONS, Christopher; RUAN, Lotus; XIONG, Ruohan; CRANDALL, Jedidiah; DEIBERT, Ron. We Chat, They Watch: How International Users Unwittingly Build up WeChat's Chinese Censorship Apparatus. Citizen Lab Research Report Nº 127. University of Toronto, May 2020. Disponible en:https://citizenlab.ca/2020/05/we-chat-they-watch/. V Acceso el 03 jul. 2022.	Informe	Referencia de CENTER FOR DEMOCRACY & TECHNOLOGY, 2022.
KOLLNIG, K., SHUBA, A., BINNS, R., VAN KLEEK, M., & SHADBOLT, N. "Are IPhones Really Better for Privacy? A Comparative Study of IOS and Android Apps." Proceedings on Privacy Enhancing Technologies, v. 2022, n. 2, 2022. https://ora.ox.ac.uk/objects/uuid:f29c7413-222e-45bf-ac0c-de927df105ab.	Artículo científico	Google Académico
KULSHRESTHA, Anunay; MAYER, Jonathan. Identifying Harmful Media in {End-to-End} Encrypted Communication: Efficient Private Membership Computation. In: 30th USENIX Security Symposium (USENIX Security 21). 2021. p. 893-910.	Trabajo publicado en anales de evento académico	Google Académico

REFERENCIA	CATEGORIA	FUENTE
LIMNIOTIS, Konstantinos. Cryptography as the Means to Protect Fundamental Human Rights. Cryptography, v. 5, n. 4, p. 34, 2021. https://doi.org/10.3390/cryptography5040034.	Artículo científico	Google Académico
MAYER, Jonathan. Content moderation for end-to-end encrypted messaging. Princeton University, 2019. http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging	Informe	Referencia de CENTER FOR DEMOCRACY & TECHNOLOGY, 2022.
NEGREIRO ACHIAGA, Maria Del Mar. Curbing the surge in online child abuse: The dual role of digital technology in fighting and facilitating its proliferation. European Parliament, Think Tank, 23 nov. 2020. Disponible en:https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360.	Informe	Referencia de ABELSON et al., 2021.
PEREIRA, Ana Bárbara Gomes; RODRIGUES, Gustavo Ramos; VIEIRA, Victor Barbieri Rodrigues. Percepções sobre criptografia e investigações criminais no Brasil: mapeamento e análise. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponible en:: https://bit.ly/3kGTde3.	Informe	Google Académico

REFERENCIA	CATEGORIA	FUENTE
REIS, Julio C. S., MELO, Philipe, GARIMELLA, Kiran, & BENEVENUTO, Fabrício. Can WhatsApp benefit from debunked fact checked stories to reduce misinformation? Harvard Kennedy School Misinformation Review. August 20, 2020. https://doi.org/10.37016/mr-2020-035	Artículo científico	Referencia de CENTER FOR DEMOCRACY & TECHNOLOGY, 2022.
ROSENZWEIG, Paul. The Law and Policy of Client-Side Scanning (Originally published by Lawfare). 2020. Joint PIJIP/TLS Research Paper Series. 58. https://digitalcommons.wcl.american.edu/research/58.	Artículo científico	Google Académico
Shenkman, C., Thakur, D., & Llansó, E. (2021). Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis. Center for Democracy & Technology. https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/	Informe	Referencia de CENTER FOR DEMOCRACY & TECHNOLOGY, 2022.
STRUPPEK, Lukas; HINTERSDORF, Dominik; NEIDER, Daniel; KERSTING, Kristian. Learning to break deep perceptual hashing: The use case neuralhash. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022. p. 58-69. Disponible en:https://doi.org/10.48550/arXiv.2111.06628.	Artículo científico	Google Académico

Apéndice 2 – Informe de análisis

- E-mail
- Año
- Referencia ABNT
- Enlace de la publicación
- Categoría
  marcar apenas una opción

  Artículo científico

  Declaración, carta abierta

  Informe

  Nota técnica

  Jurisprudencia

  Trabajo publicado en anales de evento académico

  Monografía, disertación o tesis

  Artículo de opinión

  Materia de periódico

  Post de blog

- Escopo
  marcar apenas una opción

  Rastreabilidad

  Hacking gubernamental

  Barrido por el lado del cliente

- Síntesis
  Texto de resumen elaborado por el equipo de IRIS: debe incluir una presentación breve de la propuesta del trabajo, metodología (o ausencia de indicación de metodología), eventuales referencias relevantes (citadas como base para el concepto o posicionamiento indicado en el trabajo) y enfoque dado al escopo analizado.
- Comentarios
- Citaciones
- Observaciones

iris

INSTITUTO
DE REFERENCIA
EN INTERNET
Y SOCIEDAD