Varredura pelo lado do **cliente**

uma revisão sistemática

Comunicações privadas, investigações e **direitos**

iris
INSTITUTO DE REFERÊNCIA EM INTERNET E SOCIEDADE

Varredura pelo lado do **cliente**

uma revisão sistemática

AUTORIA

Gustavo Ramos Rodrigues

Paulo Rená da Silva Santarém

Victor Barbieri Rodrigues Vieira

Wilson Guilherme Dias Pereira

REVISÃO

Lahis Pasquali Kurtz

Luiza Correa de Magalhães Dutra

REVISÃO EXTERNA

Camila Laranjeira da Silva

Roberta Battisti

PROJETO GRÁFICO, CAPA, DIAGRAMAÇÃO E FINALIZAÇÃO

Felipe Duarte

PRODUÇÃO EDITORIAL

Instituto de Referência em Internet e Sociedade

COMO CITAR EM ABNT

PEREIRA, Wilson Guilherme Dias; RODRIGUES, Gustavo Ramos; SANTARÉM, Paulo Rená da Silva; VIEIRA, Victor Barbieri Rodrigues. **Varredura pelo lado do cliente**: uma revisão sistemática. Belo Horizonte: Instituto de Referência em Internet e Sociedade, outubro de 2022. Disponível em: <bit.ly/3EAhEDF>. Acesso em: dd mmm aaaa.

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE

DIREÇÃO

Gustavo Rodrigues

Paloma Rocillo

MEMBROS

Ana Bárbara Gomes | Coordenadora de Políticas Públicas e Pesquisadora

Felipe Duarte | Coordenador de Comunicação

Fernanda Rodrigues | Coordenadora de Pesquisa e Pesquisadora

Juliana Roman | Pesquisadora

Júlia Caldeira | Pesquisadora

Lucas Samuel | Estagiário de pesquisa

Luiza Dutra | Pesquisadora

Paulo Rená da Silva Santarém | Pesquisador

Rafaela Ferreira | Estagiária de pesquisa

Thais Moreira | Estagiária de comunicação

Victor Barbieri Rodrigues Vieira | Pesquisador

Wilson Guilherme | Pesquisadore

irisbh.com.br

SUMÁRIO

Resumo executivo

O projeto **Comunicações privadas, investigações e direitos**, do Instituto de Referência em Internet e Sociedade – IRIS, busca oferecer subsídios confiáveis para que o debate político e jurídico de investigações em comunicações privadas no Brasil possa combinar segurança de tecnologias da informação e comunicação com proteção de direitos humanos e de garantias democráticas. Pretende-se analisar impactos e riscos; sistematizar conhecimento científico; e, ao final, produzir recomendações para setores público e privado. O objeto de análise são três mecanismos para investigações sobre comunicações privadas: rastreabilidade de mensagens instantâneas, *hacking* governamental, e varredura pelo lado do cliente.

Neste segundo relatório,[1] avaliou-se o cenário da **varredura pelo lado do cliente** (em inglês *client-side scanning*, com as iniciais CSS, aqui abreviada como VPLC). O termo se refere a técnicas de escaneamento realizado no dispositivo de usuários ("cliente") para identificação de instâncias de compartilhamento de materiais considerados ilícitos – especialmente envolvendo conteúdo sexual de abuso de crianças e adolescentes ou CSAM (*child sexual abuse material*) – em ambientes protegidos por criptografia segura, ao invés de realizar esse escaneamento ao nível de servidor. Por meio de uma revisão sistemática de literatura, investigou-se um total de 22 publicações selecionadas. Os achados foram organizados em contexto, conceito, funcionamento e problemas.

Primeiro, a **análise contextual** da VPLC expõe um conjunto de técnicas de detecção de conteúdo indesejado, suas limitações em sistemas com criptografia, bem como demonstra, de que forma, a proposta de 2021 da Apple para superar essa incompatibilidade, deu início às controvérsias que são o objeto central desse estudo. Segundo, explica-se como se pode **definir conceitualmente** e classificar tecnologicamente a varredura pelo lado do cliente, considerando o conjunto de técnicas. Terceiro, detalham-se, em linguagem acessível, as etapas e os procedimentos de **funcionamento** da VPLC. E, no quarto eixo, o universo de **problemas apontados** na literatura é agrupado de acordo com a natureza dos desafios à implementação dessa técnica: a) no **plano tecnológico**, questiona-se o funcionamento, a eficácia, a segurança e o escopo da VPLC, em razão dos riscos respectivos, isolados ou cumulados, de ela ser inutilizada, não alcançar os resultados propostos, abrir brechas de vulnerabilidade ou sofrer desvio de função; e b) no **plano jurídico**, discutem-se os possíveis efeitos negativos da implementação da VPLC sobre a privacidade, o sigilo das comunicações, a presunção de inocência, e a segurança pública, bem como se a técnica atende aos princípios de proporcionalidade e necessidade.

Somando análises e problematizações, na literatura revisada verifica-se a prevalência das vulnerabilidades tecnológicas não resolvidas e das carências de justificativas jurídicas compatíveis com os riscos gerados pela VPLC. As defesas da implementação, mirando o problema da difícil detecção de CSAM em sistemas com criptografia, parecem menosprezar a importância dos devidos cuidados para uma efetiva proteção contra

existentes fragilidades tecnológicas e prováveis violações de direitos das pessoas em geral, inclusive crianças e adolescentes.

Tais resultados confirmam a percepção de que a VPLC, como suposta medida alternativa à quebra da criptografia, é uma medida inadequada até mesmo para o objetivo de enfrentar a pornografia infantil. Fora de limites bem delineados, não há evidências suficientes de funcionamento tecnológico efetivamente robusto ou livre de uma alta suscetibilidade a ataques, ao passo em que não se resolveram graves riscos jurídicos gerais e pontuais, além de questões econômicas e sociais.

Apresentação

Os primeiros debates sobre a criptografia forte envolviam a inserção de mecanismos para acesso excepcional das agências estatais de investigação e persecução penal aos algoritmos criptográficos. Contudo, a sociedade civil e a comunidade técnico-científica foram bem sucedidas na defesa de que políticas de segurança pública deveriam considerar riscos tecnológicos, jurídicos e econômicos.

Esses setores demonstraram que as ferramentas de quebra da criptografia – demandadas para investigações legítimas por agentes públicos – abririam brechas para o acesso também por terceiros mal intencionados, além de não impedir que pessoas interessadas em fugir das autoridades migrem para outras plataformas que não tenham essas brechas. O resultado seria a população em geral com menos segurança e os suspeitos intocáveis.[2] Tais argumentos diminuíram as demandas por soluções como portas clandestinas (*backdoors*).[3] Surgiram, contudo, alternativas legislativas à quebra da criptografia, para dar às autoridades acesso a dados e informações supostamente necessárias para identificar e punir criminosos.

O projeto **Comunicações privadas, investigações e direitos** busca sistematizar a literatura sobre métodos alegadamente alternativos à quebra da criptografia, para nutrir o debate científico, político e jurídico sobre o tema no Brasil. Pretende-se oferecer subsídios confiáveis para que decisões políticas, regulatórias e judiciais combinem a segurança das tecnologias de informação e comunicação com a proteção de direitos humanos e garantias democráticas. Em específico, objetiva-se: 1) analisar impactos e riscos à segurança de dados e informações digitais, e direitos envolvidos; 2) sistematizar conhecimento sobre técnicas de investigação; e 3) produzir recomendações para o Estado e empresas.

Os relatórios científicos analisarão três métodos alternativos: a) rastreabilidade de mensagens instantâneas, na qual se guardam metadados da comunicação para futura identificação do caminho ou da origem de um eventual conteúdo ilícito; b) varredura pelo lado do cliente, pelo qual se analisa e compara um conteúdo em um dispositivo com bases de dados prévias, em busca de um padrão específico; e c) *hacking* governamental, pelo qual se exploram vulnerabilidades ocultas e não-intencionais de um sistema.

A partir dos resultados, o Instituto de Referência em Internet e Sociedade – IRIS pretende dialogar com diversos setores e construir posicionamentos sobre esses métodos, com base em evidências científicas e no respeito aos direitos humanos. O material será disponibilizado online, para consulta e uso geral.

1. Introdução

Na segunda metade do século XX, o debate sobre a disponibilidade pública de criptografia forte para a proteção de comunicações privadas orbitou a inserção de mecanismos de exceção, que viabilizassem o acesso de agências estatais de investigação e persecução penal ao conteúdo protegido. O tema rendeu grandes controvérsias públicas sobre efeitos jurídicos, políticos e econômicos, em conflitos na governança da criptografia forte conhecidos como guerras criptográficas (*crypto wars*), marcadas pela oposição da comunidade técnico-científica, do setor privado e de ativistas de direitos humanos na área digital contra diversos arranjos de acesso excepcional.[4]

Conquanto persista a pressão pública de autoridades de diversos países por tais mecanismos,[5] a década de 2010 viu novos tipos de propostas. Com a promessa de combinar a segurança dos sistemas e meios para investigações de dados e informações exigidos para identificar e punir criminosos, elas abarcam técnicas de *hacking* governamental, de rastreabilidade de mensagens instantâneas com criptografia e de varredura pelo lado do cliente, objeto deste estudo.

A discussão sobre mecanismos de varredura pelo lado do cliente (muitas vezes citados na literatura em inglês *client-side scanning* – CSS, aqui abreviada pelas iniciais em português VPLC) destacou-se em 2021, quando a Apple anunciou novos recursos de segurança para crianças e adolescentes, a fim de combater a pornografia infantil e o aliciamento. A proposta envolvia a comparação de imagens salvas no iCloud Photos com um banco de materiais de abuso sexual infanto-juvenil. O que se identificasse como Conteúdo de Abuso Sexual Infantil[6] (*Child Sexual Abuse Material*) – CSAM seria computado e, ultrapassada uma quantidade limite de correspondências (30, inicialmente)[7], submetida a checagem humana.[8] Esse monitoramento permanente foi criticado por conflitar com a proteção de confidencialidade oferecida pela criptografia. Se a proposta da Apple, supostamente, não reduzia seus padrões de segurança criptográfica forte, seriam os mecanismos de VPLC legais, eficientes e consistentes?

A pergunta central da presente investigação é: de que forma a literatura acadêmica pertinente vê a adequação da VPLC como meio investigativo em sistemas com criptografia forte sem mecanismos de acesso excepcional? O estudo explora riscos e desafios tecnológicos e jurídicos, organiza os principais pontos de defesa e crítica à proposta, e avalia sua viabilidade, pela revisão sistemática de literatura de 22 textos selecionados à luz do cenário atual de debate, dos idiomas português e inglês, e da técnica computacional.

Refletindo prós e contras, conforme o peso político e os parâmetros legais do debate, os resultados compõem quatro seções: contexto da controvérsia, conceito de VPLC, funcionamento da tecnologia, e problemas apontados. Ainda, o Apêndice 1 lista as obras analisadas e o Apêndice 2 replica o formulário de análise (o mesmo utilizado no primeiro relatório, sobre rastreabilidade de mensagens instantâneas).

2. Metodologia

Diversamente do notável acúmulo de estudos detalhando[9] riscos e impactos do acesso excepcional à criptografia, as supostas alternativas carecem do mesmo escrutínio. Em específico, a varredura pelo lado do cliente que alcançou proeminência pública internacional a partir do anúncio de que seria implementada pela Apple.

O alegado escopo de enfrentamento à divulgação de material de exploração ou abuso sexual infantil não prescinde de uma discussão política e jurídica – sobre se as investigações de comunicações privadas poderiam combinar segurança de TICs e proteção de direitos humanos – fundada em bases técnicas e acadêmicas consistentes, seja para ações pelo setor privado, seja para políticas públicas.

Para alcançar esse patamar de densidade e confiabilidade, realizou-se uma revisão sistemática de literatura: investigou-se o estado da arte sobre o tema, com recorte empírico em um grupo de obras selecionadas e avaliadas mediante critérios e procedimentos explícitos e organizados. Esse método se propõe a identificar eventuais lacunas em estudos acadêmicos de certo campo ou temática,[10] bem como, questões e subtemas para novas investigações e projetos. Pesquisas assim:

> *[...] são particularmente úteis para integrar as informações de um conjunto de estudos realizados separadamente sobre determinada terapêutica/intervenção, que podem apresentar resultados conflitantes e/ou coincidentes, bem como identificar temas que necessitam de evidência, auxiliando na orientação para investigações futuras.*[11]

Neste estudo, analisou-se *corpus* documental das seguintes fontes: busca por palavra-chave e coleta de referências bibliográficas de duas obras relevantes selecionadas.

Primeiro, operou-se uma busca pela palavra-chave "*client-side scanning*" na plataforma Google Acadêmico[12]. Embora o relatório adote a tradução do termo "varredura pelo lado do cliente", a busca pelo termo em inglês se justifica pela escassez de bibliografia ampla em português, bem como pela proeminência dos debates em inglês impulsionados pelas ferramentas anunciadas pela Apple. Foram encontradas 38 referências, sendo excluídas 3 entradas repetidas e 2 textos de acesso restrito.

O segundo e o terceiro subconjuntos de obras vieram das referências bibliográficas constantes de dois textos escolhidos de modo discricionário: "*Bugs in our Pockets: The Risks of Client-Side Scanning*", de ABELSON e outros,[13] e "Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta", do Center For Democracy & Technology (com tradução em português). Esses dois textos, já conhecidos da equipe, foram escolhidos pela relevância no debate sobre métodos alternativos à quebra da criptografia, e por abordarem controvérsias envolvendo a varredura pelo lado do cliente. Todas as 68 referências citadas no primeiro texto foram selecionadas; e, dos trechos do segundo texto que abordavam especificamente a VPLC, foram eliminadas 2 referências, por repetição, somando-se 9 obras ao *corpus*: um subtotal de 77.

A intenção foi aprofundar a discussão da varredura pelo lado do cliente, considerando o baixo número de resultados da primeira fase, provavelmente decorrente da "novidade" do tema, que inviabiliza a abordagem acadêmica de mais fôlego, em monografias, dissertações e teses.

Assim como no relatório anterior, admite-se que a sistematicidade do estudo é reduzida por esse passo, o que afeta a representatividade dos resultados em função da subjetividade da seleção pela equipe de pesquisa. Todavia, essa desvantagem foi tida como compensada pelo ganho de subsídios para reflexão sobre o tema, e atenuada por se tratar de um procedimento em grupo, com revisão cega pelos integrantes da equipe.

As 110 referências foram, então, sujeitas a avaliação preliminar de pertinência temática e formal. Título, resumo (se presente) e seção inicial foram lidos por dois pesquisadores, que votaram pela inclusão ou exclusão. Nessa fase, para reduzir vieses, os pesquisadores não tinham acesso aos votos um do outro. Havendo dissenso, a decisão era do terceiro pesquisador, que também não sabia como haviam votado os demais pesquisadores. Além das obras que não tratavam propriamente do tema da VPLC (49) e da barreira do idioma (1 texto em alemão), deliberou-se por excluir do *corpus* obras não acadêmicas: apresentações de telas (4), artigos de opinião (3), código de programa (1), currículo (1), entrevista (1), lista de obras (2), notícias jornalísticas (5), palestras (3), política interna (1), pronunciamento oficial (1), propostas normativas (2) e publicações de *blog* (14).

As 22 obras restantes foram então integralmente lidas, analisadas e inseridas em um formulário, com categorização (artigo, dissertação, capítulo de livro, etc); resumo; observações do pesquisador responsável; e citações em destaque. Gerou-se, então, uma síntese descritiva, orientada a identificar, em cada obra: proposta, metodologia (ou sua ausência), eventuais referências relevantes (citadas como base para o conceito ou posicionamento do trabalho), e qual a abordagem sobre a varredura pelo lado do cliente.

Ressalte-se a limitação do impacto das fases 2 e 3 sobre o resultado final: a maioria das obras de fato analisadas (68,18%) adveio da busca por palavras-chave, método não-discricionário. Essa particularidade – montante de textos selecionados conforme

um dado método, ou inseridos por via discricionária – mitiga o dano à sistematicidade e preserva a replicabilidade: o estudo pode ser reproduzido sem essas obras, ou na íntegra.

Assim, o *corpus* documental final abrange 22 publicações, listadas no Apêndice 1: 3 obras publicadas em anais de eventos, 7 artigos científicos, 3 notas técnicas e 12 relatórios.

Adicionalmente, o relatório conta com referências bibliográficas que não compuseram o *corpus de pesquisa*, por sua pertinência na contextualização do momento de emergência do debate sobre a VPLC em torno do anúncio da Apple, em 2020, ou na explicação do funcionamento das técnicas de detecção de conteúdo, ou, ainda, na fundamentação de opções metodológicas. Todavia, essas referências adicionais não compuseram o objeto da análise que buscou responder à pergunta central deste estudo.

3. Resultados

Os resultados da revisão sistemática da literatura selecionada foram dispostos em quatro eixos: contexto; conceito; funcionamento; e problemas.

3.1. Contexto

A expansão do acesso à Internet mundo afora é acompanhada de um aumento nos índices de práticas ilegais online. Dentre elas, o compartilhamento de material ilícito, a exemplo de conteúdo de abuso sexual infantil – CSAM, é tópico de constante preocupação para autoridades públicas. Nesse sentido, várias técnicas e ferramentas investigativas categorizam materiais ilícitos online, fazem moderação, identificam remetentes, e providenciam a persecução de envolvidos.

Nesse contexto, apresentam-se a seguir algumas técnicas para detecção de CSAM, a limitação de sua aplicação em sistemas com criptografia e como a Apple pretendeu apresentar soluções que impulsionaram o debate sobre a VPLC.

3.1.1. Técnicas de detecção de CSAM

Uma via simples é a análise de texto, para reconhecer palavras-chave (pela mera forma) ou, com mais complexidade, reconhecer padrões (pelo significado), no processamento de linguagem natural, cujo teor indique aliciamento ou extorsão. Há *aprendizado de máquina* para detectar atributos indicativos de CSAM em texto, imagem e vídeo, buscando padrões e com atribuição de valor semântico aos dados. E, ainda, há a classificação de CSAM por meio de *modelos de correspondência*, que atribui identificadores (*hashes*) ao material analisado e permite às plataformas avaliar imagens e vídeos compartilhados, mediante o cotejo – por verificação criptográfica de identidade ou por verificação perceptiva de semelhança (ver 5.3) – do *hash* atribuído ao conteúdo em análise com a base de *hashes* previamente atribuídos a conteúdos ilegais já conhecidos.

As ferramentas API Content Safety da Google,[14] PhotoDNA da Microsoft,[15] Safer da Thorn,[16] e PDQ Hash da Meta[17] são soluções de detecção automatizada usados em redes sociais, fóruns, bancos de imagens e vídeos muito populares:[18] *hash* perceptivo impede o retorno de perfis banidos no serviço de relacionamento social OkCupid; o SimSearchNet/ SimSearchNet++ verifica correspondência com material de desinformação, inclusive sobre COVID-19, em imagens no Instagram e no Facebook; e identificadores *hash* de áudio evitam o acionamento involuntário da "Alexa" por anúncios.

Tais técnicas de detecção atuam em conteúdo armazenado nos bancos de dados das plataformas, ou por elas transmitido, sendo denominadas de *server-side scanning* – SSS (em português, "varredura pelo lado do servidor" – VPLS).[19] São soluções prontamente aplicáveis a sistemas sem criptografia de ponta a ponta, com conteúdo acessível e checável por plataformas e autoridades estatais.

3.1.2. Detecção em sistemas com criptografia

Mas a aplicação dessas técnicas VPLS não é trivial em ambientes digitais com criptografia assimétrica, nos quais os participantes da comunicação usam protocolos de chaves públicas de cifragem e chaves privadas de decifragem. Sob tal segurança criptográfica, por definição, o conteúdo só pode ser conhecido pelos legítimos remetentes e destinatários, salvo pela inserção de eventual vulnerabilidade nos algoritmos criptográficos, medida que o Alto Comissariado das Nações Unidas aponta, há anos, como fonte de graves riscos a garantias fundamentais, como a privacidade, a liberdade de expressão e de reunião pacífica, a proteção de dados pessoais, entre outras.[20]

Em setembro de 2020, foi divulgada sem autorização uma minuta interna da Comissão Europeia[21] sobre as dificuldades de se detectar CSAM em comunicações criptografadas de ponta a ponta, mantendo privacidade e segurança. Sem abordar os aspectos políticos decorrentes, o rascunho descreveu meios tecnológicos para essa identificação (portas clandestinas, varredura pelo lado do cliente, enclave seguro, classificação, computação homomórfica, etc.) e os analisou em cinco critérios: eficácia, viabilidade, privacidade, segurança e transparência.

Tais meios de detecção de CSAM foram classificados conforme o elemento básico das comunicações criptografadas que elas afetam: a técnica, o servidor, ou o dispositivo do usuário.[22] A detecção na *técnica criptográfica* interfere no protocolo, por meio da chamada "criptografia" homomórfica, que ainda pode ser combinada com aprendizado de máquina e classificadores, aplicada integralmente nos servidores, criando *hashes* para conteúdo criptografado antes do envio, de acordo com as técnicas do quadro a seguir:

TABELA 1 MEIOS DE DETECÇÃO DE CSAM EM SISTEMAS COM CRIPTOGRAFIA			
ELEMENTO AFETADO	**TÉCNICA**	**SERVIDOR**	**DISPOSITIVO**
Modalidade utilizada	"criptografia" homomórfica	a) enclave no servidor da plataforma b) enclave no servidor de um terceiro	a) tudo no dispositivo; b) *hash* no dispositivo e verificação no servidor c) parte do *hash* no dispositivo e parte no servidor, e verificação no servidor d) classificadores no dispositivo.

Conforme o quadro, as técnicas de detecção *no servidor* podem realizar algumas ou todas as operações sem criptografia de ponta a ponta da VPLS em um reduto, ou enclave seguro a) no servidor da plataforma, ou b) no servidor de um terceiro; c) no servidor de múltiplos terceiros. Já as técnicas de detecção *no dispositivo* poderiam realizar algumas ou todas as operações sem criptografia de ponta a ponta da VPLS, sendo categorizadas em quatro modalidades: a) integralmente realizadas no dispositivo; b) identificação integral no dispositivo e correspondência no servidor; c) identificação parcial no dispositivo e identificação parcial e correspondência no servidor; d) uso de classificadores no dispositivo. Esse último conjunto recebe o nome de *client-side scanning* – CSS, ou "varredura pelo lado do cliente" – VPLC.

A comunidade técnica havia criticado os pressupostos e as considerações da minuta europeia vazada em 2020,[23] que se propunha a uma análise técnica das possibilidades de detecção, assumidamente sem enveredar pelas questões políticas e jurídicas envolvidas.[24] Em especial, lembrou-se que em julho de 2020 havia sido aprovada pelo Parlamento Europeu uma norma provisória para o tema, com vigência de 3 anos, autorizando o escaneamento voluntário do conteúdo de comunicações como esforço proativo das plataformas digitais – regra que já havia sido criticado por contrariar previsões normativas da própria União Europeia.[25]

Até então, várias modalidades alternativas à quebra da criptografia vinham sendo questionadas em seus impactos sobre o sigilo e a segurança das comunicações. Com o anúncio da Apple, a VPLC tomou o centro do debate:[26]

> *Em agosto de 2021, a Apple anunciou planos para introduzir tal sistema para seus serviços iMessage e iCloud, mas suspendeu a implementação da mudança proposta após fortes críticas de uma ampla gama de especialistas em segurança de tecnologia da informação, criptógrafos e grupos de direitos humanos.*

Embora a proposta da Apple não seja o objeto central desta pesquisa, faz-se oportuno narrar em detalhes os acontecimentos entre a apresentação e a suspensão da proposta que ensejou tanta controvérsia em torno da VPLC.

3.1.3. Apple e a proteção da infância

Em 5 de agosto de 2021, a Apple anunciou que ainda naquele ano adotaria três mudanças[27] nos seus sistemas operacionais (iOS 15, watchOS 8, iPadOS 15 e macOS Monterey), a fim de aprimorar o enfrentamento a materiais de abuso sexual infantil, ou CSAM (sigla para o termo em inglês *Child Sexual Abuse Material*).[28]

Primeiro, a "**Segurança das comunicações em mensagens**"[29] interferiria no dispositivo de crianças e adolescentes, "embaçando" as imagens com teor sexual explícito recebidas ou enviadas via iMessage. O app então pediria uma confirmação para permitir a visualização. Especificamente para crianças, com menos de 13 anos, se a família optasse por ser notificada, o iMessage exibiria no celular um segundo pedido de confirmação, com o aviso à criança de que a família seria comunicada, mas sem expor a imagem.

Com a segunda mudança, "**Detecção de CSAM**", antes de serem enviadas para o iCloud Photos, as imagens seriam convertidas em *hashes* e, no dispositivo, comparadas com os *hashes* de materiais de abuso sexual infantil de um banco de dados fornecido pela ONG estadunidense NCMEC – *National Center for Missing and Exploited Children* ("Centro Nacional para Crianças Desaparecidas e Exploradas").

Por fim, com a "**Orientação ampliada na Siri e na Busca**", a empresa adicionaria comandos que exibiriam orientações de segurança e informações de contatos para denúncia, quando houvesse pesquisas, por texto ou voz, envolvendo materiais de abuso sexual infantil.

As duas primeiras medidas atuariam tanto no dispositivo quanto no servidor: a base de dados dos *hashes* ilícitos seria embarcada no sistema operacional dos aparelhos da fabricante, o que possibilitaria a identificação; e posteriormente seria feita a verificação de correspondência nos servidores do NCMEC.

Houve muitas críticas,[30] além de algumas cartas abertas solicitando que as mudanças não fossem adotadas.[31] Apontadas como "portas clandestinas" (*backdoors*), as duas primeiras ferramentas, para iMessage e iCloud Photos, foram acusadas de fragilizar a segurança prometida pela adoção de criptografia segura. A empresa redarguiu que os riscos dos novos produtos seriam mínimos em termos de privacidade, por não expor, nem à Apple, nem a terceiros, as comunicações dos usuários: os responsáveis só saberiam da existência de risco, mas não do conteúdo, que não seria arquivado em nenhum outro local, além do próprio dispositivo que enviou ou recebeu a imagem.

As críticas seguiram e, em 3 de setembro, a Apple comunicou que adiaria a adoção da ferramenta de detecção de CSAM. O anúncio foi celebrado, mas não encerrou pedidos de que fossem abandonados completamente os "planos de vigilância".[32] Um ano depois, em outubro de 2022, a sessão "Child Safety" no site oficial[33] informa que a ferramenta de segurança no iMessage é opcional e não comunica a família nem qualquer terceiro, oferecendo à criança ou adolescente a possibilidade de buscar ajuda de alguém de sua confiança; e que, para a terceira ferramenta as orientações e informações estão também no Spotlight (sistema de busca no desktop). Não há mais nenhuma menção ao mecanismo de detecção de CSAM, cuja adoção, aparentemente, segue suspensa sem prazo previsto para retornar a agenda da empresa.

O caso da Apple ilustra a questão central deste estudo. A despeito de críticas ou elogios à proposta específica ou à postura da empresa, merece compreensão a ideia de se detectar conteúdos indesejados a partir dos dispositivos dos usuários, a fim de viabilizar meios de investigação em ambientes criptografados. Mas como se pode definir o conceito VPLC? Como funcionaria essa tecnologia? E quais os riscos?

3.2. Conceito de VPLC

Em seu universo, as ferramentas de análise automatizada de conteúdo podem ser classificadas em dois tipos de modelos: de correspondência ou de predição.[34] O primeiro modelo pode ser realizado por verificação criptográfica ou preceptiva, com o fim de reconhecer um conteúdo como idêntico ou semelhante; enquanto isso, no segundo modelo, pretende-se reconhecer as características de um conteúdo aprendido previamente pela máquina. Ambos modelos podem operar tanto em varredura pelo lado do servidor quanto pelo do cliente.

TABELA 2
MECANISMOS PARA ANÁLISE AUTOMATIZADA DE CONTEÚDO

<table>
<tr><td colspan="2">modelos de correspondência

reconhecer o conteúdo como idêntico ou suficientemente semelhante ao conteúdo visto anteriormente</td><td rowspan="2">modelos de predição
reconhecer características do conteúdo com base em aprendizado de máquina prévio</td></tr>
<tr><td>verificação criptográfica

identificação de correspondência a partir de hash altamente sensível a alterações</td><td>verificação perceptiva

identificação de correspondência a partir de hash por determinado grau de semelhança</td></tr>
<tr><td colspan="2">• varredura pelo lado do servidor: conteúdo é convertido em hash e enviados à plataforma para avaliação

• varredura pelo lado do cliente: conteúdo é convertido em hash e avaliado no próprio dispositivo</td><td>• classificação por algoritmos pelo lado do servidor

• classificação por algoritmos pelo lado do cliente</td></tr>
</table>

A **varredura pelo lado do cliente**, aqui abreviada como VPLC (muitas vezes citada no termo em inglês *client-side scanning*, com as iniciais CSS) – também citada como "correspondência de *hash* perceptivo" (*perceptual hash matching*)[35] "escaneamento do cliente" ou "filtragem na ponta"[36] – é apresentada como uma opção alternativa ao acesso excepcional, que depende da cooperação com a plataforma.

Pode-se definir a VPLC assim:[37]

> *[...] o conceito de que, através de certas formas de implementação tecnológica, um sistema poderia ser desenvolvido para digitalizar fotografias e mensagens antes de serem enviadas de um usuário (ou depois de recebidas por outro usuário) a fim de determinar se as imagens ou mensagens em questão violam proibições legais.*

Destarte, o cerne da VPLC é a filtragem ou a moderação de conteúdo no âmbito do dispositivo tecnológico particular, em um aparente caráter menos invasivo, que se pretende compatível com a criptografia de ponta a ponta e a proteção de dados pessoais.

Assim, a VPLC consiste na análise de correspondência de dados digitais, tanto em fluxo (saindo ou chegando) ou em armazenamento, realizada no dispositivo, como parte de um sistema de comunicação, criptografado ou não. Nesse sentido, a análise de correspondência é realizada com uma lista de conteúdos "indesejados", por efetiva ilicitude ou mera incompatibilidade com a política do intermediário.

É importante destacar que a varredura pelo lado do cliente consiste numa proposta tecnológica de reconhecimento de conteúdo que se pretende adotar pelas plataformas, e não numa funcionalidade já existente nos dispositivos de comunicação. Os efeitos da adoção dessa proposta se tornaram uma parte sensível da controvérsia. Debate-se intensamente sobre os métodos disponíveis para sua concretização, bem como sua eficácia para a finalidade pretendida, os riscos de segurança decorrentes e a compatibilidade da medida com os atributos da criptografia de ponta a ponta.

A VPLC se distingue da Varredura Pelo Lado do Servidor (VPLS), na medida em que esta última realiza a análise de conteúdo durante a transmissão ou a partir do armazenamento pelo intermediário do sistema de comunicação, de modo que amplia-se o risco à privacidade do usuário.[38] Por não armazenar, nem intervir na transmissão do conteúdo, a VPLC foi visualizada pela Apple como uma forma de combate a conteúdos ilegais, em específico CSAM, sem a quebra da privacidade do usuário e da criptografia de ponta a ponta.[39]

Podem-se identificar seis fases da VPLC aplicada à moderação de conteúdo, que podem ocorrer sucessiva ou simultaneamente:[40] 1) definição do que é ou não permitido no serviço; 2) detecção do conteúdo gerado pelo usuário potencialmente irregular à luz de políticas internas ou da legislação; 3) avaliação da irregularidade do conteúdo; 4) intervenção contra o conteúdo identificado como irregular; 5) recurso contra a decisão de intervenção; e 6) educação sobre a política de moderação de conteúdo.

O diferencial otimista da VPLC seria auxiliar empresas de tecnologia a atuarem contra a pornografia infantil e outros tipos de comunicações ilícitas, mas sem o custo de degradar os sistemas de criptografia forte.[41] A VPLC supostamente poderia viabilizar o combate ao conteúdo indesejado, "*mas também manter a privacidade do telefone dos usuários antes de ser criptografado*",[42] inclusive com a premissa de que a análise no dispositivo seria sistematicamente mais eficiente "*do que fornecer acesso excepcional para aplicação da lei*"[43]

Ainda, na VPLC pareceria vantajosa a "*oportunidade para ativistas e outros detectarem alterações no software do lado do cliente e entenderem seus efeitos, adicionando alguma transparência e responsabilidade*".[44] Seria possível capacitar usuários para entender e controlar sua experiência online, sem apenas abrir as portas para a censura automatizada.

3.3. Funcionamento

O exemplo mais difundido atualmente teve como justificativa a busca por conteúdo de exploração sexual infantil. Nestes casos, a VPLC compara o material com a lista de *hashes* e, se houver correspondência, o sistema pode não enviar a mensagem, relatar a tentativa (às autoridades públicas ou a uma organização da sociedade civil), ou combinar essas ações.[45] O primeiro passo para o funcionamento de tal recurso é a comparação de um conteúdo "contra uma base de dados pré-definidas de conteúdos danosos, os quais estariam sinalizados com identificadores únicos",[46] os chamados *hashes*.

As funções de *hash* são projetadas para converter um conjunto de dados (ou entrada), como uma imagem ou um arquivo de texto, em um conjunto curto de caracteres de tamanho padronizado (ou saída), chamado *hash*, algo como uma "impressão digital" de um dado conteúdo.[47] Se implementadas mediante criptografia, tais funções agregam atributos, a exemplo da inviabilidade de duas entradas distintas serem convertidas numa mesma saída.[48]

A finalidade é detectar correspondência não apenas entre cópias exatas, mas também entre mídias semelhantes (por exemplo, uma imagem em versão redimensionada):[49]

> *Os hashes perceptivos são diferentes dos hashes criptográficos, pois o primeiro muda gradualmente à medida que a imagem muda, enquanto o último muda significativamente assim que um único pixel muda. É importante ressaltar que os hashes perceptivos são projetados para detectar instâncias de uma mídia visual que são visualmente semelhantes (por exemplo, uma versão redimensionada) sem serem cópias exatas.*

Ao longo dos últimos anos inúmeras ferramentas foram desenvolvidas a partir da filtragem do tráfego e do conteúdo da web, com fundamento na existência de determinadas palavras-chave, metadados ou padrões pré-estabelecidos.[50] Por exemplo, para detectar imagens associadas a material protegido por direitos autorais ou pornografia infantil, como os algoritmos de correspondência de *hash*, "eles identificam as imagens por um código único – uma espécie de "impressão digital" para uma determinada imagem – chamado hash, e as comparam com o hash de imagens conhecidas com direitos autorais ou pornografia infantil".

Destarte, no caso da Apple, as imagens dos usuários seriam convertidas em *hashes* pelo sistema NeuralHash e comparadas com os *hashes* de conteúdo do banco de dados de materiais de abuso sexual infantil, mantido pela ONG Centro Nacional para Crianças Desaparecidas e Exploradas dos Estados Unidos e por outras organizações de segurança infanto-juvenil:[51]

> *Um sistema de detecção chamado NeuralHash cria identificadores que podem ser comparados com IDs do Centro Nacional para Crianças Desaparecidas e Exploradas e outras entidades para detectar conteúdo conhecido de CSAM nas bibliotecas de fotos do iCloud. A maioria dos provedores de nuvem já verifica as bibliotecas de usuários para essas informações - o sistema da Apple é diferente porque faz a correspondência no dispositivo e não na nuvem. (tradução livre)*

Exemplifica-se como a tecnologia funcionaria no caso da verificação de um arquivo criptografado contendo imagens de abuso sexual infantil por um aplicativo de comunicação: "A imagem seria comparada a uma lista de imagens ilegais conhecidas e interditada antes que a imagem fosse enviada".[52] No caso do WeChat, por exemplo, a censura de conteúdos em geral acontece em tempo real.[53]

No caso da Apple,[54] para proteção da privacidade do cliente, o sistema trabalharia com uma correspondência de *hashes*, ao invés de digitalizar as imagens na nuvem. O mero fato de encontrar correspondências entre os materiais, não geraria – a priori – nenhuma informação a ser prestada à empresa, nem ao usuário e tampouco às instituições de responsabilização. Entretanto, gerar-se-ia um registro para acompanhamento de outros materiais em caso de reincidência de CSAM, a partir dos denominados *vouchers* informacionais, ou "recibos". Deste modo, seria adotada a técnica *threshold secret sharing*, que só permite tratar os dados após ser excedido, por parte de um mesmo usuário, um limite de correspondências previamente estabelecidas.

Esse recibo seria armazenado com a imagem na conta do iCloud Photos. No caso das reincidências excederem um "limite de correspondências" (em quantidade não divulgada pela empresa), o usuário e o material passariam por uma verificação humana, que descriptografaria somente as correspondências de CSAM, sem dar acesso às outras imagens. Assim, pessoas designadas apenas acessariam "derivados visuais" de baixa resolução das imagens (como uma imagem em baixa resolução, em preto e branco, ou em miniatura). Se confirmado o CSAM, a conta seria suspensa e a ONG pertinente responsável seria comunicada, com um relatório do procedimento.

A etapa de verificação humana dos conteúdos detectados previne o risco do falso positivo, mas não neutraliza os riscos graves da exposição de dados que sejam considerados lícitos. Como aponta o Center For Democracy And Technology,[55] a checagem do conteúdo por seres humanos representa em igual medida um risco à privacidade, pois todo acesso por porta clandestina fragiliza a segurança do usuário, que pode passar a ser vítima não só de ataques terroristas, mas até de ataques políticos institucionais. Desta maneira, verifica-se que um meio pensado para enfrentar violências sexuais contra crianças e adolescentes pode servir para lesar outros direitos humanos.

Com a finalidade de mitigar os riscos de ataques virtuais que buscam acessar as imagens armazenadas com recibos, bem como a alteração dos dados de um recibo armazenado, a Apple informou que a base de dados obtida pelo cruzamento das imagens com o CSAM seria inserida de forma criptografada dentro do próprio sistema operacional dos aparelhos.[56] Dessa forma, a lista não poderia ser atualizada de maneira independente pela internet, o que aumentaria a segurança do sistema.

É importante destacar que para gerar a base dos h*ashes* seria preciso ainda o cuidado com os materiais utilizados para evitar conteúdos infiltrados e/ou maliciosos. Para tanto, a proposta da Apple exigia a comparação dos materiais com uma lista de hashes apontados como indesejados por duas instituições não governamentais que operassem em jurisdições diferentes,[57] não se valendo apenas da base da NCMEC. Essa medida buscava impedir o uso do sistema por autoridades estatais para fins de censura ou vigilantismo por algum governo.

A VPLC do caso Apple se mostrava ainda como uma tecnologia híbrida, de escaneamento de cliente e servidor. Assim, pode-se especular que se o usuário não usasse o iCloud Photos estaria isento do cruzamento dos dados para detecção de CSAM.

Para Erik Neuenschwander,[58] as ferramentas pretendiam responder a pressões de diversas instituições quanto ao acesso a informações criptografadas, sob alegações de combate a atividades de terrorismo ou, como é o caso, de CSAM, ao mesmo passo em que se protegesse a privacidade dos usuários.

Na proposta formulada por REIS e outros,[59] por exemplo, a VPLC poderia ser utilizada para detectar, a partir de *hashes*, conteúdos rotulados previamente como desinformação por instituições de checagem de fatos. A solução poderia ser implementada nos dispositivos de quem envia, com a vantagem de detectar e limitar a distribuição de conteúdos indesejados; nos de quem recebe, auxiliando contra o aliciamento; ou em ambos, viabilizando uma abordagem de acareação, que permitiria identificar interferências no sistema.[60]

3.4. Problemas

Na contramão do que foi alegado pela Apple, as críticas na literatura acadêmica acusam a defesa da adoção da VPLC de um otimismo ingênuo e acrítico, e apresentam preocupações de várias ordens – nenhuma delas solucionada pelas propostas de VPLC.

Do ponto de vista **sociológico**, encontrou-se o fato de que a automatização da busca por qualquer tipo de conteúdo indesejado afeta somente a difusão do material, mas não elimina a sua fonte:[61]

> *as soluções tecnológicas para detectar conteúdo problemático por si só não abordarão as questões mais amplas de, por exemplo, a distribuição de desinformação ou CSAM, que precisam identificar e abordar em seu núcleo os problemas sociais e políticos causas por trás desses fenômenos.*

Ainda, pontuou-se que interesses domésticos de cada nação ocasionariam **prejuízos econômicos,** nesse sentido a medida:[62] "poderia minar a competitividade de serviços nacionais",[63] ao reduzir o interesse geral nos produtos e serviços oferecidos pelas empresas de países cuja legislação exigisse a adoção obrigatória de VPLC.

Em abordagens mais profundas, as preocupações nas obras analisadas podem ser agrupadas em duas categorias: tecnológicas e jurídicas.

3.4.1. Aspectos tecnológicos

Na visão tecnológica, pode-se afirmar que "os sistemas de correspondência de hash CSS não são tecnicamente robustos".[64] As inconsistências não respondidas pelos proponentes da VPLC podem ser vistas em função do objeto da vulnerabilidade, ou seja, de qual o atributo do próprio sistema que é afetado por um efeito negativo: o funcionamento, a eficácia, a segurança, o escopo.
Quanto ao **funcionamento**, diversamente dos servidores dos intermediários, a falta de padronização do "lado do cliente" implicaria grande diversidade de possíveis restrições para o melhor emprego do VPLC. Logo, o mecanismo de varredura em si não poderia sequer operar por uma baixa capacidade de armazenamento ou processamento no dispositivo cliente seja por falta de atualização dos softwares; por uma conexão ruim com a Internet; ou por falta de carga na bateria.[65]
Especificamente, a construção e atualização do banco de dados de conteúdos indesejados exigiria uma comunicação entre servidores e dispositivos potencialmente falha. Segundo Hua e outros,[66] "a escala dos bancos de dados torna proibitivo tanto enviar os hashes conhecidos como ruins para o cliente quanto, se os hashes forem sensíveis, aplicar técnicas 2PC para garantir o mínimo possível sobre os vazamentos de banco de dados para os clientes".[67]

Mesmo se o sistema operar corretamente, ainda pode falhar na eficácia: conteúdos novos ou comunicados pela primeira vez não seriam detectados como irregulares, por requerer prévia coleção e rotulação[68] para o reconhecimento por correspondência. Nesse sentido, o conteúdo indesejado precisaria circular mais de uma vez para que o VPLC seja eficaz – e essa não é a regra. Por exemplo, de todas as imagens de material de abuso sexual de crianças delatadas nos EUA, 84% foram denunciadas apenas uma vez[69] – agregando-se ainda a camada de subnotificação decorrente da limitação de recursos humanos e da capacidade manual para revisar todas as denúncias, cujo número cresceu de dez mil ao ano em 1998 para quase um milhão ao mês em 2017.[70]

Em suas limitações, a VPLC não propõe analisar o significado ou o contexto, sendo indiferente a situações de uso legítimo de obras protegidas por direitos autorais, ou eventual ausência de intenção criminosa[71]. Até por isso, notou-se fácil, para qualquer usuário do sistema de comunicação, inutilizar o sistema automatizado de VPLC, ao neutralizar ou contaminar conteúdos pela manipulação de dados digitais com técnicas de gradiente e transformação, mas sem afetar de modo relevante a sua percepção por seres humanos.[72]

Logo, mudanças de luminosidade, brilho ou inserção de leves ruídos gerariam falsos negativos, ao atrapalhar a correspondência ao hash indexado; ou poderiam ser modulados para gerar falsos positivos, forçando disparo de alarmes falsos com a correspondência de hashes em materiais inócuos. Aponta-se como especialmente preocupante que, a despeito do grau de transparência sobre o funcionamento do algoritmo de geração do hash (tratando-se ou não de uma "caixa-preta"), a necessidade de ampliação dos limites de detecção (para contornar tentativas deliberadas de evitar o reconhecimento de conteúdos indesejados) permitiria a geração de muitos falsos positivos.[73] Diz-se que reconhecer conteúdos por semelhança abre o risco de falsos positivos, com a detecção equivocada de uma correspondência no sistema de hashes perceptivos.[74]

Assim, quanto menos o sistema for sensível a mudanças sutis (luminosidade, cor, sombras, inversão, etc.) – ou seja, quanto menos ele considerar ruídos, a fim de detectar a correspondência entre imagens semelhantes com poucas diferenças –, mais o sistema se sujeita a dois problemas. Primeiro, essa insensibilidade permite mais falsos positivos, ou seja, identificar como correspondente um conteúdo que na verdade não corresponde; e, segundo, ela permite ainda a adversários provocarem uma elevação artificial da demanda pela análise de imagens inofensivas, alteradas propositalmente apenas na medida em que se produza um hash semelhante ao de imagens previamente reconhecidas como indesejadas.

Também o próprio volume crescente de compartilhamento de material de exploração sexual infantil pode representar uma dificuldade para o processamento tempestivo das denúncias.[75] Não seria uma tarefa trivial definir o nível aceitável de falsos positivos adequado para a implementação da VPLC em um sistema de comunicação com criptografia de ponta a ponta.[76]

E mesmo que o sistema funcione e seja efetivo, no tocante à **segurança**, em tese, o banco de dados de *hashes*, eixo central de operação do sistema, pode ser facilmente manipulado por agentes mal-intencionados.[77]

Embora uma postura mais transparente por parte da Apple pudesse gerar mais confiança ao usuário,[78] de maneira geral, em qualquer sistema, a adição de componentes agrega mais potencial de falhas de segurança.[79] No caso da VPLC, a despeito de haver ou não interferência direta na criptografia, "há necessária redução da segurança do sistema em decorrência da ampliação da superfície de ataque",[80] de tal modo que especialistas temem "riscos similares aos do acesso excepcional via backdoor",[81] a saber, "abuso, o efeito inibitório e os danos à confiança no ecossistema digital".[82]

Mesmo a defesa da VPLC contra assédio ou desinformação[83] ressalva que o *hash* perceptivo "pode não ser adequado para todas as classes de conteúdo abusivo, como CSAM, em que o destinatário pode ser um adversário".[84] Esses modelos – ao operarem por semelhança – foram diagnosticados, na análise da ferramenta Neural Hash, por exemplo, como "altamente suscetíveis a vários ataques, alguns deles triviais, que quebram o sistema".[85] Grover e outros destacam que o VPLC "confia no dispositivo do usuário final para calcular com veracidade o hash da mensagem".[86] Por esse modo de construção, afirma-se[87] que, em ambiente adversariais, os meios de VPLC são tão inseguros e vulneráveis que nem mesmo a Apple, com todo seu esforço de engenharia, conseguiu oferecer um projeto técnico confiável[88] ou à prova de brechas em seus sistemas.[89] É sintomático que um ano após o auge da controvérsia, até novembro de 2022, a gigante de tecnologia ainda mantenha a decisão de suspender a adoção da ferramenta de detecção de CSAM.

A propósito, se para o iOS, a Apple controla a tecnologia desde o *hardware* até o *software*, no ecossistema Android numerosas empresas diferentes produzem os aparelhos celulares e as variam muito as versões ativas do sistema operacional móvel. Mesmo assim, verificou-se não haver diferenças relevantes entre Android e iOS em várias dimensões relacionadas à privacidade do usuário, inclusive pela ausência de uma regulamentação estatal: o nivelamento da segurança se dá por baixo.[90] Portanto, diante da enorme quantidade de dispositivos em uso, de muitas marcas e modelos, podem-se esperar futuras falhas na atualização dos programas de detecção, prejudicando a segurança do funcionamento do próprio mecanismo.[91]

Contra o otimismo de que seria possível "encontrar uma solução simples para o problema"[92] da difusão de conteúdos nocivos como material de abuso sexual infantil, a reticência geral de especialistas às supostas propostas alternativas[93] se estende à varredura pelo lado do cliente: "uma percepção frequente foi de que haveria um comprometimento principiológico da criptografia mesmo sem uma interferência direta no algoritmo criptográfico ou no sistema de gerenciamento de chaves".[94] A promessa de vigilância limitada é tida por ilusória: titulares dos dados pessoais analisados não poderiam prever nem auditar a ação de autoridades.[95]

Finalmente, ainda que funcione bem, com eficácia e segurança, aponta-se o risco de que a VPLC tenha seu **escopo** ampliado para outros propósitos, indo além do enfrentamento material de exploração sexual infantil,[96] para abranger outros tipos de conteúdos, tais como desinformação,[97] terrorismo ou evasão de documentos estatais,[98] e cuja ilegalidade pode não ser tão evidente ou livre de discussão.

Seria viabilizada, por exemplo, eventual censura de mensagens políticas legítimas.[99] No pior dos cenários, pela ampliação da lista de conteúdos indesejados, "*a totalidade do dicionário poderia ser incorporada a essa base, efetivamente possibilitando a decifragem total das mensagens e nulificando o propósito da criptografia*".[100] E caberia aos intermediários a tarefa de resistir à pressão por expansão ou abusos.[101]

Não se deve descuidar da real possibilidade de o objeto da moderação ser questionável, pois a natureza da VPLC não garante nem exige a legitimidade no objetivo da aplicação, como se afirma consensualmente em relação à exploração infantil: "Hashes para outro conteúdo sensível, mas legal (como político ou sexual) podem ser adicionados ao banco de dados e sem o conhecimento do usuário".[102]

3.4.2. Aspectos jurídicos

Juridicamente, há uma forma de pensar a questão "que reconhece as estruturas técnicas dos sistemas criptográficos como indissociáveis das conotações políticas que adquiriram ao longo dos anos no que tange à defesa dos direitos humanos",[103] e seriam submetidos a altos riscos de graves ameaças pontuais ou mesmo a restrições gerais e sistematizadas.

A respeito do **sigilo das comunicações** e da **privacidade**,[104] afirma-se que, "se os resultados da comparação de hash forem compartilhados com o servidor, as garantias de privacidade da criptografia de ponta a ponta são violadas".[105]

Assim, em afronta ao princípio da **presunção de inocência**, a VPLC "deteriora a finalidade da criptografia de ponta a ponta relacionada à liberdade de informação e expressão, pois o conteúdo da comunicação é filtrado por padrão".[106] Aprofundando a questão tecnológica do escopo, do ponto de vista político, abre-se uma preocupação com outras modalidades de moderação, inclusive as que revelam menos cuidado com excessos ou abusos. Ao enfraquecer os argumentos em favor de serviços com criptografia de ponta a ponta,[107] a adoção da VPLC "acaba sendo um trampolim para mecanismos de backdoor/censura mais arriscados, porque uma vez implantado o primeiro, será mais fácil implantar ou justificar a implantação do último".[108] E mesmo se fosse efetiva a contenção de riscos numa democracia, sua implementação poderia legitimar ferramentas semelhantes em países habituados a censurar e vigiar.[109]

Na **segurança pública** da coletividade, se "é possível para um ator engenhoso adivinhar o conteúdo de uma mensagem a partir de seu hash",[110] organizações criminosas poderiam usar o sistema para ilícitos. Diante da viabilidade de práticas de engenharia reversa, a existências de um sistema com garantias de proteção não podem depender da confiança na ética de quem opera o sistema de detecção de conteúdo, pois não cabe exigir ética em ambientes adversariais, tais como no contexto do enfrentamento à CSAM.

Assim, a VPLC serviria como "um **modelo para vigilância em massa**, pois pode não ser possível para o usuário ou a sociedade civil monitorar a lista de hash usada por seu telefone para garantir que ele esteja apenas relatando ou impedindo a transmissão de imagens de abuso sexual infantil"[111] . E, mais grave, diante da amplitude, vigiando dados privados "de todo mundo, o tempo todo, sem ordem judicial ou suspeita",[112] não seria possível assegurar a aplicação estritamente regular e confiável, inclusive em relação à infância e juventude, pois "esses riscos afetarão todos os usuários de plataformas de comunicação digital; todas as crianças e todos os adultos do mundo, agora e possivelmente no futuro, e as consequências são difíceis de prever".[113]

Assim, considerando os direitos humanos como um todo, “essa vigilância em massa pode resultar em um significativo efeito amedrontador na liberdade de expressão e, de fato, na própria democracia”.[114] E para a sociedade civil organizada, também “a possibilidade de desvirtuamento de função do sistema representa um risco democrático”.[115] Logo, a partir de uma visão de prós e contras em confronto,[116] a gama de riscos seria **desproporcional** aos potenciais benefícios, bem como **desnecessária** para os objetivos pretendidos.[117]

4. Conclusão

Neste segundo relatório,[118] avaliou-se o cenário da **varredura pelo lado do cliente** (muitas vezes referenciada pelo termo em inglês *client-side scanning*, ou as iniciais CSS, aqui abreviada como VPLC). O termo se refere ao uso de técnicas de escaneamento dos dispositivos de usuários (clientes) para identificação de instâncias de compartilhamento de materiais considerados ilícitos em ambientes protegidos por criptografia segura, ao invés de se realizar esse escaneamento ao nível de servidor. Por meio de uma revisão sistemática, investigou-se um total de 22 publicações selecionadas.

O primeiro aspecto notório sobre as controvérsias e críticas relacionadas a técnicas de VPLC diz respeito a **aspectos tecnológicos**. As preocupações dessa natureza apontadas ao longo do estudo podem ser resumidas em quatro elementos analisados a partir das soluções apresentadas e propostas, até o momento da redação deste relatório: funcionamento, eficácia, segurança e escopo.

Primeiramente, quanto ao **funcionamento** dessas soluções, constata-se que a primeira barreira para a viabilidade da VPLC diz respeito à impossibilidade de se empregar essa tecnologia em qualquer hardware ao nível de cliente. Por deslocar o processamento da comparação de *hashes* dos servidores das empresas provedoras para os aparelhos dos usuários finais, questões como a capacidade de armazenamento e processamento desses dispositivos, da opção do usuário por não atualizar seu sistema operacional ou mesmo da obsolescência do hardware impossibilitar essa atualização, entre outras, tornam-se empecilhos para o uso da VPLC de maneira verificavelmente ampla por autoridades públicas.

Quanto à **eficácia**, observa-se que o uso de técnicas de *hash* perceptivo – propostas na vasta maioria das soluções de VPLC até o momento – necessitam de uma base de *hashes* correspondentes ao conteúdo ilegal que se quer identificar para que possam funcionar, visto que dependem de uma comparação entre o material compartilhado pelos usuários e essa base original de *hashes* ilícitos. A composição dessas bases de *hashes* depende de denúncias iniciais e da subsequente constatação da ilicitude do conteúdo veiculado. Isto, por sua vez, gera preocupações em decorrência da constatação de que conteúdo de abuso sexual infantil (CSAM) – alvo de parcela significativa das soluções propostas de VPLC – é, na vasta maioria das vezes (84%), denunciado uma única vez.

Ainda no quesito da eficácia, aponta-se que a definição da sensibilidade das técnicas de VPLC a alterações no conteúdo analisado geram preocupações relevantes independentemente do grau de sensibilidade atribuído ao algoritmo. Níveis mais altos de sensibilidade fazem com que singelas alterações no conteúdo veiculado resultem na atribuição de *hashes* diferentes a conteúdos essencialmente idênticos – mas que foram alterados através de recortes, ajustes de saturação, cor, entre outros. Concomitantemente, níveis mais baixos de sensibilidade do algoritmo possibilitam a adulteração de conteúdos inócuos para atribuir a eles *hashes* idênticos aos de conteúdos marcados como ilícitos, possibilitando a ativação de falsos positivos nos sistemas de comparação desses *hashes*, em especial por agentes mal-intencionados.

Já no quesito **segurança**, constata-se que as bases de *hashes* ilícitos podem ser facilmente adulteradas por agentes mal-intencionados, representando assim uma ampliação na superfície de ataque de sistemas criptográficos. Isso, por sua vez, adiciona ao algoritmo criptográfico riscos similares aos relacionados à inserção de mecanismos de acesso excepcional (*backdoors*) nesses sistemas, que podem ser usurpados por terceiros não autorizados para obter acesso a todo o sistema.

Por fim, quanto ao **escopo** das técnicas de VPLC, evidencia-se a possibilidade de abuso dessas ferramentas por parte de autoridades públicas ou mesmo das próprias plataformas que as administram, a fim de identificar e repreender instâncias de compartilhamento de conteúdo por motivos ideológicos, políticos, socioculturais, entre outros. A possibilidade de ampliação do escopo do conteúdo rastreado por VPLC representa um risco significativo – em especial para comunidades e populações marginalizadas e perseguidas –, o que se opõe diametralmente às expectativas de segurança da informação e liberdade de expressão que se busca proteger através do uso de algoritmos criptográficos em um primeiro momento.

A segunda dimensão notória analisada diz respeito aos **aspectos jurídicos** nos quais as técnicas de VPLC implicam. Os apontamentos aqui trazidos sobre o tema resumem-se em repercussões para: a privacidade e o sigilo das comunicações, a presunção de inocência, a segurança pública e, por fim, a proporcionalidade e a necessidade.

Quanto à **privacidade** e o **sigilo das comunicações**, constata-se que as garantias de privacidade e sigilo da criptografia de ponta a ponta são violadas em casos em que os resultados da comparação de *hashes* sejam compartilhados com o servidor. Esse compartilhamento, contudo, é necessário para que seja possível uma verificação humana do material apontado como ilícito pelo algoritmo, para evitar a penalização de falsos positivos.

No tocante à **presunção de inocência**, ressalta-se que o uso de técnicas de VPLC representa uma quebra com essa prerrogativa constitucional e processual. Isto porque a ferramenta é aplicada a todos os usuários de uma determinada plataforma, mal-intencionados ou não, o que resulta na filtragem de todo o conteúdo compartilhado por padrão.

Por fim, ao que se refere à **proporcionalidade** e à **necessidade**, observa-se, por todo o exposto, que técnicas de VPLC representam um risco desproporcional em comparação com os benefícios obtidos. Adicionalmente, esse risco mostra-se desnecessário em relação ao objetivo almejado, tendo em vista todas as barreiras tecnológicas apontadas ao longo deste trabalho, que tornam a VPLC uma ferramenta pouco eficaz para o combate aos ilícitos que se pretende repreender através dessa técnica.

Finalmente, cabe evidenciar preocupações adicionais que decorrem do uso de VPLC. Uma delas diz respeito à dimensão **sociológica**: as técnicas aqui analisadas representam mecanismos de combate à disseminação de conteúdo ilícito, mas não eliminam a fonte criadora de materiais dessa natureza. A outra diz respeito aos **prejuízos econômicos** que podem ser causados em decorrência da obrigação legal de filtragem massiva de conteúdos por parte das plataformas, o que poderia resultar na impossibilidade de plataformas de pequeno porte atuarem nesse mercado e, assim, resultar em uma concentração de mercado ainda mais intensa por grandes provedores de aplicação.

A conclusão desta análise aponta que o uso de técnicas de VPLC para o combate à disseminação de conteúdos ilícitos em ambientes criptografados – como CSAM, material relativo a terrorismo, entre outros – mostra-se uma medida inadequada por diversos motivos. Não apenas as soluções de VPLC descritas até o momento apresentam diversas falhas e brechas de um ponto de vista tecnológico, como também representam um enfraquecimento de diversas garantias jurídicas consagradas como direitos fundamentais – tais quais o direito à privacidade, à liberdade de expressão, à presunção de inocência, entre outros. Nesse sentido, a VPLC configura-se como uma ferramenta potencialmente tão danosa quanto à própria quebra da criptografia segura através da inserção de mecanismos de acesso excepcional (*backdoors*) em algoritmos criptográficos.

Conquanto a metodologia adotada ofereça sistematização e consistência aos argumentos analisados, notou-se que a limitação a obras de caráter acadêmico não consegue abarcar os debates mais intensos, no calor do momento, que acontecem inevitavelmente fora do universo de artigos científicos e estudos investigativos. Esse ponto cego inerente acaba exigindo fontes adicionais sobre a cronologia dos acontecimentos, reações pela imprensa e eventuais pronunciamentos oficiais.

Ao longo do presente estudo, foi realizada uma revisão bibliográfica extensiva sobre os debates que permeiam a proposta de varredura pelo lado do cliente, bem como suas repercussões sociais, jurídicas e políticas. Espera-se que a análise realizada possa ser utilizada como base para o aprofundamento das discussões em trabalhos futuros. Pode-se indagar sobre a implementação exclusiva em favor do interesse do cliente, considerando a distinção a ambientes adversariais, quando a ferramenta deveria operar contra o interesse do proprietário do dispositivo. Outras propostas concretas de VPLC podem ser consideradas, incluindo algum eventual novo anúncio da Apple, desde que haja compromisso com medidas que cuidem das questões apontadas em abordagem ampla, por exemplo, por meio de sistemas com código fonte integralmente aberto.

NOTAS

1 O primeiro relatório, sobre o panorama da rastreabilidade de mensagens instantâneas, foi publicado em 18 de maio deste ano (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, maio de 2022. Disponível em: https://bit.ly/3yLlb0P. Acesso em: 30 ago 2022).

2 Uma discussão em detalhes sobre os diversos aspectos pertinentes e as percepções quanto às propostas de inserção de mecanismos de acesso excepcional em ambientes com criptografia foi objeto de um estudo prévio conduzido pelo 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. Disponível em: https://bit.ly/3kGTde3. Acesso em: 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 jan. 2022. Disponível em: https://irisbh.com.br/portas-clandestinas-uma-traducao-mais-precisa- para-debatermos-backdoors-em-criptografia/. Acesso em: 10 out. 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. Acesso em: 28 abr. 2022.

6 O termo infantil aqui se refere a pessoas com menos de 18 anos de idade. Desde 1988 o Brasil abandonou o "sistema menorista" da doutrina da situação irregular – expressa no Código Mello Matos (Decreto nº 17.943-A/1927) e no Código de Menores (Lei nº 6.697/1979) – e adotou o paradigma da proteção integral de crianças e adolescentes, pelo qual integrantes desse segmento social são reconhecidos como sujeitos de direito em situação peculiar de desenvolvimento, destinatários de proteção e assistência pelo Estado, pela família e pela sociedade, com prioridade absoluta a suas garantias e direitos básicos (artigo 227 da Constituição da República). O ECA (Lei nº 8.069/1989) nunca usa o termo "menor", preconizando a expressão "crianças e adolescentes", a fim de não reforçar a lógica perniciosa e excludente da perspectiva anterior (BRANCHER, 2000: 126).

7 Depois das primeiras reações negativas ao anúncio, o número foi informado pelo Vice-Presidente de engenharia de software Craig Federighi em entrevista ao 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. Disponível em https://www.wsj.com/articles/apple-executive-defends-tools-to-fight-child-porn-acknowledges-privacy-backlash-11628859600. Acesso em 13 out. 2022) e só então constou de um 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. Disponível em https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acesso em 13 out. 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" ("A segunda proteção é a revisão humana: não há denúncia automática no sistema da Apple. Todas as correspondências positivas devem ser confirmadas visualmente pela Appla como contendo CSAM antes de a Apple desativar a conta e enviar uma denúncia à organização de proteção infantil", em tradução 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. Disponível em https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acesso em 13 out. 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. Disponível em: http://revista.ibict.br/fiinf/article/view/4835. Acesso em: 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 O Google Acadêmico, ou Google Scholar, é acessível no endereço 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. Acesso em 19/05/2022.

14 Por meio da API Content Safety, a Google identifica novas imagens de CSAM por

meio de classificadores com inteligência 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}**. Bruxelas, 11 mai. 2022. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acesso em 13 out. 2022).

15 Ferramenta mais difundida do seu tipo, o PhotoDNA tem duas etapas: detecção e criação do hash. Primeiro, ele identifica imagens acima de determinado tamanho e, ignorando o texto, analisa se ela é conhecida. Segundo ele converte a imagem original para uma versão em escala de cinza de baixa resolução, aplica um filtro de alta-frequência e divide em quadrantes do qual são extraídas medidas estatísticas que geram o hash, uma "assinatura" que permite reconhecer imagens similares submetidas ao mesmo processo, mas não permite que se obtenha a imagem original em regresso a partir do 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}**. Bruxelas, 11 mai. 2022. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acesso em 13 out. 2022).

16 O Safer é uma solução modular que identifica, remove e denuncia imagens de CSAM. Ele opera contra CSAM conhecido, por modelo de correspondência, ou CSAM potencialmente novo e não denunciado, por modelo preditivo de classificador – uma tecnologia de aprendizado de máquina treinada pela empresa Thorn com centenas de milhares de imagens (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}**. Bruxelas, 11 mai. 2022. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acesso em 13 out. 2022).

17 O PDQ da Meta opera por um algoritmo de hash perceptivo de funcionamento semelhante ao PhotoDNA, também 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}**. Bruxelas, 11 mai. 2022. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acesso em 13 out. 2022).

18 MAYER, Jonathan. **Content moderation for end-to-end encrypted messaging.** Princeton University, 2019. P. 42. Disponível em http://cyberlaw.stanford.edu/

publications/content-moderation-end-end-encrypted-messaging. Acesso em 13 out. 2022.

19 Pode operar por a) enclaves seguros no servidor da plataforma; b) correspondência única de terceiros; ou c. Correspondência de vários terceiros. As explicações de cada técnica refogem ao objetivo deste estudo, e podem ser encontradas em EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications**: draft document, September 2020. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acesso em 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. Disponível em https://digitallibrary.un.org/record/1638475. Acesso em 13 out. 2022.

21 O rascunho "Technical solutions to detect child sexual abuse in end-to-end encrypted communications" ("Soluções Técnicas para Detectar Abuso Sexual de Crianças em Comunicações Criptografadas de Ponta a Ponta", em tradução literal) (EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document**, September 2020. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acesso em 13 out. 2022) foi posteriormente incorporado como parte do "Annex 9: Encryption and the fight against child sexual abuse" (Anexo 9: Criptografia e luta contra o abuso sexual infantil) em uma proposta pública da Comissão Europeia para a regulação, no âmbito da UE, da prevenção e combate ao 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. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acesso em 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. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acesso em 13 out. 2022.

23 GLOBAL ENCRYPTION COALITION. **Breaking Encryption Myths: What the European Commission's leaked report got wrong about online security**. 19 nov. 2020. Disponível em https://www.globalencryption.org/2020/11/breaking-encryption-myths/. Acesso em 13 out. 2022; e EDRI - EUROPEAN DIGITAL RIGHTS. **Is surveilling children really protecting them? Our concerns on the interim CSAM regulation**. 24 set. 2020.

Disponível em https://edri.org/our-work/is-surveilling-children-really-protecting-them-our-concerns-on-the-interim-csam-regulation/. Acesso em 13 out. 2022.

24 Constou na minuta uma ressalva: *"Este artigo visa oferecer uma primeira avaliação técnica para auxiliar a identificar possíveis soluções. Trabalho adicional substantivo, para além do escopo deste artigo, provavelmente seria necessário para avaliação, desenvolvimento e aplicação futuras de soluções técnicas através da infraestrutura das empresas"* (EU, 2020: 2). A linha foi mantida na versão pública da proposta: *"Este documento visa mapear possíveis soluções que possam garantir a privacidade das comunicações eletrônicas (incluindo a privacidade das crianças) e a proteção das crianças contra o abuso e a exploração sexual. As soluções exploradas são de natureza puramente técnica, e* ***este artigo não toma uma posição sobre o 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. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acesso em 13 out. 2022. – grifo nosso).

25 POHL, Hartmut. Against surveillance of digital communications in Europe. **Gesellschaft Für Informatik**, 8 nov. 2021. Disponível em https://gi.de/meldung/against-surveillance-of-digital-communications-in-europe. Acesso em 13 ou. 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)**. Genebra: UN, 13 July 2018. 18 p. Disponível em https://digitallibrary.un.org/record/1638475. Acesso em 13 out. 2022. P. 8.

27 A apresentação simultânea das três ferramentas, dificultando a distinção entre os novos recursos, foi criticada como uma possível causa de problemas de compreensão pelos usuários, quanto a sua segurança e privacidade. O chefe de privacidade da Apple, Erik Neuenschwander, respondeu se tratar de um conjunto de recursos que atuariam simultaneamente. 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. Disponível em: https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acesso em: 20 set. 2022.

28 O documento em PDF "Expanded Protections for Children — Technology Summary", atualmente disponível no site da empresa (APPLE. **Expanded Protections for Children — Technology Summary**. Agosto de 2021. Disponível em https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf. Acesso em 13 out. 2022), é uma versão levemente ampliada do texto originalmente publicado em .html no endereço https://www.apple.com/child-safety/, conforme registro no Internet Archive (https://web.archive.org/web/20210805191220/https://www.apple.com/child-

safety/). A empresa detalhou ainda as ferramentas "Segurança das comunicações em mensagens" e "Detecção de CSAM" no 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. Disponível em https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acesso em 13 out. 2022.).

29 APPLE. **Expanded Protections for Children — Technology Summary**. Agosto de 2021. Disponível em https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf. Acesso em 13 out. 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. Disponível em https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acesso em 13 out. 2022. p. 2-4.

30 Luiza Brandão listou, por exemplo, críticas da International Association of Privacy Professionals – IAPP, da Global Encryption Coalition, da Internet Society – ISOC e da 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. Disponível em: https://irisbh.com.br/apple-e-o-mito-privacidade-x-seguranca/. Acesso em: 19 set. 2022). Também merece destaque a crítica da empresa WhatsApp (PROVENZANO, Brianna. WhatsApp acusa Apple de criar sistema de vigilância ao verificar abuso infantil em fotos. Uol - Giz_br, [S. l.], p. -, 8 ago. 2021. Disponível em: https://gizmodo.uol.com.br/whatsapp-apple-sistema-abuso-infantil-vigilancia/. Acesso em: 20 set. 2022).

31 No dia seguinte, uma carta aberta compilou críticas e pediu a suspensão imediata das mudanças propostas e a renovação do compromisso da Apple com a criptografia de ponta a ponta e a privacidade: até meados de setembro de 2022, a carta foi assinada por mais de 30 organizações e 8.700 indivíduos (**AN OPEN LETTER AGAINST APPLE'S PRIVACY-INVASIVE CONTENT SCANNING TECHNOLOGY**. 6 ago. 2021. Disponível em: https://appleprivacyletter.com/. Acesso em: 20 set. 2022).

A ONG estadunidense Electronic Frontier Foundation (EFF) organizou uma campanha denominada "Diga à Apple: Não Escaneie Nossos Telefones" ("Tell Apple: Don't Scan Our Phones"). Ver ELECTRONIC FRONTIER FOUNDATION – EFF. **Tell Apple: Don't Scan Our Phones. Action Center.** 1 set. 2021. Disponível em: https://act.eff.org/action/tell-apple-don-t-scan-our-phones. Acesso em 19/09/2022.

Também nos EUA, a ong CDT classificou a mudança como uma ameaça à segurança e à privacidade. Ver CENTER FOR DEMOCRACY AND TECHNOLOGY – CDT. **CDT: Apple's Changes to Messaging and Photo Services Threaten Users' Security and Privacy.** 5 ago. 2021. Disponível em https://cdt.org/press/cdt-apples-changes-to-messaging-and-

photo-services-threaten-users-security-and-privacy/. Acesso em 19/09/2022.

E em 19 de agosto, noventa organizações da sociedade civil de vários países de todos os continentes (incluindo a EFF, a CDT e este Instituto de Referência em Internet e Sociedade) assinaram uma carta aberta instando a Apple a "abandonar os planos (...) de construir recursos de vigilância em iPhones, iPads e outros produtos". 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/. Acesso em 13 out. 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 A página https://www.apple.com/child-safety tem o 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. Disponível em https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acesso em 13 out. 2022. CENTER FOR DEMOCRACY & TECHNOLOGY. **Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta**. Tradução: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acesso em 13 out. 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. Disponível em https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acesso em 13 out. 2022.; MAYER, Jonathan. **Content moderation for end-to-end encrypted messaging**. Princeton University, 2019. P. 42. Disponível em http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Disponível em: https://bit.ly/3kGTde3. Acesso em: 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. Acesso em 13 out. 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. Acesso em 13 out. 2022. p. 1.

39 APPLE. **Expanded Protections for Children: frequently Asked Questions. v1.1.** Agosto de 2021. Disponível em: https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Frequently_Asked_Questions.pdf. Acesso em 13 out. 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**. Tradução: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em 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://digitalcommons.wcl.american.edu/research/58. Acesso em 13 out. 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. Disponível em https://citizenlab.ca/2020/05/we-chat-they-watch/. Acesso em 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. Disponível em 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. Acesso em 13 out. 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. Disponível em 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. Disponível em: https://bit.ly/3kGTde3. Acesso em: 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/. Acesso em 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. Disponível em https://citizenlab.ca/2020/05/we-chat-they-watch/. Acesso em 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. Disponível em https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acesso em 13 out. 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/. Acesso em 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. Disponível em: https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acesso em: 20 set. 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. Acesso em 13 out. 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. Disponível em https://citizenlab.ca/2020/05/we-chat-they-watch/. Acesso em 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. Disponível em 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. Disponível em https://cdt.org/insights/breaking-encryption-myths-what-the-european-commissions-leaked-report-got-wrong-about-online-security/. Acesso em 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. Disponível em 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. Disponível em 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/. Acesso em: 20 set. 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. Disponível em https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acesso em 13 out. 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. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acesso em 13 out. 2022. p. 7.

61 LUMNIOTIS, 2021: 27)

62 Sobre possíveis impactos econômicos negativos da previsão legal de mecanismos de enfraquecimento da criptografia, ver BARKER, George. LEHR, William. LONEY, Mark. SICKER, Douglas. **O Impacto Econômico das Leis que Enfraquecem a Criptografia**. Law & Economics Consulting Associates (LECA). Tradução de Paulo Rená da Silva Santarém. Reston, VA: Internet Society, 2021. Disponível em https://www.isoc.org.br/files/The-Economic-Impact-of-Laws-the-Weaken-Encryption-PT.pdf. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Acesso em 13 out. 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**. Tradução: SANTARÉM, Paulo Rená da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acesso em 13 out. 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. Acesso em 13 out. 2022. p. 1.

67 Em sistemas processamento de bancos de dados, um protocolo de 2PC ou confirmação de duas-fases (two-phase commit) opera um conjunto de mudanças em um sistema distribuído: os resultados pretendidos, em bloco, são efetivados ou abortados na segunda etapa de acordo com as respostas, respectivamente, positivas ou negativas dos participantes na primeira etapa de verificação preparatória.

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/. Acesso em 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. Disponível em https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acesso em 13 out. 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. Disponível em https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360. Acesso em 13 out. 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**. Tradução: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acesso em 13 out. 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. Disponível em https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf. Acesso em 13 out. 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/. Acesso em 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. Acesso em 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/. Acesso em 13 out. 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. Disponível em https://doi.org/10.48550/arXiv.2111.06628. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acesso em 13 out. 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. Acesso em 13 out. 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. Disponível em https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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**. Tradução: SANTARÉM, Paulo Rená da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acesso em 13 out. 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. Disponível em https://doi.org/10.48550/arXiv.2111.06628. Acesso em 13 out. 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Acesso em 13 out. 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. Disponível em https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 19 abr. 2022. p. 37.

83 "Aplicativos de mensagens criptografadas podem usá-lo para ajudar a alertar os usuários sobre conteúdo malicioso, com privacidade significativamente melhor do que as abordagens que enviam dados em texto simples para servidores de terceiros. Em outra configuração, plataformas de mídia social que atualmente consultam dados de texto simples de seus usuários a serviços de terceiros para ajudar a identificar abusos podem usar nossas técnicas para melhorar a privacidade de seus usuários" (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. Acesso em 13 out. 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. Acesso em 13 out. 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. Disponível em https://doi.org/10.48550/arXiv.2111.06628. Acesso em 13 out. 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. Disponível em http://nujslawreview.org/2021/07/09/the-ministry-and-the-trace-subverting-end-to-end-encryption/. Acesso em 13 out. 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. Acesso em 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. Acesso em 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. Acesso em 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. Acesso em 13 out. 2022.

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

92 ROSENZWEIG, Paul. T**he 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. Acesso em 13 out. 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. Acesso em 13 out. 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Acesso em 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. Disponível

em https://doi.org/10.48550/arXiv.2111.06628. Acesso em 13 out. 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. Disponível em https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acesso em 13 out. 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. Disponível em https://citizenlab.ca/2020/05/we-chat-they-watch/. Acesso em 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. Acesso em 13 out. 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Acesso em 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. Disponível em 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. Disponível em: https://bit.ly/3kGTde3. Acesso em: 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. Disponível em https://doi.org/10.48550/arXiv.2111.06628. Acesso em 13 out. 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**. Tradução: SANTARÉM, Paulo Rená da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acesso em 13 out. 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. Disponível em 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: P**roceeding of the 30th USENIX Security Symposium, August 11-13, 2021**. p. 893-910. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Disponível em 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. Acesso em 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. Disponível em 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. Acesso em 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. Disponível em https://bit.ly/3kGTde3. Acesso em: 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Acesso em 19/05/2022. p. 35

118 O primeiro relatório, sobre o panorama da rastreabilidade de mensagens instantâneas, foi publicado em 18 de maio deste ano (RODRIGUES, Gustavo Ramos; SANTARÉM, Paulo Rená da Silva; VIEIRA, Victor Barbieri Rodrigues. **Comunicações privadas, investigações e direitos: rastreabilidade de mensagens instantâneas**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, maio de 2022. Disponível em: https://bit.ly/3yLlb0P. Acesso em: 30 ago 2022).

Referências

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. Acesso em 19/05/2022.

AN OPEN LETTER AGAINST APPLE'S PRIVACY-INVASIVE CONTENT SCANNING TECHNOLOGY. 6 ago. 2021. Disponível em: https://appleprivacyletter.com/. Acesso em: 20 set. 2022.

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

APPLE. **Expanded Protections for Children — Technology Summary.** Agosto de 2021. Disponível em https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf. Acesso em 13 out. 2022.

APPLE. **Expanded Protections for Children: frequently Asked Questions. v1.1.** Agosto de 2021. Disponível em: https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Frequently_Asked_Questions.pdf. Acesso em 13 out. 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. Disponível em https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf. Acesso em 13 out. 2022.

BARKER, George. LEHR, William. LONEY, Mark. SICKER, Douglas. **O Impacto Econômico das Leis que Enfraquecem a Criptografia.** Law & Economics Consulting Associates (LECA). Tradução de Paulo Renáda Silva Santarém. Reston, VA: Internet Society, 2021. Disponível em https://www.isoc.org.br/files/The-Economic-Impact-of-Laws-the-Weaken-Encryption-PT.pdf. Acesso em 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. Disponível em https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf. Acesso em 13 out. 2022.

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

Messaging and Photo Services Threaten Users' Security and Privacy. 5 ago. 2021. Disponível em https://cdt.org/press/cdt-apples-changes-to-messaging-and-photo-services-threaten-users-security-and-privacy/. Acesso em 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. Disponível em https://cdt.org/insights/breaking-encryption-myths-what-the-european-commissions-leaked-report-got-wrong-about-online-security/. Acesso em 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.** Tradução: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 fev. 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/. Acesso em 13 out. 2022.

COHN, Cindy. **Delays Aren't Good Enough—Apple Must Abandon Its Surveillance Plans.** Electronic Frontier Foundation. 3 set. 2021. 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.

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/. Acesso em 05 jul. 2022.

EDRI - EUROPEAN DIGITAL RIGHTS. **Is surveilling children really protecting them? Our concerns on the interim CSAM regulation.** 24 set. 2020. Disponível em https://edri.org/our-work/is-surveilling-children-really-protecting-them-our-concerns-on-the-interim-csam-regulation/. Acesso em 13 out. 2022.

ELECTRONIC FRONTIER FOUNDATION - EFF. **Tell Apple: Don't Scan Our Phones. Action Center.** 1 set. 2021. Disponível em https://act.eff.org/action/tell-apple-don-t-scan-our-phones. Acesso em 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}**. Bruxelas, 11 mai. 2022. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022SC0209. Acesso em 13 out. 2022.

EU - EUROPEAN UNION. European Commission. **Technical solutions to detect child sexual abuse in end-to-end encrypted communications: draft document**, September 2020. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf. Acesso em 13 out. 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/. Acesso em 13 out. 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, set. 2019. P. 58. 7 - 73. Disponível em http://revista.ibict.br/fiinf/article/view/4835. Acesso em: 10 jun. 2021.

GLOBAL ENCRYPTION COALITION. **Breaking Encryption Myths: What the European Commission's leaked report got wrong about online security.** 19 nov. 2020. Disponível em https://www.globalencryption.org/2020/11/breaking-encryption-myths/. Acesso em 13 out. 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/. Acesso em 13 out. 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. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity22/presentation/jain. Acesso em 13 out. 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. Disponível em https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html. Acesso em 13 out. 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. Disponível em https://citizenlab.ca/2020/05/we-chat-they-watch/. Acesso em 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. Acesso em 13 out. 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. Disponível em https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha. Acesso em 13 out. 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. Acesso em 13 out. 2022.

MAYER, Jonathan. **Content moderation for end-to-end encrypted messaging.** Princeton University, 2019. Disponível em http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging. Acesso em 13 out. 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. Disponível em https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360. Acesso em 13 out. 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.** Genebra: ONU, 4 ago. 2022. Disponível em https://digitallibrary.un.org/record/3985679?ln=en. Acesso em 13 out. 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. Disponível em: https://techcrunch.com/2021/08/10/interview-apples-head-of-privacy-details-child-abuse-detection-and-messages-safety-features/. Acesso em: 20 set. 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. Disponível em: https://bit.ly/3kGTde3. Acesso em 13 out. 2022.

POHL, Hartmut. Against surveillance of digital communications in Europe. **Gesellschaft Für Informatik**, 8 nov. 2021. Disponível em https://gi.de/meldung/against-surveillance-of-digital-communications-in-europe. Acesso em 13 ou. 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. Disponível em https://misinforeview.hks.harvard.edu/article/can-whatsapp-benefit-from-debunked-fact-checked-stories-to-reduce-misinformation/. Acesso em 13 out. 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. Acesso em: 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, maio de 2022. Disponível em: https://bit.ly/3yLlb0P. Acesso em: 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. Acesso em 13 out. 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 jan. 2022. Disponível em: https://irisbh.com.br/portas-clandestinas-uma-traducao-mais-precisa-para-debatermos-backdoors-em-criptografia/. Acesso em 13 out. 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. Disponível em https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acesso em 13 out. 2022.

STERN, Joanna; HIGGINS, Tim. Apple Executive Defends Tools to Fight Child Porn, Acknowledges Privacy Backlash. **The Wall Street Journal.** 13. aug. 2021. Disponível em https://www.wsj.com/articles/apple-executive-defends-tools-to-fight-child-porn-acknowledges-privacy-backlash-11628859600. Acesso em 13 out. 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. Disponível em https://doi.org/10.48550/arXiv.2111.06628. Acesso em 13 out. 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)**. Genebra: UN, 13 July 2018. 18 p. Disponível em https://digitallibrary.un.org/record/1638475. Acesso em 13 out. 2022.

Apêndice 1 - Corpus total de textos analisados

REFERÊNCIA	CATEGORIA	FONTE
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. Acesso em 19/05/2022.	Relatório	Google Acadêmico
APPLE. Expanded Protections for Children, August 2021. Disponível em https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Technology_Summary.pdf	Relatório	Referência 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. Disponível em https://www.apple.com/child-safety/pdf/Security_Threat_Model_Review_of_Apple_Child_Safety_Features.pdf.	Relatório	Referência 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. Disponível em https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b6555a1018a750f39028005bfdb9f35eaee4b947.pdf.	Trabalho publicado em anais de evento acadêmico	Referência de ABELSON et al., 2021.

REFERÊNCIA	CATEGORIA	FONTE
CENTER FOR DEMOCRACY & TECHNOLOGY. Olhando de fora para dentro: abordagens para moderação de conteúdo em sistemas com criptografia de ponta a ponta. Tradução: SANTARÉM, Paulo Renã da Silva. VIEIRA, Victor Barbieri Rodrigues. Instituto de Referência em Internet e Sociedade. 15 de fevereiro de 2022. Disponível em https://irisbh.com.br/publicacoes/abordagens-para-a-moderacao-de-conteudo-em-sistemas-com-criptografia-de-ponta-a-ponta/	Relatório	Inserção discricionária
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/. Acesso mais recente em 05 jul. 2022.	Relatório	Referência 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. Disponível em https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf.	Relatório	Referência de ABELSON et al., 2021.

REFERÊNCIA	CATEGORIA	FONTE
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/	Artigo científico	Google Acadêmico
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	Artigo científico	Google Acadêmico
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. Disponível em https://www.usenix.org/conference/usenixsecurity22/presentation/jain.	Trabalho publicado em anais de evento acadêmico	Google Acadêmico

REFERÊNCIA	CATEGORIA	FONTE
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. Disponível em https://www.unicef-irc.org/publications/1152-encryption-privacy-and-childrens-right-to-protection-from-harm.html.	Relatório	Google Acadêmico
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. Disponível em https://citizenlab.ca/2020/05/we-chat-they-watch/. Acesso em 03 jul. 2022.	Relatório	Referência 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.	Artigo científico	Google Acadêmico

REFERÊNCIA	CATEGORIA	FONTE
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.	Trabalho publicado em anais de evento acadêmico	Google Acadêmico
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.	Artigo científico	Google Acadêmico
MAYER, Jonathan. Content moderation for end-to-end encrypted messaging. Princeton University, 2019. http://cyberlaw.stanford.edu/publications/content-moderation-end-end-encrypted-messaging	Relatório	Referência 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. Disponível em https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659360	Relatório	Referência de ABELSON et al., 2021.

REFERÊNCIA	CATEGORIA	FONTE
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. Disponível em: https://bit.ly/3kGTde3.	Relatório	Google Acadêmico
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	Artigo científico	Referência 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.	Artigo científico	Google Acadêmico

REFERÊNCIA	CATEGORIA	FONTE
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/	Relatório	Referência 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. Disponível em https://doi.org/10.48550/arXiv.2111.06628.	Artigo científico	Google Acadêmico

Apêndice 2
Formulário de análise

- E-mail
- Ano
- Referência ABNT
- Link da publicação
- Categoria

marcar apenas uma opção

Artigo científico
Declaração, carta aberta
Relatório
Nota técnica
Jurisprudência
Trabalho publicado em anais de evento acadêmico
Monografia, dissertação ou tese
Artigo de opinião
Matéria de jornal
Post de blog

- Escopo

marcar apenas uma opção

Rastreabilidade
Hacking governamental
Varredura pelo lado do cliente

- Síntese

Texto de resumo elaborado pela equipe do IRIS: deve incluir uma apresentação breve da proposta do trabalho, metodologia (ou ausência de indicação de metodologia), eventuais referências relevantes (citadas como base para o conceito ou posicionamento indicado no trabalho) e abordagem dada ao escopo analisado.

- Comentários
- Citações
- Observações

iris

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE