Comments on Client-Side Scanning as a CSAM Detection Mechanism

The Messaging, Malware, and Mobile Anti-Abuse Working Group (M[3]AAWG, https://www.m3aawg.org/) appreciates the opportunity to submit comments on client-side scanning as a CSAM detection mechanism. M[3]AAWG is a technology-neutral global industry association. As a working body, we focus on operational issues of internet abuse, including technology, industry collaboration, and public policy. With more than 200 institutional members worldwide, we bring together online community stakeholders and develop best practices and cooperative approaches for fighting online abuse.

General Comments

M[3]AAWG vehemently supports efforts to fight Child Sexual Abuse Material (CSAM). We recognize the devastating harm it causes and work to equip the industry with effective tools and frameworks to combat CSAM, including best common practices documents such as the M[3]AAWG Disposition of Child Sexual Abuse Materials BCP Revised Aug 2021. Our position is grounded in the belief that tackling CSAM requires solutions that are both operationally effective and respectful of fundamental rights, ensuring that interventions do not create new avenues for abuse or significantly deleterious side effects.

The European Union's proposed Child Sexual Abuse Regulation (CSAR) includes provisions for mandatory client-side scanning. As security defenders, we believe these provisions, although well-intended, are seriously problematic.

First, such proposals in practice undermine the confidentiality of end-to-end encrypted (E2EE) communications, a cornerstone of technical security. History has shown that weakening encryption, whether through intentional backdoors or scanning before encryption, compromises the privacy and security of all users. Once such vulnerabilities are introduced, they inevitably become attractive targets for malicious actors, exposing individuals and organizations to increased risks of cybercrime, espionage, and exploitation. Strong encryption safeguards not only personal messages but also critical infrastructure, healthcare records, financial transactions, and the global economy.

Second, the use of closed-source or proprietary subsystems (such as so-called "perceptual hashing" of images), means that industry, network providers/ISPs, academia and society cannot validate either the efficacy nor the error rates inherent in client-side scanning. Calls to modify client devices without transparency must be rejected- our communities rely on the trust and integrity of mobile devices more than ever, given the extent to which people and businesses depend on the integrity of those mobile devices. There also exist many systems that do not use the most widely

used operating systems that are readily available to circumvent any proposed controls - for example the proposed controls could put network operators/ISPs in a position where they would be expected to undertake major, and likely fruitless, efforts to control their customers data flows in ways that would be antithetical for their businesses.

Third, the system designed to handle reports of potential misbehavior is likely to be overwhelmed given the massive scale of interpersonal messaging and the certainty of some level of false positives generated by image analysis mechanisms. This could significantly damage law enforcement, which in many places already struggles to keep up with CSAM reporting.

Fourth, the system will not be able to capture all CSAM. In addition to inevitable false positives, image analysis mechanisms also produce false negatives, in which an analysed image with CSAM is not detected. Such images may easily be generated, found, and circulated, undermining the entire client-side scanning system. Furthermore, as stated in the second point above, efficacy and error rates cannot be reliably measured.

Fifth, image manipulation is often as simple, and in some cases simpler, than image analysis. Once image analysis mechanisms are deployed, it becomes nearly trivial to generate images that appear innocuous to humans but are detected and reported as CSAM. This creates a number of potential attack vectors, including mass generation of invalid reports or victim framing.

Lastly, if client-side scanning is mandated in some jurisdictions, other jurisdictions or private actors will inevitably reuse that infrastructure and include other types of content scanning, creating a ready-made system and template for censorship.

Conclusion

We urge policymakers to seek alternative approaches to address criminal activity directly while preserving robust encryption for law-abiding users. We must ensure that the chosen methods to combat abuse do not inadvertently create systemic weaknesses that could harm billions of innocent users and erode the very protections we seek to strengthen.

We appreciate the opportunity to submit these comments, and we welcome further opportunities to engage as needed to answer any questions during this process. Please address any inquiries to M^{3}AAWG Executive Director Amy Cadagin at comments@m3aawg.org.

Sincerely,

Amy Cadagin
Executive Director
Messaging Malware Mobile Anti-Abuse Working Group (M^{3}AAWG)
comments@m3aawg.org