Briefing: end-to-end encryption and child sexual abuse material

5Rights and Professor Hany Farid

December 2019

Summary

A plan by Facebook, and others, to implement end-to-end encryption across their services, including on Instagram Direct and Facebook Messenger, is set to be a boon for child predators and abusers. Specifically, it could cripple the global effort to disrupt the online distribution of child sexual abuse material (CSAM).

5Rights Foundation is not opposed to end-to-end encryption. We are simply calling on all companies not to implement changes until they can demonstrate that protections for children will be maintained or improved. Regrettably, most companies have failed to make this commitment.

Background

In 2018, tech companies flagged 45 million photos and videos as child sexual abuse material, contained within over 18 million reports to the CyberTipline at the National Center for Missing and Exploited Children (NCMEC). Perhaps more startling than this volume of more than 5,000 images and videos per hour is that last year's reports alone account for 1/3 of all reports received by the CyberTipline since its inception in 1998 -- the global distribution of CSAM is growing exponentially.

Many major technology companies have deployed technology that has proven effective at disrupting the global distribution of known CSAM. This technology, the most prominent example being photoDNA, works by extracting a distinct digital signature (a 'hash') from known CSAM and comparing these signatures against images sent online. Flagged content can then be instantaneously removed and reported.

This type of robust hashing technology is similar to that used to detect other harmful digital content like viruses and malware. Since its development in 2009, and its eventual world-wide deployment, photoDNA remains one of the most effective strategies for combatting child sexual abuse online.

The efficacy of this technology is now under threat.

The Problem

Earlier this year, Facebook's Mark Zuckerberg announced that he is implementing end-to-end encryption on *all* his platforms, not just on Whatsapp which is already end-to-end encrypted. In announcing the decision, Mr. Zuckerberg conceded that it came at a cost.

> *"Encryption is a powerful tool for privacy, but that includes the privacy of people doing bad things. When billions of people use a service to connect, some of them are going to misuse it for truly terrible things like child exploitation, terrorism, and extortion."*

This casual lack of concern fails to address the seriousness or scale of CSAM, not least given that Facebook Messenger accounted for 12million of the 18million reports received by the CyberTipline in 2018. Facebook is by no means the only problem, however. In a November 2019 article entitled *Child Abusers Run Rampant as Tech Companies Look the Other Way,* the New York Times revealed that Apple does not scan for known CSAM on either its encrypted iMessage service or its cloud storage. Dropbox, Google and Microsoft – all extremely well-placed to tackle the distribution of CSAM – only scan for images when they are shared through their services, not when they are uploaded.

Much of the problem exists already, therefore. But the move towards end-to-end encryption has brought the issue into sharp focus. Without intervention, the adoption of end-to-end encryption would completely cripple the efficacy of the decade-long photoDNA program. This is horrific for the millions of young children impacted by CSAM, and completely unnecessary.

The Solution

First and foremost, we are not opposed to end-to-end encryption. While many companies seek to excuse their inaction by positing a false choice between privacy and safety, the reality is that photoDNA is perfectly compatible with end-to-end encryption – it does not involve 'prizing open' an encrypted message or creating a 'backdoor' in the system.

PhotoDNA can be modified to be implemented at the point a message/image is sent, contrary to the current approach where it is implemented 'server-side'. In this 'client-side' implementation, the distinct signature is extracted prior to encryption and transmitted alongside the encrypted message. Because no identifying information can be extracted from this signature, it does not reveal any details about the encrypted image while allowing for the monitoring of known CSAM. Homomorphic technology can also perform image hashing on encrypted data without the need to decrypt the data.

The Ask

Mr. Zuckerberg has repeatedly expressed his desire to "get it right" this time. The technology exists to get it right. We now just need the will to do so.

Any steps towards end-to-end encryption must not put children at greater risk. We ask for:

1. A commitment from all internet companies to scan relevant services for known CSAM, and to implement end-to-end encryption in a way that allows photoDNA and similar tools to operate. This includes existing end-to-end encrypted services like WhatsApp and iMessage.
2. A commitment from advertisers to pull their ad spend from any platform that has not made the above commitment.
3. A commitment from both national governments and international institutions to mandate that end-to-end encryption must allow for the detection and disruption of CSAM.

The Bigger Picture

Keeping successful programs like photoDNA in place should only be the first step in ensuring that children are safe online, including in encrypted environments.

The technology sector should be more aggressive in developing and deploying new technologies to limit the spread of CSAM and protect children. For example, technology exists to expand photoDNA from its current image-based analysis to video. This is essential as we know that much of the most recent CSAM is video-based. Technology also exists to identify new CSAM (e.g. Two Hat Security's CEASE.ai tool) and to scan for text-based grooming activity on both private channels and open networks. It is vital that the move towards end-to-end encryption does not undermine progress in these areas.

About Hany Farid

Professor Hany Farid worked with Microsoft to develop photoDNA before its release and has been responsible since for a number of ground-breaking improvements to the technology. He is a Professor at the University of California, Berkeley with a joint appointment in Electrical Engineering and Computer Science and the School of Information. Farid's research focuses on digital forensics, image analysis, and human perception.

About 5Rights

The digital world was imagined as one in which all users would be equal, yet 1/3rd of internet users are children. Nearly one billion children are therefore growing up in an environment that systematically fails to recognise their age and the protections, privileges, legal frameworks and rights that together constitute the concept of childhood.

Working closely with children, we operate in the engine room of the digital world: supporting enforceable regulation and international agreements, developing technical standards and protocols, and helping businesses re-imagine the design of their digital services.

5Rights fights for a digital environment that anticipates the presence and meets the needs of all children, so they can access it ***knowledgeably, creatively, and fearlessly***.

Jay Harman
Policy Lead, 5Rights
jay@5rightsfoundation.com | 020 7502 3818 | 5rightsfoundation.com

FAQs

What is child sexual abuse material?
CSAM refers to sexualised content in any form that involves a child. This includes both real and simulated acts of child sexual abuse and any representation of the sexual parts of a child for primarily sexual purposes. Such material is documented evidence of a crime because it involves children who cannot consent. The age of most of these victims ranges from only a few months old to 12 years of age.

Can PhotoDNA be repurposed for other types of content?
No. PhotoDNA is only available to trusted online service providers, businesses hosting user-generated content, and select organisations or agencies dedicated to tackling child exploitation. Use of photoDNA is only granted to those who have been through a strict vetting process to secure a license, which legally restricts the use of the technology to the identification of CSAM.

If we make exceptions for CSAM, what's to stop other countries – possibly with repressive regimes – making exceptions for other things?
Per licensing agreements, photoDNA can only be used for the identification of CSAM and not for other purposes (see above).

Doesn't this amount to improper surveillance?
No. PhotoDNA works by creating a distinct digital signature (a 'hash') of an image, which is then compared to a database of hashes of images already identified and confirmed to be CSAM. This privacy-preserving hash is not reversible, and cannot be used to recreate the original image. This is the same type of technology that is routinely used to stop the spread of spam, viruses, and malware. Simply put, photoDNA can only identify content that has already been determined to be CSAM, and is unable to extract any meaningful information from content that is not CSAM.

Are these technologies actually effective at limiting the distribution of CSAM online?
Yes. In 2018 alone, NCMEC's CyberTipline received over 18 million reports of CSAM. At a rate of 2000 per hour, these 18 million reports account for 1/3 of all reports received by the CyberTipline since its inception in 1998. More than 95% of all of these reports are generated by photoDNA.

What is the error rate of photoDNA?
PhotoDNA was designed to have an error rate (mis-identifying non-CSAM as CSAM) of 1 in 50 to 100 billion. This technology has been deployed by Adobe, Dropbox, Facebook, Google, Microsoft, Twitter, and many more for over a decade and has been proven to be highly effective and accurate.

What happens when photoDNA makes a mistake?
If photoDNA incorrectly flags an image, then the account is temporarily frozen until a user-generated appeal is filed and the flagged content is reviewed by a human moderator. Any improperly frozen accounts are then reinstated.

Does photoDNA and similar technologies threaten the privacy of users who wish to use encrypted services.

No. Messages remain fully encrypted and the photoDNA hash is privacy preserving meaning that the hash is not reversible, and cannot be used to recreate the original image. PhotoDNA can only identify content that has already been determined to be CSAM, and is unable to extract any meaningful information from content that is not CSAM.

Can the problem of CSAM be solved with Artificial Intelligence?

No. Despite the remarkable advances of the last few years, even the best AI-based image recognition systems operate at an accuracy of around 90%. With an error rate of 1/10 as compared to photoDNA's 1/100,000,000,000, current AI systems are orders of magnitude away from where they need to be to operate at Internet-scale. In addition, any such AI-based system would be ineffective within a fully encrypted system.

What is the timeline for this being implemented?

Facebook has stated that the implementation of E2E encryption **by default** will not happen on Facebook Messenger and Instagram Direct before 2020, and most likely not before 2021. However, it is already possible for *users* to enable E2E encryption in Facebook Messenger.

What about DNS-over-HTTPS?

The implementation of DNS-over-HTTPS is a related, but different, issue and requires different solutions. The issues are similar, however - without intervention, implementing DNS-over-HTTPS could severely weaken the ability to tackle CSAM online. The IWF, whose 'URL list' would be badly impacted by DNS-over HTTPS, has prepared a useful briefing on the issue, which you can read here.

How are countries around the world responding to the issue of E2E encryption?

Australia, Germany, India and the USA have all proposed or passed legislation/policy pertaining to E2E encryption.

In the USA, this is limited to statements by White House officials that legislation is necessary to stop tech companies using encryption to undermine law enforcement, while in Australia legislation was passed in December 2018. The Telecommunications (Assistance and Access) Act 2018, which requires companies to give law enforcement access to encrypted data if companies are able, or to build capacity to do so if they can't already.

No country has moved to solve the specific issue of continuing to scan for CSAM on private channels.