Breaking encryption myths
What the European Commission's leaked report got wrong about online security

In September 2020, a draft European Commission report called "Technical solutions to detect child sexual abuse in end-to-end encrypted communications" was leaked.[1] The report gave an analysis of different ways to spot illegal content in private communications that use end-to-end encryption.

Encryption is one of our strongest tools to help keep billions of people worldwide secure online. But is there a way to scan personal information without putting the safety and security of all users at risk?

To help separate fact from fiction, a group of expert members of the Global Encryption Coalition analysed the report.

Background

Preventing crime and keeping people secure online is a universal priority. While curbing the spread of child sexual abuse material (CSAM) online is an important goal towards this end, the European Commission's leaked report outlines a handful of so-called "content moderation solutions" that would put all users, including children, at far greater risk of harm. This is because each of the content detection methods would require breaking end-to-end encrypted systems with a form of backdoor access to encrypted communications.

Breaking end-to-end encryption to curb objectionable content online is like trying to solve one problem by creating 1,000 more. Insecure communications make users more vulnerable to the crimes we collectively are trying to prevent.

While the report alludes to the idea that some may be less risky than others[2], each method presents serious security and privacy risks for billions of users worldwide.

Detection methods

The European Commission's leaked report assesses three types of methods to spot prohibited material in end-to-end encrypted communications.

1. **Traditional backdoors:** A communications platform that allows third party access to the content of encrypted communications.[3] For instance, key escrow[4], access to a server where the data is held upon receipt, or a "middle-box" which decrypts the data at a central server and then re-encrypts it for sending to the intended recipient[5]. The report specifically assessed:
    - Non-end-to-end encrypted communications
    - End-to-end encrypted communications with exceptional access

1 https://www.politico.eu/wp-content/uploads/2020/09/SKM_C45820090717470-1_new.pdf
2 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Page 21. Chart.
3 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Pages 3, 4, 6.
4 https://academic.oup.com/cybersecurity/article/1/1/69/2367066
5 https://www.internetsociety.org/blog/2019/11/what-is-a-man-in-the-middle-mitm-attack/

2. **Client-side scanning:** Referred to in the leaked report as "device related," these access methods scan message contents on the user's phone, tablet or mobile.[6] Videos and images, for example, are scanned for matches against a database of prohibited content before, and sometimes after, the message is sent to the recipient.[7] If the message contents match known prohibited content, the full message would be sent to a third party for manual review. The report specifically assessed:
    - All detection done on-device
    - On-device full hashing with matching at server
    - On-device partial hashing with remaining hashing and matching at server
    - On device use of classifiers
3. **Secure enclaves and homomorphic computation:** Referred to in the Draft as "server related" and "encryption related," an encrypted message's content can only be "seen" by the computation in a server or on the user's device[8]. Yet, if the message is found to contain prohibited content it will be forwarded to a third party for manual review. The report assessed:
    - Secure enclaves in the ESP server
    - Single third-party matching
    - Multiple third parties matching
    - On-device homomorphic encryption with server-side hashing and matching

The dangers of backdoor access to encryption

The three types of content moderation methods assessed in the report involve different technical approaches, but they share one crucial thing in common: they put the security of billions of people and nations worldwide at risk. That's because each of them involves creating a form of "backdoor access" to confirm whether material is actually prohibited. The mainstream technical expert consensus[9] is that those backdoors are likely to become available to third party adversaries who will be able to use them.

Manual review puts user safety and privacy at risk

The report repeatedly mentions methods which include a manual review step. However, it doesn't say how a manual review would be performed.[10] Manual review is a major point of weakness that applies to all the methods proposed. False positives are a concern with all automated content moderation,[11] making manual review unavoidable.

- **Manual review doesn't just happen; it requires a backdoor.** When an automated scan detects what looks like prohibited material, it goes for manual review.[12] This means a person (third party) will ultimately have to review the content. However, a manual review would require the creation of a technical feature that allows third party access to the contents of encrypted communications. In other words, they have to build in a backdoor.

6 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Pages 7-13.
7 https://blog.cryptographyengineering.com/2019/12/08/on-client-side-media-scanning/
8 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Pages 13-20.
9 http://dspace.mit.edu/bitstream/handle/1721.1/97690/MIT-CSAIL-TR-2015-026.pdf
10 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Page 23. Paragraph 1.
11 https://www.bbc.com/news/54467384
12 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Figures 1, 4, 5, 6, 7, 8, 9, 10, 11.

Furthermore, the report fails to point out the huge risks that come with manual review. Its approach to threat analysis is like comparing the security of different types of walls on a house while ignoring its unlocked door.

Because each method they assessed involves manual review, it inherently breaks the end-to-end encryption, thereby weakening the security of the system and putting the safety and privacy of its users at greater risk.

Exceptional access creates perpetual vulnerability

The leaked report makes multiple references to "end-to-end encrypted communications with exceptional access." End-to-end encrypted communications cannot have exceptional access; that is a complete misnomer. More critically, the report fails to acknowledge that exceptional access means that communications are not secure.[13] As the report notes, "all the content of communication could in practice be accessed by the electronic service provider at any point using the exceptional access mechanism."[14] That means anyone with access to the mechanism – whether lawful or not – could use it to access the content.[15]

Here are the facts:

- **An encryption backdoor by any other name is still a backdoor.** Sometimes called "exceptional access", an encryption backdoor is a way for a third party to access the contents of communications. This might sound harmless, but the results can be disastrous.
- **Any backdoor is a security nightmare.** Creating "backdoor access" would put children at far greater risk by creating new opportunities for criminals and authoritarian governments to further victimize and exploit vulnerabilities in communications. This is because encryption backdoors can be opened by anyone who finds them—including criminals, terrorist organizations, hostile intelligence, and other adversaries.[16]
- **Encryption backdoors are intentional vulnerabilities built into a system.**[17] They make it easier for other parties, including criminals, foreign governments and other unauthorised parties to get access to data that was supposed to be encrypted.[18] The leaked report, however, makes a misleading risk analysis of the security of these systems, often labelling them as "medium" or "low".[19] These risk assessments are underestimated as there are ample examples of these complex systems being compromised.[20, 21]
- **There is no way to make a door that only the "good guys" can open and the "bad guys" cannot.** Creating a backdoor weakens the security of the whole system and puts all its users at risk.[22]

Client-side scanning and server-related methods as analysed in the report are not alternatives to encryption backdoors. They *are* backdoors.

13 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Page 6. Paragraph 4.
14 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Page 6. Paragraph 5.
15 https://www.justsecurity.org/53316/criminalize-security-criminals-secure/
16 https://uk.reuters.com/article/us-usa-security-congress-insight/spy-agency-ducks-questions-about-back-doors-in-tech-products-idUKKBN27D1CS
17 https://www.lawfareblog.com/if-we-build-it-they-will-break
18 https://docs.zoho.com/file/eykeu5682e11c122046debcc3f0e6e16c4e13
19 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Page 21. Chart.
20 https://www.nbcnews.com/tech/security/spy-agency-ducks-questions-back-doors-tech-products-rcna167
21 https://www.mattblaze.org/papers/eesproto.pdf
22 https://cdt.org/insights/issue-brief-a-backdoor-to-encryption-for-government-surveillance/

Weakening security to detect prohibited material is irresponsible

The report doesn't just propose flawed methods for confirming whether material is actually prohibited. It suggests risky approaches for spotting that content in the first place.

Client-side scanning

Every client-side scanning approach mentioned in the report opens up risk for users[23]. And that's without including those potentially dangerous manual review steps, discussed above.

While client-side scanning methods can sometimes be beneficial in specific voluntary applications by users, such as for performing virus scans on a computer, the mandatory application of client-side scanning to encrypted communications is not. Client-side scanning "increases the "attack surface" for encrypted communications by creating additional ways to interfere with communications by manipulating the database of prohibited content."[24] While prohibited content is scanned on the user device, the database of hashed images can be held either at a central server or also on the user device. As noted in the leaked report, technology and policy constraints inhibit the feasibility of an all-on-device method.[25] The database is more likely to be held in a central server, not a user's device. In the central server model, the hash of each image the user wants to send will be known by the server.[26] Information about the image or video could be exposed in transit, and attackers could use rainbow tables[27] to gain more information about the image or video being hashed.

Like databases of user keys present in a key escrow system[28], the central database in a client-side scanning system would present an attractive target for bad actors. They could, for example, add hashed content to the central database.[29] If the manual review mechanism is also hijacked, attackers could get the power to check it under manual review when it comes up as a match to known content. They could potentially see content before it is encrypted and sent, and they could track exactly what content is being sent, when, where and to whom it is sent. A client-side scanning mechanism could be used by authorities to scan for content beyond the original purpose it was created for, leading to concerns around censorship and mass surveillance.[30]

Secure enclaves and homomorphic computation

Secure enclaves and homomorphic encryption are important innovations in user privacy. They were designed primarily for use in cloud computing to protect the privacy of user data while still allowing computations to be performed on that data.

However, by allowing private user data to be scanned via direct access by servers and their providers, the methods outlined in the report break the privacy expectations of users of end-to-end encrypted communication systems.

A pattern is emerging. All four secure enclaves and homomorphic computation methods the report outlines raise similar concerns as the methods discussed above, such as hashed fingerprints or classifiers in client-side scanning. Even if the hashes are

23 https://www.lawfareblog.com/law-and-policy-client-side-scanning

24 https://www.internetsociety.org/resources/doc/2020/fact-sheet-client-side-scanning/

25 *Technical solutions to detect child sexual abuse in end-to-end encrypted communications.* Page 8. Paragraph 2.

26 https://www.eff.org/deeplinks/2019/11/why-adding-client-side-scanning-breaks-end-end-encryption

27 https://www.lifewire.com/rainbow-tables-your-passwords-worst-nightmare-2487288

28 http://dspace.mit.edu/bitstream/handle/1721.1/97690/MIT-CSAIL-TR-2015-026.pdf

29 https://www.lawfareblog.com/law-and-policy-client-side-scanning

30 http://cyberlaw.stanford.edu/blog/2020/05/client-side-scanning-and-winnie-pooh-redux-plus-some-thoughts-zoom

made with cutting-edge secure computational technology, the problem persists. In the "on-device homomorphic encryption with server-side hashing and matching" method, it's unclear if end-user devices would be able to perform homomorphic computations, potentially exacerbating existing inequalities in user access to powerful devices.

What would breaking encryption mean for the future of online security in the EU?

Our in-depth assessment of the methods outlined in the report highlights that all methods of third-party access to encrypted content break end-to-end encryption. If used, each and every one of the methods described would threaten both the security and privacy of billions of users, including children, online.

The European Commission's report failed to highlight these risks.

Important omissions, like the assessment or explanation of a manual review step, undermine the usefulness of the report. It cannot claim to assess ways to detect prohibited content on end-to-end encrypted communications, when the real risks are not outlined. Similarly, it fails to explore the security impact of these access methods on users. This leaves dangerous vulnerabilities unmentioned, like the potential for bad actors to manipulate databases used in client-side scanning methods.[31, 32]

Conclusion

EU policymakers and lawmakers need to understand the real impact of content moderation methods if they are to make sound decisions to keep citizens safe online. This leaked report fails to outline the serious risks of requiring communications service providers to detect prohibited content. These requirements would force service providers to undermine the security of their end-to-end encrypted services, jeopardizing the safety of the billions of people who rely on them each day. Put simply, there is no way to break encryption without making everyone, including children, more vulnerable.

31 https://www.lawfareblog.com/law-and-policy-client-side-scanning

32 https://blog.cryptographyengineering.com/2019/12/08/on-client-side-media-scanning/

Signatories*

Georgia Bullen, Executive Director, Simply Secure

Jon Callas, Director of Technology Projects, EFF

L. Jean Camp, Indiana University

Richard Compton, Chairman, M3AAWG DDoS SIG

Amanda DeLuke, Deliverability Analyst, Higher Logic LLC

Nuno Guerreiro de Sousa, Technologist, Privacy International

Prof. Sven Dietrich, City University of New York, USA

Peter Goldstein, Chief Technology Officer and Co-Founder, Valimail

Joseph Lorenzo Hall, Senior Vice President for a Strong Internet, Internet Society

J. Alex Halderman, Professor of Computer Science and Engineering; Director, Center for Computer Security and Society, University of Michigan

Fen Osler Hampson, Chancellor's Professor, Carleton University

Dr. Sven Herpig, Director for International Cybersecurity, Stiftung Neue Verantwortung

J. C. Jones, Cryptography Engineer, Insufficient.Coffee

Joseph Kiniry, Principal Scientist, Galois, CEO and Chief Scientist, Free & Fair

Mallory Knodel, Chief Technology Officer, Center for Democracy and Technology

Petri Koistinen, Nitor

Chelsea Holland Komlo, University of Waterloo

Istvan Lam, CEO, Tresorit

Susan Landau, The Fletcher School and Tufts School of Engineering

Sascha Meinrath, Director, X-Lab, Palmer Chair in Telecommunications, Penn State University

Nat Meysenburg, Technologist, New America's Open Technology Institute

Lauren Meyer, VP of Industry Relations & Compliance, Kickbox

Jake Moore, Cybersecurity Specialist, ESET

Jakub Olexa, CEO, Mailkit

Riana Pfefferkorn, Stanford University

Ryan Polk, Senior Policy Advisor, Internet Society

Hannah Quay-de la Vallee, Senior Technologist, Center for Democracy and Technology

Greg Robinson, Privacy and Compliance Director, Higher Logic LLC

Guillaume Séjourné - Product Manager - Vade Secure and member of the M3AAWG Board of Directors

Adam Shostack, author of Threat Modeling: Designing for Security

Christopher Weatherhead, Technology Lead, Privacy International

Mohammed A. Zaman, Deployment Consultant UK, dmarcian Europe

*Affiliations provided for identification purposes only