Statement to the Committee on Digital Affairs on the CSA Regulation (“chat control”)

Elina Eickstädt

Chaos Computer Club

23 February 2023

Preliminary remarks

In principle, the European Commission’s proposed Regulation to combat child sexual abuse pursues an important objective. Without a doubt, better help must be provided to the victims of child abuse, and the dissemination of material depicting child abuse must be prevented. However, the draft Regulation contains measures which would mean the surveillance of all communication content and undermine fundamental principles of confidential, secure digital communication. In its present form, the Regulation would create an unprecedented surveillance infrastructure, one which is neither rights-compliant nor technically feasible. The draft Regulation is unsuited to achieving its stated aim, and in practice it would create new problems for law enforcement, rather than ensuring that serious criminal offences can be tackled in a targeted and effective manner. Technology can only ever be a supportive tool in solving complex societal problems. This statement begins by discussing the technical basis for implementing the detection orders called for by the Regulation. It then highlights the implications for the privacy of all members of the public and the dangers of the age verification proposal. The final section describes how the implementation of access blocking is unrealistic from a technical perspective, and discusses the problematic role of the planned EU Centre. On a more general note, I would like to draw attention to the fact that a legal framework has already been created with various measures to improve the protection of children online, in the form of the Digital Services Act (DSA). Systematically implementing the DSA could solve many of the problems raised by the Commission.

Proposed measures and technical implementation

In essence, the Regulation aims to ensure that the content of all interpersonal communication (social media, chat services, but also hosting and cloud providers) is scanned, without this being dependent on a reasonable suspicion of wrongdoing. In particular, the draft Regulation contains obligations to detect known and unknown material depicting child abuse, and to detect the attempted solicitation of children, known as grooming.

Detection of known material

Known material is detected using what is known as “perceptual hashing”. Various implementations of this technology already exist, such as PhotoDNA, which was developed by Microsoft. Meta uses the PDQ hash function, while Apple uses NeuralHash. The hash of an image can be thought of as a fingerprint. An image is fed into the algorithm, and a hash value specific to the image is computed on the basis of various characteristics. As the hashes are generated on the basis of various aspects of the image, the same hash is produced even if there are small changes, and so the detection of relevant visual material is relatively reliable. Comparing fingerprints is less resource-intensive than comparing an entire image. This approach also means that it is not necessary to keep a central database of CSAM, only of the hashes, and this reduces the potential for misuse.

While this technology can, in principle, be regarded as very reliable, there are two key disadvantages which must be kept in mind. Firstly, this approach requires a database which must be maintained centrally. The database administrator has the power to decide which images are detected and flagged. This can potentially be abused. In addition, the perceptual hashing method is not completely error-free. Innocuous images can be manipulated to have a hash which matches that of an image marked as “illegal”. This enables attackers to have images flagged without having control of the database.[1]

Detection of unknown material

While the Regulation is ostensibly technology-neutral, it is very clear from the impact assessment published with the Regulation that systems based on artificial intelligence (AI) are to be used. Filtering systems based on AI or machine learning (ML) are suggested for the detection of unknown material. This involves a machine learning model being “trained” using known material. This model can then identify new material with a given degree of certainty. A large amount of data, consisting of both confirmed illegal material and legal material, is needed to train these models. It is safe to assume that only major companies have the resources to develop such models. Particularly when it comes to content moderation, this tends to result in a lack of transparency, as large companies treat the models they develop as a trade secret. In other words, this approach would further strengthen the dominance of the major tech companies. Users are left in the dark regarding the basis on which decisions are taken. This technology also has a single-digit error rate. If we conservatively assume an error rate of 1%, scanning one million images would produce 10,000 false positives; these would then have to be reviewed by providers or the EU Centre. To avoid tens or hundreds of thousands of false positives being produced each day, large companies are already using content moderators to pre-filter the reports. In other words, private, confidential images would initially pass through many hands, even if they are wrongly identified and reported by the algorithm. The use of AI in connection with highly confidential communications creates more potential problems than it solves.

Detection of grooming

The detection of grooming requires a detailed analysis of all text in chats. This means that every single message must be examined for suspicious patterns before it is sent or as part of the sending process. Attempts to moderate hate speech or extremism on social platforms, for example by using keywords, routinely lead to content being wrongly blocked, as keywords do not consider context. Detecting grooming is far more complex than detecting known visual material. When moderating text content, it is essential to consider the context if this is to be implemented in a rights-compliant manner.

Content moderation technologies are now much more advanced, but involve similar risks. What is known as “natural language processing” is used to detect illegal content. Again, this involves machine learning. In this case, the model is trained not with known images, but rather with known posts, messages or language patterns which are frequently used in grooming cases. But once again the error rates are very high, around 5 to 10% in the case of purely text-based analyses.[2] Roughly 10 billion text messages are sent in the EU each day, meaning that up to 1 billion messages would be wrongly flagged.[3] Human intervention would

[1] Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning – https://arxiv.org/abs/2106.09820

[2] ExtremeBB: Enabling Large-Scale Research into Extremism, the Manosphere and Their Correlation by Online Forum Data – https://arxiv.org/pdf/2111.04479.pdf

[3] Chat Control or Child Protection? – https://arxiv.org/abs/2210.08958

once again be needed, and private, non-criminal content would pass through many hands before being marked as unobjectionable.

None of these three methods are entirely error-free.[45] Users would have to live with the constant knowledge that all shared content, whether text or an image, could potentially end up in the hands of moderators or law enforcement agencies. The knowledge that they are being constantly monitored has an enormous impact on how users express themselves – even leading them to censor themselves (chilling effect). Current cases show that even if the material turns out to have been wrongly flagged, providers often permanently block affected accounts, or it takes a very long time and users have to jump through a great many hoops to have their accounts restored.[6]

Risks for private, confidential communication

The aforementioned technical tools are solely for the detection of material in general. The draft Regulation calls for detection to take place in all forms of interpersonal communication and by all hosting services. This includes publicly viewable platforms, private and business cloud storage services, but also chat services used for private and confidential communication, such as Signal, WhatsApp, Threema or email. Service providers offering end-to-end encryption could be forced to use technologies which break or circumvent the encryption. It should be noted that the following conditions must be met if we want to ensure trustworthy communication:

- The user's own device must be uncompromised and must not send content to third parties.
- Encryption must be secure so that we do not have to trust the net, which in this case means internet providers and similar actors.

The impact assessment[7] published with the draft Regulation indicates what technologies are being envisaged for implementation. It proposes that material is monitored using "on-device full hashing with matching at server". The hashing is carried out via the methods set out above. Client-side scanning, as it is known, is suggested for the scanning of material in encrypted communications.

Client-side scanning: Client-side scanning (CSS) is a new technology which enables law enforcement agencies and security authorities to circumvent encryption. Unlike in the case of earlier proposals, the authorities do not receive a key to a backdoor. CSS takes place directly on the device and analyses all content prior to encryption.[8] Suspicious material is then transmitted directly to third parties, e.g. content moderators or law enforcement agencies. Not only does CSS break the principle of end-to-end encrypted communication, namely that users can determine who has access to the content they send; it also means that surveillance software is hosted on the user's own device. Among other things, this may also make the device more vulnerable to "zero day exploits", which can be used by attackers. This need not be a faceless, abstract attacker; it can also be someone close to the user. A secure device for communication is especially important for victims of any kind of abuse, in

[4] Analysis of the reliability of perceptual hashing – https://www.hackerfactor.com/blog/index.php?/archives/2022/10/03.html

[5] ExtremeBB: Enabling Large-Scale Research into Extremism, the Manosphere and Their Correlation by Online Forum Data – https://arxiv.org/pdf/2111.04479.pdf

[6] Wrongly suspected father loses access to Google accounts – https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html

[7] Impact assessment – https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12726-Fighting-child-sexual-abuse-detection-removal-and-reporting-of-illegal-content-online_en

[8] Bugs in our Pockets: The Risks of Client-Side Scanning – https://arxiv.org/abs/2110.07450

particular. It was not for nothing that Apple backed away from a similar proposal it floated in 2021.[9] CSS inherently creates serious security and privacy risks for society as a whole, while the support it can offer law enforcement agencies is highly problematic, given the high error rate and lack of transparency. In addition, it should be noted that a monopoly may develop regarding the provision of these technologies. The Regulation requires small and medium-sized companies to implement detection orders as well, but of course they will not have the resources to develop the necessary technologies themselves. The EU Centre referred to in the Regulation will hardly be able to solve this problem; while it is supposed to make recommendations regarding the implementation of detection orders, this merely consists of a list of suitable technologies (Article 50 (1)). The Regulation also creates an incentive to develop further CSS-based surveillance technologies. This is highly questionable, especially against the backdrop of the surveillance scandal surrounding the Pegasus spyware and its developer, NSO Group.

All in all, the proposed Regulation suspends two fundamental rights: the privacy of telecommunications[10] and the fundamental right to protection of the confidentiality and integrity of information technology systems.[11] It goes so far as to turn this right on its head: protection from surveillance is replaced by mandatory surveillance without any suspicion of wrongdoing or concrete danger. Users lose control over what data they share with whom, undermining their trust in their own device on a fundamental level.

Age verification

The draft Regulation provides for age verification as a tool which should be used whenever there is a potential risk that the application can be used for the dissemination of CSAM or for grooming. It also contains an age verification obligation for app store providers (Article 6 (1)), which is intended to ensure that minors do not have access to such applications. Quite simply, this means a sweeping obligation to verify users' ages, as the risk can never be ruled out entirely. If age verification is implemented across the board, as envisaged in the draft Regulation, then the EU will, with Germany's involvement, be paving the way for online anonymity to be largely abolished. The CCC has already repeatedly emphasised that the right to anonymity is an essential prerequisite for the exercise of key fundamental rights.

Right to anonymity[12]

Anonymity is an important good, both in the real world and online. An important part of the political opinion-forming process is for the public to be able to obtain information and engage in discussion without feeling that they are being monitored or persecuted. Authenticity online must not come at the expense of anonymity, and must not be achieved by treating people as suspects to be identified. Operators of anonymous means of communication, such as Tor or virtual private networks (VPNs), must no longer face persecution and reprisals. To this end, it must be clearly established in law that they may not be sued for statements made via their services.

[9] Apple backs away from client-side scanning: https://netzpolitik.org/2022/chatkontrolle-apple-macht-rueckzieher-beim-client-side-scanning/

[10] Privacy of telecommunications – https://www.bfdi.bund.de/DE/Buerger/Inhalte/Telefon-Internet/TelekommunikationAllg/Fernmeldegeheimnis.html

[11] Protection of the confidentiality and integrity of information technology systems – https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2008/02/rs20080227_1bvr037007en.html

[12] CCC publishes drafting guidance for digital affairs in the new government programme – https://www.ccc.de/de/updates/2021/ccc-formulierungshilfe-regierungsprogramm

In addition to age verification using identity documents, age verification using biometric data is often suggested as a possible solution. Biometric data is highly sensitive (GDPR Article 9 paragraph (1)); the use of such a system would lead to the systematic processing and collection of the biometric data of children and young people, which is classed as special-category personal data. The problem with biometric data is that it is difficult to change, and so it can be used to uniquely identify people forever. Last year, CCC members managed to buy devices on eBay that had been used to collect the biometric data of local employees in Afghanistan. 2600 highly sensitive data sets were found on the devices; in the wrong hands, this data could have massively endangered the lives of these people. This shows the dramatic consequences of the irresponsible handling of biometric data.[13]

There is no such thing as a rights-compliant approach to age verification which does not undermine the right to anonymity. This was already acknowledged in the DSA (see recital 71). The solutions proposed in the DSA to mitigate the risks without resorting to age verification should be considered.

A lack of understanding of open-source software

The open-source community would also be particularly hard hit by mandatory age verification. Once again, this issue reveals that lawmakers do not understand open-source software, despite regarding it as an important societal good.

The infrastructure for open-source distributions is usually similar to that for the Arch Linux distribution. This Linux distribution makes it a particular priority to minimise the data it collects about users. Both the distribution itself and the packages (software applications) can be downloaded from a large number of “mirrors”. A mirror is a site which offers a mirrored storage location for files, thus ensuring they are available from multiple locations online. The mirrors are divided into three tiers: mirrors in Tier 0 contain the software packages produced by the Arch team themselves. Tier 1 mirrors are copies of Tier 0, while Tier 2 mirrors are copies of Tier 1. This decentralised approach to software distribution also ensures that developers cannot centrally collect information about their users, or track software package downloads.

To implement age verification, distributions – and open-source app stores such as F-Droid – would have to move away from this decentralised, data-minimising model. Furthermore, it would still be easy to circumvent the age verification, as the sources of all software packages (Git repositories) are publicly viewable. The ability to independently compile programs is a key prerequisite for the transparency of open-source software. To comply with the Regulation, however, the providers and users of open-source software would have to forego this. Levente Polyák from the Arch Linux distribution has said, for example, that implementing age verification would require a centralisation which is alien to the development of free software and which would be impossible to implement.[14]

Blocking URLs is not technically possible

The Regulation provides for webpages to be blocked by internet service providers (ISPs), but rather than the current domain-based approach, this is to be carried out on the basis of URLs (uniform resource locators). The Regulation divides blocking into two steps. First, the ISP has to determine whether the URL is even being accessed by users; then the URL is to be blocked. This approach is not technically possible, for several reasons. Firstly, modern transport encryption (https) means that the ISP does not know what specific URLs are

[13] Biometric data, a ticking time bomb – https://www.ccc.de/en/updates/2022/afghanistan-biometrie

[14] Chat control: an acute threat to open software – https://netzpolitik.org/2022/chatkontrolle-akute-gefahr-fuer-offene-software/

accessed by users. Secondly, targeted blocking of URLs is not possible without breaking secure transport encryption; additional surveillance tools would have to be used, such as deep packet inspection. Breaking https encryption in any way must be rejected as a matter of principle, as it is essential to the security of all online transactions, such as online banking, for example. Domain-based blocking is not very targeted, and leads, for example, to a block imposed on a shared hosting platform also applying to all other content hosted by the platform.[15] In Austria, there have already been cases where this type of blocking has led to much of the internet no longer being accessible.[16]

Role of the EU Centre

The creation of a European version of NCMEC (National Center for Missing and Exploited Children), which is responsible for managing a central database of abuse material and coordinating reports in the United States, is a key element of the Regulation. However, this EU Centre is to have a much wider remit. For example, the Centre is not only to be responsible for managing the image database, but also to play an important role in forwarding data to Europol and investigating authorities. The forwarding of data is defined very broadly, however: Article 48 (3) states that a report must be forwarded if it is “not manifestly unfounded”. Based on the technologies evaluated above, there is a significant risk here that sensitive data will be unlawfully forwarded to law enforcement agencies. Information provided by Irish law enforcement agencies in response to a query from the Irish Council of Civil Liberties (ICCL) shows that the data received from NCMEC is already handled irresponsibly. In this case, law enforcement agencies retained users’ data even after determining that they had been wrongly flagged.[17]

Conclusion

The draft Regulation falls far short of its aim of combating the dissemination of child abuse material. It completely ignores useful measures such as strengthening investigative capacities and ensuring that institutions which actively work to protect children are properly resourced. In its current form, it is actually counterproductive, as the huge number of false positives that will inevitably result from the Regulation could overwhelm the reporting structures and thus make it even more difficult to investigate criminals. This was also made clear by the police of the Netherlands at a parliamentary hearing, where they drew attention to the risk of being overwhelmed by the volume of data generated[18]:

“[...] Soon we will have a new law [in the Netherlands] which will criminalise sexting with minors. Dealing with that will already be a challenge [...] The current volume of reports is already hard to process now; a further regulation would lead to complete overload.”

The mountains of irrelevant material will keep officers from important investigative work. Investigations are not getting results, and the material found is not even being deleted, as several cases and parliamentary interpellations have shown.[19] Resolving these issues effectively should be the most important objective in the fight against child abuse.

[15] Protect the Stack: Why Infrastructure Providers Should Not Police Content – https://protectthestack.org

[16] Consequences of access blocking – https://netzpolitik.org/2022/overblocking-netzsperren-klemmen-in-oesterreich-legale-webseiten-ab/

[17] An Garda Síochána unlawfully retains files on innocent people who it has already cleared of producing or sharing of child sex abuse material – https://www.iccl.ie/news/an-garda-siochana-unlawfully-retains-files-on-innocent-people-who-it-has-already-cleared-of-producing-or-sharing-of-child-sex-abuse-material/

[18] Hearing at the Parliament of the Netherlands – https://debatgemist.tweedekamer.nl/node/29579

[19] Lack of a legal basis to delete material –
https://www.tagesschau.de/investigativ/panorama/kinderpornografie-loeschung-101.html

Any technical means of implementing the Regulation would mean the creation of an unprecedented surveillance infrastructure which interferes deeply in IT security principles and deprives users of any control over their digital communications. The proposal for a Regulation should be fundamentally rejected. If we want to help children and young people as quickly as possible, it would make more sense to develop better alternatives in partnership with child protection and technology experts.