European Parliament

Directorate-General for Parliamentary Research Services
Directorate for Impact Assessment and European Added Value
Ex-Ante Impact Assessment Unit
Head of Unit

Brussels, 03 February 2021

Note for the attention of the LIBE Committee Secretariat following the shadows meeting of 25 January 2021 on the proposal for a temporary derogation from the e-Privacy Directive for the purpose of fighting online child sexual abuse

Subject: *Answers by the contractor to the questions on the targeted substitute impact assessment on the Commission Proposal on a temporary derogation from the e-Privacy Directive for the purpose of fighting online child sexual abuse, COM (2020) 568 final, 2020/0259 (COD)*

Dear colleagues,

Following the presentation of the draft final targeted impact assessment commissioned by the EPRS at the LIBE shadows meeting of 25 January 2021 and the request by Shadow Rapporteur Mr Breyer for the author of the impact assessment study to provide the answers to the questions received in writing, we attach for the Rapporteur's and Shadow Rapporteurs' information the answers given by the contractor, Professor Jeanne Pia Mifsud Bonnici.

Yours sincerely,

Alexia Maniaki-Griva

2. If indiscriminate scanning of content is disproportionate, why would indiscriminate scanning with safeguards not be? Is there any case-law to support this assumption? Specifically, in the cited case La Quadrature, the ECJ did not say that safeguards make indiscriminate analysis proportionate?

Answer:
It is correct that indiscriminate scanning is disproportionate in any case. But the study proposes additional safeguards to limit the scanning so that it is not indiscriminate. Under section 6.9, the study proposes that the scanning of text-based communication should be subject to three strict limitations:

- *It should be carried out only when there is an already suspicion of soliciting child abuse or distributing CSAM (this was further elaborated on in answer to the questions raised by* ***Mr Zarzalejos Nieto*** *found at the end of this document),*
- *It should be restricted in time, and*
- *It should be subject to periodic review by DPAs.*

It should also be noted that these restrictions are in addition to the other proposed safeguards.

The case La Quadrature du Net (C-511/18 and C-512/18) is mentioned on page 35 to support the argument advanced by the EDPS, stating that "Even if the technology used is limited to the use of "relevant key indicators", …the deployment of such general and indiscriminate analysis is excessive."

3. The study finds that "**the techniques used to detect text-based child grooming** involve indiscriminate monitoring and automated analysis of the private messages of all users" and "should be limited to private messages of persons already under suspicion of soliciting child abuse or distributing CSAM" (p. 47). However further up the study states that searching for previously unknown images also constitutes indiscriminate monitoring (p. 35). Does the search for previously unknown images therefore also need to be limited to suspects?

Answer:
In terms of technology, there is a clear distinction between hashing algorithms and machine learning/artificial intelligence. The detection of completely unknown images and videos would require AI/ML, but the technologies investigated here (Facebook's PDQ and MTK+PDQF, and Thorn's Safer) use perceptual hashing algorithms. They do not look for 'completely unknown' images and videos containing CSAM, but rather images and videos that are similar to those known to contain CSAM. What they are really looking for is CSAM that would not be caught by regular hashing algorithms if they are distorted, edited, or contain only portions of images/videos known to contain CSAM.

Thorn's Safer presents a combination. Along with perceptual hashes, Safer has also incorporated machine learning algorithms to create 'classifiers' that are used to detect CSAM images and videos that were entirely unknown. The method for training and *implementing these algorithms has not been elaborated on by official publications, but a third-party source states that they used "datasets containing thousands of images"[1] to train the image classifier. The Commission has mentioned some of these classifiers as being nudity, shapes or colours,[2] but this has not been confirmed in Thorn's public sources. However, the privacy-intrusiveness of the machine learning component of the tool cannot be estimated, since documentation on its training or implementation is unavailable. Due to their similar mechanisms, the machine learning component's privacy concerns need to be evaluated in the same manner as Microsoft's Project Artemis.*

Therefore, searching for completely unknown images would need to be limited to suspects to the extent that they indiscriminately collect original images and videos and process them using AI/ML. For the technologies that use perceptual hashing algorithms, this would not be needed. To be clear, this distinction applies to any technology used by NI-ICS providers to the extent that they utilise AI/ML algorithms to detect completely unknown CSA images and videos. For tools that use both AI/ML and hashing algorithms, each component needs to be measured against different standards so that the rights of users are safeguarded.

4. The assessment correctly recommends: "**When they transfer personal data to third countries or international organisations, NI-ICS providers should ensure that third countries have a level of protection essentially equivalent to that guaranteed by EU** law." (p. 44) Would you agree that this GDPR requirement means that US corporations can no longer disclose content and communications data to NCMEC because the US is a third country without an equivalent level of protection?

Answer:
There are situations where providers are confronted with conflicting legal requirements. US corporations have a duty under US law to report CSAM once they have actual knowledge of it. Strictly speaking, any reporting of personal data to third countries can be carried out only when equivalent protections are ensured. In keeping with the limited Terms of Reference for the impact assessment, we did not check whether there is equivalent level of protection and did not delve deeper into this.

Our suggestion would be to include a similar obligation of a duty to report, for e.g., to Europol (to the extent that this can be done within the Europol's mandate; currently, they cannot receive reports from private actors). From the exchange of written questions and responses with Europol, we understand that they have a database which is currently second only to NCMEC[3]. If Europol's mandate were to allow it and

[1] Goswami, A. [2020], 'Thorn's 'Safer' uses perceptual hashing and machine learning algorithms to identify child sexual abuse material (CSAM)', *Marktechpost*, September 7 [online] Available at: https://www.marktechpost.com/2020/09/07/thorns-safer-uses-perceptual-hashing-and-machine-learning-algorithms-to-identify-child-sexual-abuse-material-csam/ (Accessed 28 January 2021)

[2] Questions for written answer to the Commission by the EP rapporteur Birgit Sippel, S&D, and her shadows (28 September 2020) (file with authors) p.56.

[3] Text of question asked to Europol "*Is there an additional benefit (from EUROPOL's/law enforcement perspective) for a database of CSAM reports administered by a European organisation/public authority?*

There are significant benefits to having a single database of CSAM reports provided by OTT providers to EU Member States. Europol currently receives and securely processes and stores these for some Member States. Europol has one of the largest volumes of CT referral material securely stored globally, second only to NCMEC. This has led to significant advances in different CSE investigations because this data can be cross-matched against existing data from non-related CSE investigations received from Member States and Third Country partners. As a result, children have been identified and suspects identified and arrested. To maximise the potential of this data, *the Proposed Regulation were to include it, a good solution would probably be to include a duty to report to Europol that is similar to the duty to report to NCMEC. In addition, from the perspective of efficiency of investigation, this may actually be more preferable at this moment in time, as currently NI-ICS providers send information to NCMEC, which NCMEC then passes on to US authorities for transmission to Europol. This could reduce the time of the process.*

5. The assessment suggests that **communications content should be retained even where no alleged child grooming** is detected, in order to allow for later historical analysis, "in line with the CJEU's judgments regarding the Data Retention Directive" (p. 46). Is it not very clear from the CJEU's case-law that the general and indiscriminate retention of communications data (thus all the more of communications content) is disproportionate?

Answer:
From our understanding, and specifically with regard to Microsoft Project Artemis, the detection of child grooming and abuse is only conducted through an analysis of historical text-based chat conversations. Conversations on Microsoft's services (such as Xbox Live) are not analysed while they are in progress. Therefore, the use of Microsoft Project Artemis necessitates the retention of communications data for the detection of text-based CSAM. The data retention safeguards suggested on p.46 are an attempt at drawing from the CJEU's case-law to strike a balance between the rights and interests at issue: protecting the rights of users of NI-ICS while combating child sexual abuse.

In general, AI/ML algorithms for the detection of CSAM work in two steps: (1) retention of data; and (2) analysis of that data. Each step requires separate safeguards to ensure that the rights of users are protected:

(1) For data retention: It is allowed only to the extent that it is strictly necessary for the detection of CSA, for a limited time and subject to effective review (either judicial or administrative).
(2) For analysis: strict limitations on scope and time – please see the response to Q.2 above.

6. The Proposed Regulation should specifically state that its scope does not include **end-to-end encryption**." (p. 47) Should this exemption from the scope be made in the text of the regulation or is it sufficient to have in a recital only?

Answer:
From a legal certainty perspective, we think that it would be better if it were in the text of the Proposed Regulation. However, this is not something we researched as it was beyond the scope of the Terms of Reference of the impact assessment.

in other words, to use it for both short and long-term strategic analysis is an aim of the GRACE (Global Response Against Child Exploitation) H2020 funded project."