Council of the European Union
General Secretariat

Brussels, 08 May 2024

WK 6697/2024 INIT

LIMITE

JAI	**FREMP**
ENFOPOL	**TELECOM**
CRIMORG	**COMPET**
IXIM	**MI**
DATAPROTECT	**CONSOM**
CYBER	**DIGIT**
COPEN	**CODEC**

This is a paper intended for a specific community of recipients. Handling and further distribution are under the sole responsibility of community members.

WORKING DOCUMENT

From:	Presidency
To:	Law Enforcement Working Party (Police)
Subject:	Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse - Presentation of the new approach proposed by the Presidency

The Presidency provides delegations in the Annex with the presentation of its new approach on the CSA Regulation made at the meeting of the Law Enforcement Working Party (Police) on 8 May 2024.

CSA Regulation

New approach on targeting detection orders

beEU
belgium24.eu

Keeping the current approach on enhanced risk assessment and categorization

- Risk categorization of services (clear description of parameters, thresholds and terms): see document WK 3036/2024 REV 3

- 3 risk categories (high, medium, low risk)

- Measures will differ depending on the identified risk category

New approach on targeting detection orders

- **New scope**
  - Detection only on visual content (images and videos) and URLs
  - No detection on audio and text
  - Known CSAM, new CSAM and grooming (through CSAM) remain in the scope of detection orders

- **Protection of end-to-end encryption**
  - No detection on E2EE data
  - Proposed wording in article 1(5): "*Without prejudice to Article 10a, this Regulation shall not prohibit or make impossible end-to-end encryption, implemented by the relevant information society services or by the users. This Regulation shall not create any obligation to decrypt or create access to end-to-end encrypted data, or that would prevent providers from offering end-to-end encrypted services*."

- **Detection prior to transmission (upload moderation) and subject to user consent**
  - For all interpersonal communications services
  - Using different detection technologies for known CSAM (cryptographic/perceptual hashing) and new CSAM (artificial intelligence)

User consent

- How would the users give their consent?
  - The terms and conditions of the service provider could inform the user of the fact that detection prior to transmission could be deployed if the provider receives a detection order. The user would then have the choice, either:
    - To consent to the detection of visual content and URLs, in which case the user would have access to the entirety of the functionalities of that service, and notably to share visual content and URLs, or;
    - Not to consent to the detection of visual content and URLs, in which case the user would have access to all the functionalities of the service, except sharing visual content and URLs.

1. Known CSAM

- Detection based on cryptographic / perceptual hashing
  - Cryptographic hashing and perceptual hashing technologies create a unique fingerprint (hash) for each image they scan and compare this fingerprint against the hashes in a database (indicators) of known CSAM. **Cryptographic hashing** is only able to detect two exact images, while the added value of perceptual hashing, such as Photo DNA, Facebook's PDQ hash function and Apple's NeuralHash function, lays in the fact that the generated hash is robust against common image transformations such as resizing, compression, blurring, noise, etc. while also being sensitive to perceptual similarity. **Perceptual hashing** is able to identify "similar" CSAM to known CSAM based on same/similar hash values (even if they are not entirely identical at the pixel level).

- Detection 'prior to transmission'
  - The service provider will scan images and videos for CSAM (and URLs) when the user uploads them and sends them to the receiver, on the condition that the user gives its <u>consent</u> for such detection.

- Proposed wording in Article 10a of the Regulation:
  - *"In order to implement this Regulation, providers of interpersonal communications services shall install and operate technologies to detect, prior to transmission, the dissemination of known child sexual abuse material or of new child sexual abuse material"*

2. New CSAM

- Detection based on artificial intelligence

  - Based on machine learning algorithms trained on large datasets of CSAM to learn patterns and features associated with CSAM (after confirmation if it is CSAM or not). This however requires the analysis of the visual content of images and videos rather than hashing.

  - The use of artificial intelligence results in more false positives than hashing. This could be countered by:

    - Delayed reporting after two consecutive hits;
    - Pseudonymization prior to human review: pseudonymization will be applied to the metadata related to personal information, but even face blurring techniques could be considered. When there is a match of possible new CSAM, the provider would pseudonymize the metadata included in the report to the EU Centre, so that the EU Centre would not be able to know to whom the data belongs until it verifies that the content is not manifestly unfounded. In this case, the EU Centre would tell the provider and the provider would share with the EU Centre the metadata in the clear, so that the EU Centre can forward it to law enforcement.

3. Solicitation of children

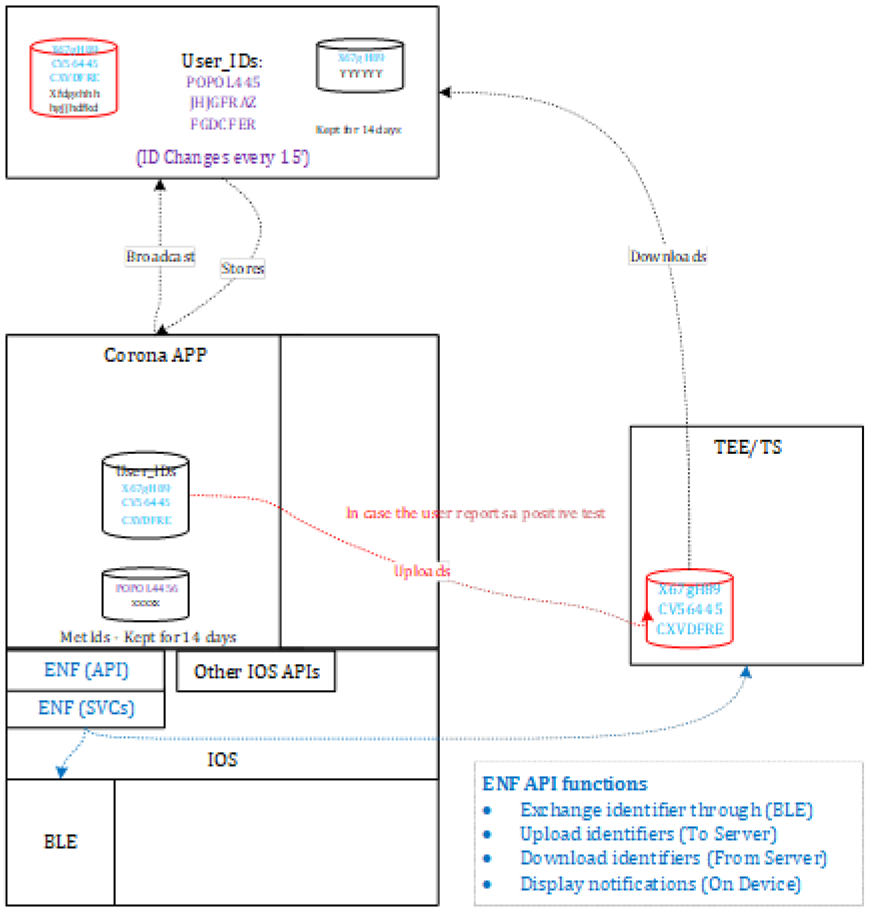
- Based on the detection of known and new CSA visual content.

- Grooming usually entails that at some stage of the “iter criminis” images or videos with sexual connotation are exchanged. Their detection would allow to capture a fairly large part of grooming attempts.

- Only in a few cases, there is no exchange of images/videos during the grooming process.

Appendix 1.1 How the coronavirus application works and interacts with iOS

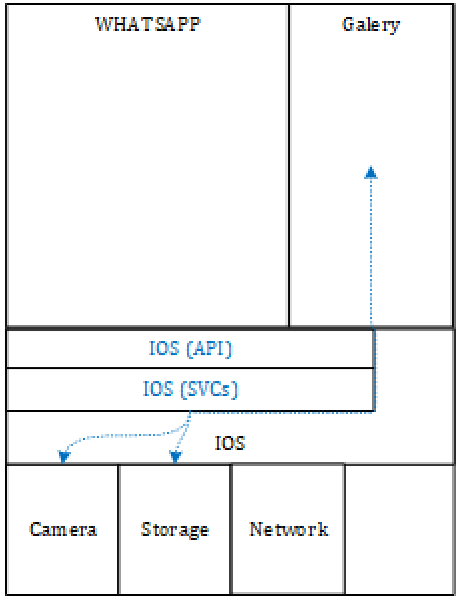
- The coronavirus application is a mobile app that helps users track their exposure to COVID-19 and get notified if they have been in contact with someone who tested positive.
- User gets a unique identifier updated every 15 minutes, transmitted via Bluetooth Low Energy to nearby devices with the app.
- No personal data is linked to this identifier, which is stored locally for 14 days on users' devices.
- If a user tests positive, it alerts other devices with the app.
- If close contact with a confirmed case, the user is notified and provided with relevant health guidance.
- The app operates within iOS through the Exposure Notification framework, ensuring privacy, consent, and minimal impact on device performance.
- Users can enable/disable features and permissions as needed, with no collection of personal data beyond the encrypted identifiers used solely for contact tracing.



ENF API functions
- Exchange identifier through (BLE)
- Upload identifiers (To Server)
- Download identifiers (From Server)
- Display notifications (On Device)

Appendix 1.2 Example: How to send a picture through WhatsApp

- When a user sends a picture through WhatsApp, they can either choose an existing picture from their gallery or take a new picture using the camera. In both cases, WSP needs to access the device's storage and camera, which are controlled by the operating system (OS). To do this, WSP needs to have the appropriate permissions from the user and the OS.
- When permissions are granted, WSP can use the OS's application programming interface (API) to communicate with the device's hardware and software.
- After compressing and encrypting the picture, WSP sends it to the WSP server, who forwards it to the recipient's device, where it is decrypted and decompressed by the WSP app.
- The recipient can then view the picture on their device



Access requires permission from the OS and from the user

AV Foundation API functions

- Open camera
- Capture image (from Camera)
- Save image to storage

PHOTOS API functions

- Access Galery APP
- Select image
- Read image (from storage)

Appendix 1.3 : High-level message sending process (1)

To detect images in the context of sending messages only, the scan of the images is to be triggered.
- As part of a message creation flow started from the application
- By the OS that needs to act as a relay
- The trusted APP is invoked by an online or offline (message-bus) communication

Use-case 1: Import an existing picture from the gallery (from the messaging application) :

In this case, the scanning could be implemented into the get image from storage IOS function which triggers a call either directly to the scanning application or indirectly through a message BUS. If the encrypted image or its hash have been obtained by the scanning application and stored in a safe place, the sending of the message can continue.

Messaging APP	OS	Scanning APP
Create message (Messaging APP)		
Add image from Galery		
API CALL	Access gallery	
API CALL	Select image	
API CALL	Read image	
	Get image from storage	
	API CALL	Scan image (In trusted APP)
		Store image on secure enclave
Send message	Return image	Hash and encrypt image
		Homomorphic function à Result
		IF CSAM detected trigger next steps of procedure

Appendix 1.3 : High-level message sending process (2)

Use-case 2: Take a picture from the camera in the messaging application :

In this case, the scanning could be implemented into the get image from camera IOS function

Messaging APP	OS	Scanning APP
Create message (Messaging APP)		
Add image from Camera		
API CALL	**Open camera**	
API CALL	**Capture Image**	
	Get image from camera	
	API CALL	**Scan image**
		Store image on secure enclave
	Save image to storage	Hash and encrypt image
Send message		Homomorphic function → Result
		IF CSAM detected trigger next steps of procedure

RISK CATEGORISATION	DETECTION ORDER (DO)	MITIGATION MEASURE(S)(MM)	PROVIDER FLAGGING NEED FOR DETECTION ORDER	FREQUENCY OF (RE)CATEGORISATION
RISK ++ HIGH	**DO** - Known, New CSAM and Grooming (through detection of CSAM) included - only images/videos & URL - including services using E2EE - Delayed and pseudonymized reporting for new CSAM	Obligatory additionnal MM Sanction(s) ++	Flagging, by the provider, of possible need to be subject to a detection order	Up to 12 months
RISK + MEDIUM	None	Obligatory additionnal MM Sanction(s) +	Yes	Up to 24 months
RISK - LOW	None	Recommended additional MM	Yes	Up to 36 months