Client-Side Scanning, Privacy and CSAM

`How to Reduce the Sharing of Child Sexual Abuse Material on Messaging Platforms`

Context

At several recent events discussing the apparent dichotomy of privacy and safety, it has become clear that several assumptions are circulating regarding the use of so-called "client-side scanning" with messaging applications, particularly those that employ "end-to-end" encryption.

The purpose of this short document is to address some of those assumptions, explaining a number of related processes at the same time.

What is Client-Side Scanning?

Firstly, to clarify the term. Client-side scanning generally refers to some form of scanning or processing of a message before its transmission. In the case of the prevention of the circulation of CSAM, this would normally involve checking whether any attached files contained known CSAM content through a process of "hash matching".

What is Hash Matching?

Hash matching is a process where images are converted to digital fingerprints ("hashes"), which are then compared with a database of known CSAM images. Any matches indicate the presence of illegal material. Importantly, hash matching works without exposing the actual content during the matching process, preserving user privacy and data security. The hash functions used are one-way, meaning the original content cannot be reconstructed from the hash value.

Does Client-Side Scanning Break or Weaken Encryption?

Client-side scanning takes place before any message is encrypted, therefore, it does not have any impact on encryption. It does not require encryption to be weakened in any way, for example, it does not need the creation of a backdoor.

Does Client-Side Scanning Enable Governments to Read Your Messages?

One claim I've heard is that client-side scanning is the equivalent of allowing a third party, usually a government or law enforcement agency, to stand behind you and read your messages as you create them. In the panel discussion, this suggestion was attributed to an article in the Dutch media; a more detailed reference was not provided.

To make sense, this would suggest that the capability to undertake any client-side scanning was inserted by a government or other third party. Whilst this could be done, it would require the application provider to be complicit, which is unlikely given the privacy policies that many operate.

Does Client-Side Scanning Happen Already?

Many messaging applications already employ client-side scanning techniques for user convenience functions. For example, an application will typically detect the presence of a URL as it is typed and insert an image of the relevant page.

When providing the image, an application may fetch the page directly from the device, sharing the IP address in the process, or do so via a central server, potentially allowing the platform operator to garner data on user activity. It is possible to provide this functionality without compromising user privacy if care is taken by the platform operator. Inspection of their privacy policy should indicate whether this is the case.

Blocking the Circulation of CSAM versus Privacy and Freedom of Expression

The final objection that is often raised relates to the potential impact on privacy or freedom of expression, noting that privacy concerns may, in turn, impact freedom of expression. As noted previously, client-side scanning functionality provided by the platform operator need not have any impact on user privacy or freedom of expression.

The use of hash matching to identify attempts to share known CSAM is undertaken without any inspection of the accompanying message as it is only the inclusion of an image file that is of interest. And as noted previously, the use of hash matching ensures that the actual image is not exposed during the matching process.

If a match occurs, the platform operator is required in many jurisdictions to report the user to the relevant law enforcement agency. Whilst this is technically a breach of the privacy of that user, we should remember that privacy is a qualified right because it can be lawfully restricted in democratic societies to protect broader public interests or other fundamental rights, in this case, the rights of the victims of CSAM.

Why is this important?

Many of us underestimate the scale of online child sexual abuse and exploitation (CSAE). Estimates by Childlight suggest that there are 300 million victims per annum (see Into the Light report 2024, page 3), that’s nearly 14% of children in the world. And research by Protect Children has highlighted the widespread use of messaging platforms by pedophiles to share CSAM (see Tech Platforms Used by Online Child Sexual Abuse Offenders, page 11), .

Conclusions

Client-side scanning is already in widespread use and can be deployed in a manner which preserves privacy. When used in combination with hash matching, it offers a privacy-preserving method to reduce the volume of known CSAM being shared via messaging platforms, in turn helping to reduce the continual re-victimisation of the victims of CSAE.