Australian Government | eSafetyCommissioner

Meeting Agenda – Apple developments

Apple representatives	[redacted]
eSafety representatives	Julie Inman Grant, [redacted] (possible [redacted], [redacted]
Meeting date	28 September 2021
Note taker	[redacted]

1. **Introductions**

2. **Platform/service update - new product and feature developments - expanded protections for children**

Update on timing/next steps for introducing the expanded protections for children

a. What engagement have you had with privacy and advocacy groups following the recent announcements to delay the rollout? What do new consultation measures entail? Who will you be engaging with?

b. What consultation was undertaken for the original expanded protections for children?

c. What are Apple's intended next steps? Are Apple seeking to cancel the developments entirely, simply roll out nearly identical features after a delay or find a middle ground?

d. What is the best avenue for eSafety to engage and actively participate in the consultation process?

Expanded protections for children

e. Which child safety organisations are providing the hashed images to Apple?

f. Who is responsible for the threshold secret sharing and determining the threshold number of positive CSAM detections required for escalation to Apple?

[redacted]

[redacted] [redacted]

[redacted] [redacted]

i. [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

[redacted] [redacted]

Commissioner Briefing Note –

Apple's Expanded Protections for Children developments

To	Julie Inman Grant
From	[redacted], [redacted]
Date	7 September 2021
Subject	Apple Expanded Protections for Children developments
Meeting date	Tuesday, 28 September 2021

Purpose

This briefing provides additional information about Apple's recently announced Expanded Protections for Children and questions to be raised at eSafety's next meeting with Apple on **Tuesday, 28 September 2021.**

Timeline of events

- In August 2021 Apple announced the forthcoming release of expanded protections for children. The proposed developments included Scanning iCloud images for CSAM, Communication safety in Messages and Expanding guidance in Siri and Search.
- eSafety met with Apple on Friday, 6 August 2021 to discuss the developments. A file note of the meeting is available here.
- Following the meeting, eSafety compiled a product briefing on the developments, surfacing gaps in Scanning iCloud images for CSAM and two key areas for further discussion – being CSAM detection thresholds and the sources of hashed images that Apple would rely on.
- On 3 September 2021, Apple announced that the proposed developments would be delayed to make improvements based on feedback from customers, advocacy groups, researchers, and others.
- eSafety will meet with Apple on Tuesday, 28 September to discuss concern raised in the product briefing and revised timing release of the developments.

Key points for discussion

1. **Community consultation and revised timing for release of the developments** – Apple has delayed the rollout of child-safety features over privacy concerns. Limited information is known about the consultation Apple undertook when developing the suite of expanded protections for children. Apple has advised they will engage in further community consultation prior to release of the developments.
2. **Sources of hashed images** – seek clarification regarding the child safety organisations providing images and validation – eSafety understand NCMEC provide hashes but have no further confirmed information about the other child safety organisations participating or their remit/geographical location.

3. **CSAM detection thresholds** – seek clarification about who is determining the threshold secret sharing number of positive CSAM detections required for escalation to Apple (Noting Apple has advised that they are not aware of the specific number required).
4. [redacted]

Key questions for Apple

Update on timing/next steps for introducing the expanded protections for children

- What engagement have you had with privacy and advocacy groups following the recent announcements to delay the rollout?? What do new consultation measures entail? Who will you be engaging with?
- What are Apple's intended next steps? Are Apple seeking to cancel the developments entirely, simply roll out nearly identical features after a delay or find a middle ground?
- What consultation was undertaken for the original expanded protections for children?
- What is the best avenue for eSafety to engage and actively participate in the consultation process?

Expanded protections for children

- Which child safety organisations are providing the hashed images to Apple?
- Who is responsible for the threshold secret sharing and determining the threshold number of positive CSAM detections required for escalation to Apple?

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]

[redacted]

[redacted]

[redacted]

Action 2: Seek clarification on revised timing for release.

Specific information about the sources of hashed images Apple will rely on

- The system performs on-device matching using a database of known CSAM image hashes provided by NCMEC and other child-safety organisations.
- This set of image hashes is based on images acquired and validated to be CSAM by at least two child safety organisations.
- eSafety understand NCMEC provide hashes but have no further insights about the other child safety organisations participating or their remit/geographical location.
- Anne Collier has published an article suggesting Apple's technology only detects copies of known images already sitting in the databases of NCMEC and the IWF.

Action 3: Seek clarification on the child safety organisations providing hashed images.

Specific information about CSAM detection and detection thresholds

- Apple introduced a new system for stopping child-abuse imagery on iOS devices.
- CSAM Detection enables Apple to accurately identify and report iCloud users who store known CSAM in their iCloud Photos accounts.
- Instead of scanning images in the cloud, the system performs **on-device matching** using a database of known CSAM image hashes provided by NCMEC and other child-safety organizations.
- This set of image hashes is based on images acquired and validated to be CSAM by at **least two child safety organisations**. Apple further transforms this database into an unreadable set of hashes, which is securely stored on users' devices.
- The hashing technology, called **NeuralHash,** analyses an image and converts it to a unique number specific to that image. Only another image that appears nearly identical can produce the same number; for example, images that differ in size or transcoded quality will still have the same NeuralHash value.
- Before an image is stored in iCloud Photos, an on-device matching process is performed for that image against the database of known CSAM hashes. This matching process is powered by a cryptographic technology called private set intersection, which determines whether there is a match without revealing the result. The device creates a cryptographic safety voucher that encodes the match result. It also encrypts the image's NeuralHash and a visual derivative. This voucher is uploaded to iCloud Photos along with the image.

- Using another technology called **threshold secret sharing**, the system ensures that the contents of the safety vouchers cannot be interpreted by Apple unless the iCloud Photos account crosses a threshold of known CSAM content.
- Only when the threshold is exceeded does the cryptographic technology allow Apple to interpret the contents of the safety vouchers associated with the matching CSAM images.
- Apple servers flag accounts exceeding a threshold number of images that match a known database of CSAM image hashes. Apple manually review each report to confirm there is a match, disables the user's account, and sends a report to NCMEC.
- Apple can't access metadata or visual derivatives for matched CSAM images until a threshold of matches is exceeded for an iCloud Photos account.
- Apple has provided that they do not know the specific threshold number.

Action 4: Seek clarification about who is responsible for determining the threshold secret sharing number of positive CSAM detections required for escalation to Apple.

[redacted]

- [redacted]
- [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]

[redacted]

- [redacted]
    - [redacted]
    - [redacted]
    - [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
    - [redacted]
    - [redacted]

Meeting Note – Apple developments

Apple representatives	[redacted]
eSafety representatives	Julie Inman Grant, [redacted] (possible [redacted], [redacted]
Meeting date	28 September 2021
Note taker	[redacted]