Apple's Child-safety Photo-scanning Plans?

Aditya Gollapudi & Betsy Pu

BEFORE WE START:

- We will mostly be talking about Child Abuse from a statisitcal and technical perspective but if you ever feel uncomfortable at any time please feel free to walk out - we won't mind
- Please try your best to be acknowledge other people's experiences
- Speak from an I perspective
- Assume others are speaking in good faith

The scale of CSAM and Child Abuse

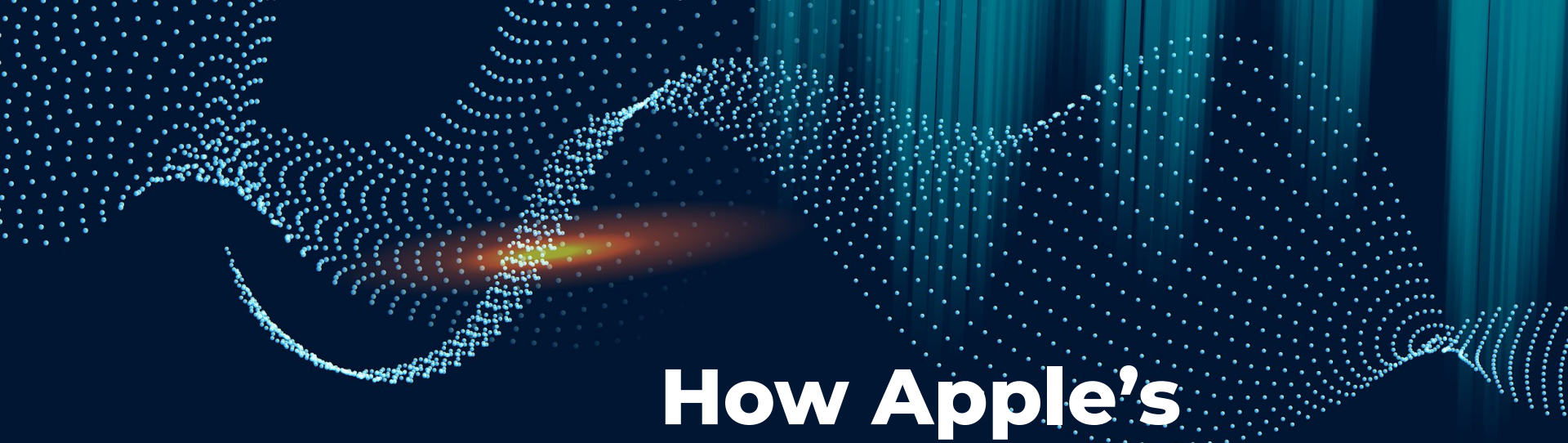
a broad and possibly inaccurate summary

Order of magnitude

- In 2017 there were 9.4M reported cases of CSAI
- By 2019 it had jumped to 45M, the problem is growing exponentially
- 68% of abuse in Asia, 19% the Americas, 6% Europe, and 7% Africa
    - But hosting tells a different story with a whopping 52%* being hosted in the Netherlands, and 90% in Europe
        - The rest is mostly hosted in the US
- 1/9 girls and 1/53 boys experience sexual assault before they reach the age of 18
- Jump to 1/5 and 1/20 for sexual abuse

Legal Background on CSAM - in the US

- CSAM is not protected by default under the 1st amendment (Ferber)
- It is illegal to own or distribute child pornography or material marketed as child pornography
- Artistic work that do not claim to include minors (e.g. Romeo and Juliet) are fine
- Academic work is fine
- Drawn and animated works remain a gray area
- The courts have repeatedly struck down laws such as the CPPA and COPA that seek to significantly raise standards on platforms to regulate whether children may view pornography



How Apple's CSAM detection works

a brief technical overview

CSAM detection

in plain english:

- NeuralHash
  - for all images in a database of CSAM, Apple computes a fingerprint of the image
  - it sends
- PSI (Private Set Intersection) protocol
  - on Apple devices, computation is done that identifies which images are in the CSAM db, but the device can't tell
  - iCloud servers can decrypt the info of intersected images... but only if:
- threshold secret sharing
  - enough (~30 images) match
- this entire process is triggered upon UPLOADING to iPhoto cloud

Under this system...

Apple CAN/DOES

- know the identity of accounts that have exceeded the threshold of flagged images
- manually review all reports before forwarding to NCMEC, looking at “visual derivatives” (~low res versions*) of the flagged images

Apple CANNOT

- learn anything about images that do not match the known CSAM database
- access metadata or visual derivatives for matched CSAM images until a threshold of matches is exceeded for an iCloud Photos account
- know how many images are flagged for an account (if it is under the threshold)*
- know anything about images NOT uploaded to iCloud

DICEY?

- control over the database contents is an administrative decision with no technical guarantees
- Can hide what is in the DB but cannot hide the size of the DB*

What do other companies do?

- Companies are required to report any CSAM they find to NCMEC
  - **But they are not required to look for it**
- The vast majority of the tech industry uses photoDNA a technology developed by Microsoft
  - Used by Law Enforcement, NCMEC/VIC, Microsoft, Google, Reddit, Discord, Adobe, Facebook, Twitter, etc.
  - Like Apple's system uses a "fingerprint"
  - Extends to video by taking a key frame from each scene
- Google/Youtube have a proprietary ML system known as CSAI match
  - Also used by Adobe, Twitter, Reddit

Concerns

- technologically easy for this system to be used to flag other content, besides CSAM
- security risks when introducing more complexity
- complexity in jurisdiction, what if every country has their own idea of monitoring?
- potential technological abuses

Apple

- Used before images are uploaded to iCloud
- Only exposes low res images
- No data* is revealed before the threshold is crossed (30 images)

PhotoDNA

- Largely used for public images
- Except for oneDrive
- Alerts as soon as a single image is flagged
- Extended to cover other topics
- Terrorism, porn, upskirts, etc.

CSAI Match

- Public video
- Detects videos outside of an existing database
- Manual review of the video in original form
- Similar tech is used to detect broad categories of undesirable content

“We have faced demands to build and deploy government-mandated changes that degrade the privacy of users before, and have steadfastly refused those demands.”

— Apple

Apple's iCloud partner in China will store user data on servers of state-run telecom

GCBD is contracting out some data storage needs to China Telecom

By Nick Statt | @nickstatt | Jul 18, 2018, 2:37pm EDT

Apple bows to China by censoring Taiwan flag emoji

Apple removes thousands of games from the Chinese App Store, alarming observers

More than 47,000 games were removed

By Jay Peters | @jaypeters | Aug 18, 2020, 6:34pm EDT

Apple blackmailed into removing Russian tactical voting app Navalny [U]

Ben Lovejoy - Sep. 17th 2021 5:36 am PT @benlovejoy

Not a uniquely Apple problem

Australia passes surveillance bill that lets police take over accounts, alter, and delete data

Critics say new police powers are too broad

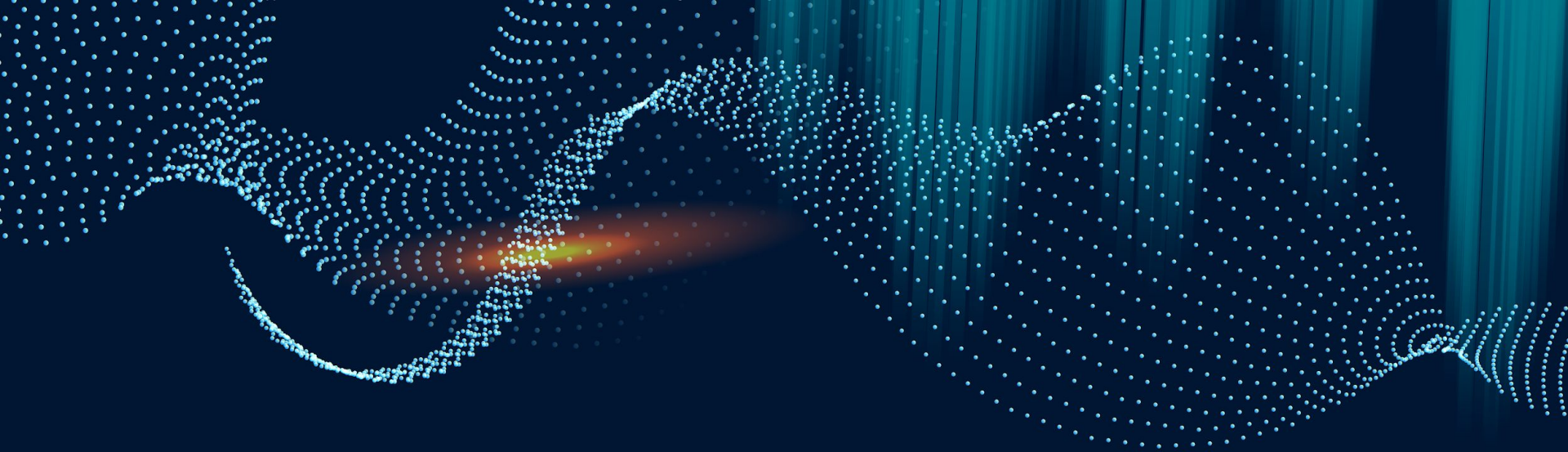
By Daniel Sims September 1, 2021, 8:12 PM | 51 comments

ProtonMail logged IP address of French activist after order by Swiss authorities

Natasha Lomas, Romain Dillet / 7:46 AM EDT • September 6, 2021

Comment

some aussies on this business



CRYPTO WARS

recent events are just a
continuation of this old debate

1990’s: Export regulation of crypto

- “encryption is a munition pls don’t take it out of the country”

2003: “Patriot Act II”

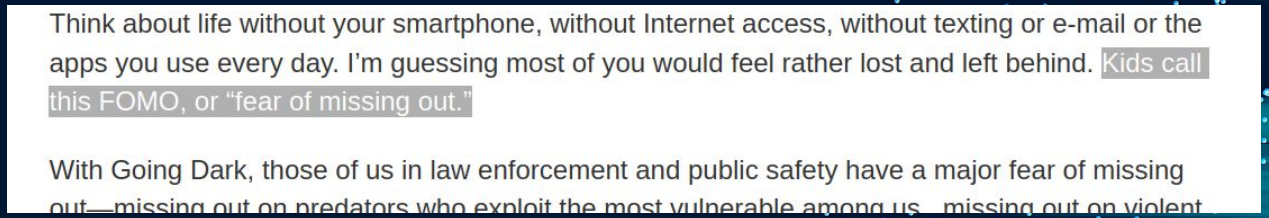
- Domestic Security Enhancement Act of 2003
- centered around **terrorism**
- included a provision adding a minimum of five years to the sentence of any convicted felon who used encryption to conceal “incriminating communication or information” related to their crime
- Never became law, but was govt’s first foray into modern encryption regulation

2007: NSA-backed encryption backdoor discovered

- the next several years are a battle between NIST and the industry on strong encryption

2008+: “Going Dark” rhetoric

- 2014 FOMO by FBI director James Comey



Think about life without your smartphone, without Internet access, without texting or e-mail or the apps you use every day. I’m guessing most of you would feel rather lost and left behind. Kids call this FOMO, or “fear of missing out.”

With Going Dark, those of us in law enforcement and public safety have a major fear of missing out—missing out on predators who exploit the most vulnerable among us, missing out on violent

2013: Snowden reveal

- For years, the NSA and the FBI directly accessed U.S. technology companies' servers to scoop up their users' data without a warrant.
- accelerated a trend towards strong encryption

2015: San Bernardino Shooting

- Apple refuses to decrypt phone

DISCUSSION

- given similar systems like PhotoDNA and Google's CSAI match, why is the biggest backlash hitting Apple for their actions?
- Is the progression towards more regulation that pushes against privacy/encryption inevitable?
- how would you want to design systems that balance privacy and certain forms of accountability/safety?
- to what extent should the debate around exceptional access be centered around CSAM, as opposed to other issues like violence, drug trafficking, extremism, terrorism, hate speech, domestic abuse? Is it appropriate for CSAM to be the singular target of Apple's system? Do we expect this to remain the case?