Apple's Plan to "Think Different" About Encryption Opens a Backdoor to Your Private Life

العربية 简体中文 FRANÇAIS DEUTSCH POLSKI РУССКИЙ ESPAÑOL

Apple has announced impending changes to its operating systems that include new "protections for children" features in iCloud and iMessage. If you've spent any time following the Crypto Wars, you know what this means: Apple is planning to build a backdoor into its data storage system and its messaging system.

Child exploitation is a serious problem, and Apple isn't the first tech company to bend its privacy-protective stance in an attempt to combat it. But that choice will come at a high price for overall user privacy. Apple can explain at length how its technical implementation will preserve privacy and security in its proposed backdoor, but at the end of the day, even a thoroughly documented, carefully thought-out, and narrowly-scoped backdoor is still a backdoor.

To say that we are disappointed by Apple's plans is an understatement. Apple has historically been a champion of end-to-end encryption, for all of the same reasons that EFF has articulated time and time again. Apple's compromise on end-to-end encryption may appease government agencies in the U.S. and abroad, but it is a shocking about-face for users who have relied on the company's leadership in privacy and security.

There are two main features that the company is planning to install in every Apple device. One is a scanning feature that will scan all photos as they get uploaded into iCloud Photos to see if they match a photo in the database of known child sexual abuse material (CSAM) maintained by the National Center for Missing & Exploited Children (NCMEC). The other feature scans all iMessage images sent or received by child accounts—that is, accounts designated as owned by a minor—for sexually explicit material, and if the child is young enough, notifies the parent when these images are sent or received. This feature can be turned on or off by parents.

When Apple releases these "client-side scanning" functionalities, users of iCloud Photos, child users of iMessage, and anyone who talks to a minor through iMessage will have to carefully consider their privacy and security priorities in light of the changes, and possibly be unable to safely use what until this development is one of the preeminent encrypted messengers.

Apple Is Opening the Door to Broader Abuses

We've said it before, and we'll say it again now: it's impossible to build a client-side scanning system that can only be used for sexually explicit images sent or received by children. As a consequence, even a well-intentioned effort to build such a system will break key promises of the messenger's encryption itself and open the door to broader abuses.

That's not a slippery slope; that's a fully built system just waiting for external pressure to make the slightest

All it would take to widen the narrow backdoor that Apple is building is an expansion of the machine learning parameters to look for additional types of content, or a tweak of the configuration flags to scan, not just children's, but anyone's accounts. That's not a slippery slope; that's a fully built system just waiting for external pressure to make the slightest change. Take the example of India, where recently passed rules include dangerous requirements for platforms to identify the origins of messages and pre-screen content. New laws in Ethiopia requiring content takedowns of "misinformation" in 24 hours may apply to messaging services. And many other countries—often those with authoritarian governments—have passed similar laws. Apple's changes would enable such screening, takedown, and reporting in its end-to-end messaging. The abuse cases are easy to imagine: governments that outlaw homosexuality might require the classifier to be trained to restrict apparent LGBTQ+ content, or an authoritarian regime might demand the classifier be able to spot popular satirical images or protest flyers.

We've already seen this mission creep in action. One of the technologies originally built to scan and hash child sexual abuse imagery has been repurposed to create a database of "terrorist" content that companies can contribute to and access for the purpose of banning such content. The database, managed by the Global Internet Forum to Counter Terrorism (GIFCT), is troublingly without external oversight, despite calls from civil society. While it's therefore impossible to know whether the database has overreached, we do know that platforms regularly flag critical content as "terrorism," including documentation of violence and repression, counterspeech, art, and satire.

Image Scanning on iCloud Photos: A Decrease in Privacy

Apple's plan for scanning photos that get uploaded into iCloud Photos is similar in some ways to Microsoft's PhotoDNA. The main product difference is that Apple's scanning will happen on-device. The (unauditable) database of processed CSAM images will be distributed in the operating system (OS), the processed images transformed so that users cannot see what the image is, and matching done on those transformed images using private set intersection where the device will not know whether a match has been found. This means that when the features are rolled out, a version of the NCMEC CSAM database will be uploaded onto every single iPhone. The result of the matching will be sent up to Apple, but Apple can only tell that matches were found once a sufficient number of photos have matched a preset threshold.

Once a certain number of photos are detected, the photos in question will be sent to human reviewers within Apple, who determine that the photos are in fact part of the CSAM database. If confirmed by the human reviewer, those photos will be sent to NCMEC, and the user's account disabled. Again, the bottom line here is that whatever privacy and security aspects are in the technical details, all photos uploaded to iCloud will be scanned.

Make no mistake: this is a decrease in privacy for all iCloud Photos users, not an improvement.

Currently, although Apple holds the keys to view Photos stored in iCloud Photos, it does not scan these images. Civil liberties organizations have asked the company to remove its ability to do so. But Apple is choosing the opposite approach and giving itself *more* knowledge of users' content.

Machine Learning and Parental Notifications in iMessage: A Shift Away From Strong Encryption

Apple's second main new feature is two kinds of notifications based on scanning photos sent or received by iMessage. To implement these notifications, Apple will be rolling out an on-device machine learning classifier designed to detect "sexually explicit images." According to Apple, these features will be limited (at launch) to U.S. users under 18 who have been enrolled in a Family Account. In these new processes, if an account held by a child under 13 wishes to send an image that the on-device machine learning classifier determines is a sexually explicit image, a notification will pop up, telling the under-13 child that their parent will be notified of this content. If the under-13 child still chooses to send the content, they have to accept that the "parent" will be notified, and the image will be irrevocably saved to the parental controls section of their phone for the parent to view later. For users between the ages of 13 and 17, a similar warning notification will pop up, though without the parental notification.

Similarly, if the under-13 child *receives* an image that iMessage deems to be "sexually explicit", before being allowed to view the photo, a notification will pop up that tells the under-13 child that their parent will be notified that they are receiving a sexually explicit image. Again, if the under-13 user accepts the image, the parent is notified and the image is saved to the phone. Users between 13 and 17 years old will similarly receive a warning notification, but a notification about this action will not be sent to their parent's device.

This means that if—for instance—a minor using an iPhone without these features turned on sends a photo to another minor who does have the features enabled, they do not receive a notification that iMessage considers their image to be "explicit" or that the recipient's parent will be notified. The recipient's parents will be informed of the content without the sender consenting to their involvement. Additionally, once sent or received, the "sexually explicit image" cannot be deleted from the under-13 user's device.

Whether sending or receiving such content, the under-13 user has the option to decline without the parent being notified. Nevertheless, these notifications give the sense that Apple is watching over the user's shoulder—and in the case of under-13s, that's essentially what Apple has given parents the ability to do.

These notifications give the sense that Apple is watching over the user's shoulder—and in the case of under-13s, that's essentially what Apple has given parents the ability to do.

It is also important to note that Apple has chosen to use the notoriously difficult-to-audit technology of machine learning classifiers to determine what constitutes a sexually explicit image. We know from years of documentation and research that machine-learning technologies, used without human oversight, have a habit of wrongfully classifying content, including supposedly "sexually explicit" content. When blogging platform Tumblr instituted a filter for sexual content in 2018, it famously caught all sorts of other imagery in the net, including pictures of Pomeranian puppies, selfies of fully-clothed individuals, and more. Facebook's attempts to police nudity have resulted in the removal of pictures of famous statues such as Copenhagen's Little Mermaid. These filters have a history of chilling expression, and there's plenty of reason to believe that Apple's will do the same.

Since the detection of a "sexually explicit image" will be using on-device machine learning to scan the contents of messages, Apple will no longer be able to honestly call iMessage "end-to-end encrypted." Apple and its proponents may argue that scanning before or after a message is encrypted or decrypted keeps the "end-to-end" promise intact, but that would be semantic maneuvering to cover up a tectonic shift in the company's stance toward strong encryption.

Whatever Apple Calls It, It's No Longer Secure Messaging

As a reminder, a secure messaging system is a system where no one but the user and their intended recipients can read the messages or otherwise analyze their contents to infer what they are talking about. Despite messages passing through a server, an end-to-end encrypted message will not allow the server to know the contents of a message. When that same server has a channel for revealing information about the contents of a significant portion of messages, that's not end-to-end encryption. In this case, while Apple will never see the images sent or received by the user, it has still created the classifier that scans the images that would provide the notifications to the parent. Therefore, it would now be possible for Apple to add new training data to the classifier sent to users' devices or send notifications to a wider audience, easily censoring and chilling speech.

But even without such expansions, this system will give parents who do not have the best interests of their children in mind one more way to monitor and control them, limiting the internet's potential for expanding the world of those whose lives would otherwise be restricted. And because family sharing plans may be organized by abusive partners, it's not a stretch to imagine using this feature as a form of stalkerware.

People have the right to communicate privately without backdoors or censorship, including when those

people are minors. Apple should make the right decision: keep these backdoors off of users' devices.

JOIN THE NATIONWIDE PROTEST

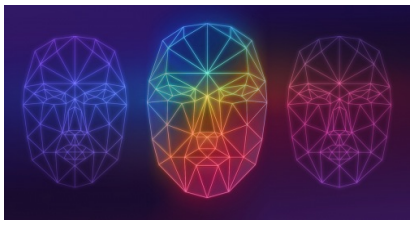
TELL APPLE: DON'T SCAN OUR PHONES

Read further on this topic:

- If You Build It, They Will Come: Apple Has Opened the Backdoor to Increased Surveillance and Censorship Around the World
- How LGBTQ Content is Censored Under the Guise of "Sexually Explicit"
- EFF Joins Global Coalition Asking Apple CEO Tim Cook to Stop Phone-Scanning
- Apple's Plan to Scan Photos in Messages Turns Young People Into Privacy Pawns
- 25,000 EFF Supporters Have Told Apple Not To Scan Their Phones
- Delays Aren't Good Enough—Apple Must Abandon Its Surveillance Plans
- Don't Stop Now: Join EFF, Fight for the Future at Apple Protests Nationwide
- Protestors Nationwide Rally to Tell Apple: "Don't Break Your Promise!"
- Why EFF Flew a Plane Over Apple's Headquarters

JOIN EFF LISTS

RELATED UPDATES



DEEPLINKS BLOG BY JASON KELLEY, INDIA MCKINNEY, MATTHEW GUARIGLIA | FEBRUARY 9, 2022

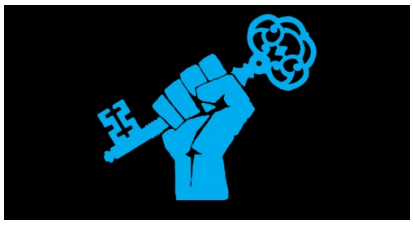
Victory! ID.me to Drop Facial Recognition Requirement for Government Services

DEEPLINKS BLOG BY ALEXIS HANCOCK, JON CALLAS | FEBRUARY 9, 2022

What the Duck? Why an EU Proposal to Require "QWACs" Will Hurt Internet Security

DEEPLINKS BLOG BY JASON KELLEY | FEBRUARY 8, 2022

If EARN IT Passes, What Happens On Your iPhone Won't Stay On Your iPhone

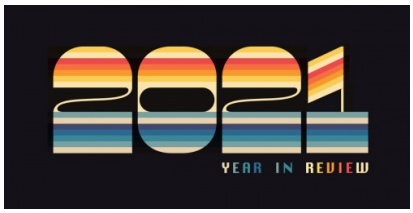


DEEPLINKS BLOG BY JOE MULLIN | JANUARY 21, 2022

The U.K. Paid $724,000 For A Creepy Campaign To Convince People That Encryption is Bad. It Won't Work.

PRESS RELEASE | JANUARY 6, 2022

"Worst in Show Awards" Livestreams Friday: EFF's Cindy Cohn and Cory Doctorow Will Unveil Most Privacy-Defective, Least Secure Consumer Tech Products at CES



DEEPLINKS BLOG BY VERIDIANA ALIMONTI | DECEMBER 27, 2021

The Battle for Communications Privacy in Latin America: 2021 in Review

DEEPLINKS BLOG BY EVA GALPERIN | DECEMBER 25, 2021

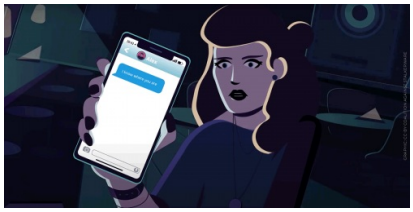
Stalkerware: 2021 in Review

DEEPLINKS BLOG BY ALEXIS HANCOCK | DECEMBER 15, 2021

EU's Digital Identity Framework Endangers Browser Security

DEEPLINKS BLOG BY KAREN GULLO, EVA GALPERIN | DECEMBER 15, 2021

Apple's Android App to Scan for AirTags is a Necessary Step Forward, But More Anti-Stalking Mitigations Are Needed



DEEPLINKS BLOG BY EVA GALPERIN | NOVEMBER 25, 2021

Coalition Against Stalkerware Celebrates Two Years of Work to Keep Technology Safe for All