Текст · 6 августа 2021, 06:34
Александр Бородихин, Давид Френкель,

Ключи к фото. Как Apple собралась бороться с детской порнографией в айфонах

Что случилось

«Наша цель в *Apple* — создавать технологии, которые расширяют возможности и обогащают жизнь людей — при этом помогая им оставаться в безопасности, — говорится в пресс-релизе компании. — Мы хотим помочь защитить детей от маньяков[1], которые с помощью средств связи ищут жертв для сексуальной эксплуатации, и ограничить распространение материалов с сексуальным насилием над детьми».

Новые функции коснутся трех сервисов.

Во-первых, эпловское приложение *Messages* научится ограждать несовершеннолетних от откровенных фото: если на телефон ребенка придет такое фото, его «распознают» офлайновые алгоритмы машинного обучения, картинку скроют, а посмотреть ее можно будет только после двух экранов с предупреждениями — в том числе о том, что родители получат уведомление о просмотре подозрительного фото (функция доступна только для детей младше 13 лет, ее уже раскритиковали как потенциально угрожающую ЛГБТ-подросткам).

Во-вторых, голосовой помощник *Siri* и айфоновский поиск начнут реагировать на запросы, связанные с детской порнографией: вместо ответа телефон

предложит взрослому анонимно обратиться за помощью.

Самым спорным оказался третий механизм: он позволяет айфону сопоставлять загружаемый в *iCloud* фотоархив с базой изображений, на которых запечатлены сцены насилия над детьми или откровенные фото с несовершеннолетними.

«Эта инновационная технология», отмечается в пресс-релизе *Apple,* позволит оперативно передавать информацию о детском порно правоохранительным органам, одновременно предоставляя «значительные преимущества в плане конфиденциальности по сравнению с существующими методами»: корпорация сможет узнать о содержимом фотографий только в том случае, если в их облачный аккаунт *iCloud* будет загружен уже известный файл со сценами сексуального насилия. Если пользователь не пользуется *iCloud,* его фото анализироваться не будут.

«Метод *Apple* разработан с учетом конфиденциальности пользователей. Вместо анализа изображений в облаке, операционная система сопоставляет фото прямо на устройстве, используя базу данных хэшей известных изображений», — подчеркивают разработчики. Они опубликовали технические документы с описанием алгоритма, чтобы развеять сомнения в безопасности «обычных» пользователей: предусмотрены разнообразные отсечки для ложных срабатываний и проведены границы между пользовательским устройством и эпповским облаком.

Как это работает

Искать подозрительные фотографии *Apple* планирует по базе изображений с различными сценами насилия над детьми, которую собирает американская некоммерческая организация Национальный центр пропавших и эксплуатируемых детей *(NCMEC);* в будущем могут добавиться другие партнеры.

Чтобы сделать это прямо на устройстве пользователя, *Apple* загрузит на него специальную версию базы, где вместо самих фотографий будут находиться только «хэши» — числа, «описывающие» снимок при помощи алгоритма так, чтобы одинаковые значения получались только из очень похожих фотографий. Хэши шифруются с помощью ключа, который известен только *Apple* и не передается на сам айфон.

Аналогичные хэши айфон независимо от *Apple* вычисляет для всех фотографий пользователя. Алгоритм получения «хэшей» называется *NeuralHash* и учитывает не конкретные пиксели, а именно содержание изображения: сначала с помощью нейронной сети он получает большой ряд чисел, «описывающих» изображение — так чтобы у похожих визуально и по содержанию картинок эти числа были близки. Затем ряд преобразуется в одно целое число — собственно, «хэш». Это число должно совпадать для одной и той же фотографии разного размера или разрешения, даже если фотографию каким-то образом обрезать.

Затем для каждой фотографии создается «ваучер безопасности», содержащий шифрованные результаты сравнения с базой снимков детской порнографии, метаданные фото и некое производное

изображение для предпросмотра, подробности о котором *Apple* не раскрывает.

Поскольку значения в базе, с которой ведется сравнение, зашифрованы, телефон не может сам сравнить «хэши» своих фотографий с «хэшами» из базы, это происходит уже на серверах *Apple.* Дальше разработчики предусмотрели еще один реверанс в сторону безопасности: данные в базе на телефоне отсортированы таким образом, чтобы для любого «хэша» можно было найти его предполагаемый порядковый номер в базе на сервере; айфон затем использует данные, хранящиеся по этому номеру, и сам «хэш» в качестве ключа шифрования «ваучера безопасности». На сервера *Apple* отправляется лишь часть исходного «хэша», по которой можно восстановить расшифрованное значение из базы и попытаться расшифровать «ваучер».

Если «хэш» полностью идентичен, то есть фотография действительно совпадает со снимком из базы, то расшифровка пройдет успешно. Если же «хэш» отличается, значит картинка в базе не найдена, и расшифровать «ваучер» не получится. То есть, как уверяет *Apple,* компания не получит доступ к фото, пока не найдет аналогичного в базе детского порно.

Еще одна технология, призванная обезопасить пользователей, называется *Threshold Secret Sharing*[2].

Идея заключается в том, чтобы расшифровать данные «ваучеров» можно было только после определенного количества совпадений с базой — большее некоего заданного порога. Порог *Apple* не раскрывает, но указывает, что вероятность ошибки — 1 к триллиону.

Здесь интересна логика авторов: метаданные «ваучера» шифруются еще одним ключом, который генерируется на айфоне и состоит из нескольких частей, и число этих частей завязано на пороговое число. Внутри каждого зашифрованного «ваучера» на сервер передается лишь часть этого ключа, поэтому расшифровать содержимое «ваучера» полностью *Apple* сможет только тогда, когда количество совпавших с базой фотографий превысит порог.

Наконец, чтобы *Apple* не могла самостоятельно узнать, сколько прямо сейчас у конкретного пользователя на айфоне совпадений фото с базой, время от времени телефон будет создавать фейковый «ваучер»: он будет проходить проверку на совпадение с базой, но не будет содержать необходимой части ключа, то есть не приближать пользователя к порогу совпадений.

Если же порог все же будет преодолен, *Apple* будет считать аккаунт пользователя подозрительным, в связи с чем оставляет за собой право расшифровать данные из его *iCloud* и передать на проверку людям — если подозрения подтвердятся, пользователя заблокируют и о нем сообщат правоохранительным органам.

Apple обратилась к трем независимым специалистам по криптографии, чтобы те отрецензировали разработку, и они не нашли серьезных проблем с ее реализацией.

Идея использовать криптографические алгоритмы для автоматизации поиска детской порнографии не нова: ФБР использует их с начала нулевых, а в 2009 году появилась самая известная разработка в этом направлении — PhotoDNA корпорации Microsoft (она также использует хэши и базу NCMEC для сравнения). При этом пока успехов в борьбе с распространением таких материалов не было: только в 2018 году технологические компании насчитали в интернете не менее 45 миллионов фото и видео, попадающих под определение детской порнографии.

Поисковые гиганты Google и Bing используют для фильтрации поисковой выдачи базу данных, собираемую британской Internet Watch Foundation. Российский «Яндекс», который специалисты обвиняют в неудовлетворительной фильтрации детского порно, подключиться к базе IWF не может: по данным Fortune, это связано с запретом российских властей на совместные проекты с иностранными НКО.

В России аналог базы NCMEC пыталась создать в 2014 году Лига безопасного интернета Елены Мизулиной, которая ссылалась на американский опыт. Тогда в организации утверждали, что в базе оказались «более 11 терабайт» фото и видео, которым были присвоены «уникальные маркеры» для дальнейшего сопоставления. Дальнейших упоминаний этой инициативы обнаружить не удалось; на фоне обсуждения базы ЛБИ в даркнете появилось объявление о продаже базы, но ее размер не превышал 241 гигабайта.

А что не так?

Специалисты по кибербезопасности предсказуемо встретили инициативу *Apple* с возмущением: многочисленные попытки обезличить анализ фото не отменяют главной проблемы — самого факта постоянного доступа к пользовательским файлам. Даже если файлы, которые анализирует алгоритм, загружаются в облачное хранилище, фактически пользователь не делится ими ни с кем.

«Независимо от того, каковы долгосрочные планы *Apple,* они посылают очень четкий сигнал. По их (очень влиятельному) мнению, создавать системы, сканирующие телефоны пользователей на предмет запрещенного контента, безопасно. Такой сигнал они посылают правительствам, конкурирующим сервисам, Китаю, вам лично. Окажутся они правы или нет в этом вопросе, едва ли важно, — сокрушается профессор Университета Джонса Хопкинса Мэтью Грин, который первым рассказал в твиттере о готовящемся нововведении. — Отныне ничего святого, теперь правительства [разных стран] будут требовать того же ото всех, а когда мы поймем, что это была ошибка, будет слишком поздно».

Embed Wrapper Not Realised
Embed Wrapper Not Realised

Embed Comment Not Realised

«Всякий раз, когда технологическая компания делает уступки в вопросах приватности пользователя для достижения политических или социальных целей, но ограничивает внедрение технологии какой-нибудь невероятно сложной непонятной криптографией, примерно через пять минут политики и активисты

начинают жаловаться, что сделано недостаточно, и компания может сделать еще больше для помощи детям, для борьбы с терроризмом или с наркокартелями, просто отказавшись от этой заумной криптографии», — пишет британский журналист и активист Дэнни О'Брайен.

«*Apple* пересматривает свою позицию по приватности и толкает нас в 1984 год», — еще более жесток в формулировках один из разработчиков сквозного шифрования для *Facebook Messenger* Алек Маффет, который уверен, что корпорация делает «огромный шаг назад в вопросе личной приватности».

Правозащитная организация *Electronic Frontier Foundation* уже выступила с громкой отповедью *Apple*, заклеймив новую технологию бэкдором[3]

и предупредив, что компания открывает путь дальнейшим нарушениям пользовательских свобод.

Использование слова бэкдор тут символично: *Apple* на протяжении последних лет активно участвовала в публичной дискуссии о безопасности компьютерных систем на стороне пользователя и вступила в противостояние с властями США, требовавшими дать следователям доступ к айфонам через бэкдоры в операционных системах. В 2015-2016 годах *Apple* отказалась помогать во взломе айфона стрелка из американского Сан-Бернардино, судилась с ФБР и подчеркивала, что защиту устройств нельзя компрометировать искусственными лазейками.

Те заявления корпорации резко контрастируют с сегодняшними нововведениями и звучат скорее

как приведенные выше цитаты ее критиков: «Последствия требований правительства ужасают. Если правительству позволить пользоваться [законом], чтобы облегчить разблокировку вашего *iPhone,* оно получит возможность проникнуть в устройство любого человека, чтобы получить его данные».

Редактор: Дмитрий Трещанин

---

1. В оригинале используется слово predator (хищник) — таким термином в англоязычной среде принято называть сексуальных преступников.
2. Порог секретности.
3. Встраиваемая разработчиком в код приложения «лазейка», позволяющая ему при необходимости получить оперативный доступ к защищенной системе; специалистами по безопасности считается серьезной уязвимостью.