CANADIAN CENTRE for CHILD PROTECTION®
Helping families. Protecting children.

ARGUMENTE VON WISSENSCHAFTLERN, WELCHE DIE EU AUFFORDERN, IHREN REGULIERUNGSVORSCHLAG ZU CSA ZURÜCKZUZIEHEN, SIND PRAXISFERN UND HALTEN EINER PRÜFUNG NICHT STAND

11. Juli 2023

von Lloyd Richardson, Direktor der IT-Abteilung des Canadian Centre for Child Protection

Letzte Woche forderte eine internationale Gruppe von Wissenschaftlern die EU auf, ihre Bemühungen, um Regulierungsmaßnahmen aufzugeben, welche Technologieunternehmen verpflichten würden, Anstrengungen zu unternehmen, um die Verbreitung von Missbrauchsdarstellungen (CSAM) und Versuche von Straftätern, Kinder auf ihren Plattformen sexuell zu missbrauchen (Cybergrooming), aufzudecken.

Der Widerstand der Gruppe beruht im Wesentlichen auf der Auffassung, man solle sich gar nicht erst die Mühe machen, ein System zur Verhinderung der sexuellen Ausbeutung von Kindern im Internet einzuführen, wenn es nicht perfekt sei. Dies ist eine sehr verbreitete Ansicht, wenn es darum geht, Schutzmaßnahmen in digitalen Räumen einzuführen. Einige der Behauptungen sind auch falsch, irreführend oder beides.

Unser Team im Canadian Centre for Child Protection verfügt über mehr als 10 Jahre praktischer Erfahrung mit vielen der von der Gruppe kritisierten Technologien. Unser Fachwissen ist nicht theoretisch, sondern angewandt: Wir haben dafür gesorgt, dass mit diesen bewährten Mitteln Millionen von Missbrauchsdarstellungen aus dem Internet entfernt wurden. Lassen Sie uns vor diesem Hintergrund mit einigen der Behauptungen in dem offenen Brief auseinandersetzen:

„Darüber hinaus ist es auch möglich, ein legales Bild zu erstellen, das fälschlicherweise als illegales Material erkannt wird, da es den gleichen Hashwert wie ein in der Datenbank vorhandenes Bild hat (falsch positiv). Dies ist auch ohne Kenntnis der Hash-Datenbank möglich. Ein solcher Angriff könnte dazu verwendet werden, unschuldige Nutzer in die Falle zu locken und die Strafverfolgungsbehörden mit Falschmeldungen zu überschwemmen, wodurch Ressourcen von den eigentlichen Ermittlungen zum sexuellen Missbrauch von Kindern abgezogen würden.“

Es ist möglich, unschuldige Benutzer in die Falle zu locken und die Feuerwehren mit Falschmeldungen zu überfluten, indem man Feueralarme auslöst. Anstatt die Feuerwehren abzuschaffen, sollten wir diese Art von unsozialem Verhalten durch gesetzliche Sanktionen verhindern. Außerdem ist die Erstellung eines falsch positiven Bildes ohne Zugriff auf die Hauptdatenbank viel zeit- und ressourcenaufwändiger als das Auslösen eines Feueralarms.

„Als Wissenschaftler gehen wir nicht davon aus, dass es in den nächsten 10 bis 20 Jahren möglich sein wird, eine skalierbare Lösung zu entwickeln, die auf den Geräten der Nutzer läuft, ohne dass illegale Informationen durchsickern, und die bekannte Inhalte (oder Inhalte, die von bekannten Inhalten abgeleitet sind oder mit ihnen in Zusammenhang stehen) zuverlässig, d. h. mit einer akzeptablen Anzahl von falsch positiven und negativen Ergebnissen, erkennen kann.“

Dies ist das Kernproblem: Eine „akzeptable Anzahl“ von falsch positiven und negativen Ergebnissen wird nie angegeben. Wenn die „annehmbare Anzahl“ gleich Null ist, wird die CSAM-Erkennung einem höheren Standard unterworfen als alle anderen Technologien zur Eindämmung von Missbrauch. Denken Sie an Radarpistolen, Verkehrskameras, Sicherheitsscanner an Flughäfen, Drogenspürhunde: Jede Erkennungsmethode birgt das Risiko falsch positiver Ergebnisse, aber wir geben diese Technologien nicht auf; stattdessen bauen wir Protokolle und Richtlinien um sie herum, um diese Limitierungen und Risiken zu berücksichtigen und sicherzustellen, dass falsch positive Ergebnisse keine Konsequenzen für unschuldige Personen haben. Dasselbe Argument wurde gegen das PhotoDNA-Scannen von Bildern vorgebracht; das Scannen wurde vor Jahren eingeführt, und die Apokalypse der Fehlalarme ist nie eingetreten.

„Bei dem Ausmaß, in dem private Kommunikation online ausgetauscht wird, würde selbst das Scannen der in der EU ausgetauschten Nachrichten bei nur einem App-Anbieter bedeuten, dass jeden Tag Millionen von Fehlern erzeugt würden."

In dem Schreiben heißt es: „Es hat keine offene und objektive Bewertung stattgefunden, die ihre Wirksamkeit nachweist", worauf stützt sich also diese Einschätzung?

„Falsch positive Ergebnisse sind auch beim Einsatz von Erkennungstechnologien unvermeidlich - selbst bei bekanntem CSAM-Material."

Es stimmt, dass kein Detektionssystem fehlerfrei ist, aber es gibt unzählige Beispiele für Detektionssysteme, bei denen Sicherheit und Schutz sorgfältig gegen die Fehlerrate abgewogen wurden.

„Die einzige Möglichkeit, dies auf eine akzeptable Fehlermarge zu reduzieren, wäre, nur in engen und wirklich gezielten Fällen zu scannen, in denen ein vorheriger Verdacht besteht."

Auch hier wird offengelassen, was „akzeptabel" wäre, und wir sollen es als „Null" interpretieren.

„.... die große Anzahl von Menschen, die benötigt werden, um Millionen von Texten und Bildern zu überprüfen."

Wieder eine Zahl ohne Quellenangabe.

„Zweitens kann ein Virus anhand einer kleinen eindeutigen Zeichensequenz erkannt werden, was bei einem Bild oder Video nicht der Fall ist: Es wäre sehr einfach, eine eindeutige Zeichensequenz zu verändern, was sich auf die Erkennung des Hashwertes, aber nicht auf die Abbildung selbst auswirkt; während dieses Vorgehen bei einem Virus den Code unbrauchbar machen würde."

Laut Industrie nicht wahr; siehe z.B. https://www.kaspersky.com/resource-center/definitions/what- is-a-polymorphic-virus. „Eine im letzten Jahr veröffentlichte Studie zeigt, dass erstaunliche 97 Prozent der analysierten Viren polymorphe Eigenschaften haben". Virenscanner verwenden ungenaue Abgleiche und heuristisches Scannen wie jede andere Erkennungstechnologie auch.

„Solche Tools würden angeblich funktionieren, indem sie Inhalte auf dem Gerät des Nutzers scannen, bevor diese verschlüsselt oder nachdem sie entschlüsselt wurden, und anschlagen, wenn illegales Material gefunden wird. Das könnte man mit dem Anbringen von Videokameras in unseren Häusern gleichsetzen, die jedes Gespräch abhören und Berichte senden, wenn wir über unerlaubte Themen sprechen."

In einem CSS-Regime werden nur Medien im Transit (d. h. auf dem Weg vom Gerät zum externen Server) auf CSAM „geprüft". Beachten Sie, dass „geprüft" die Terminologie ist, die WhatsApp für „scannen" verwendet, wenn es beschreibt, wie es CSS auf Malware durchführt; siehe https://faq.whatsapp.com/667552568038157/.

Medien, die sich auf dem Gerät des Nutzers befinden, werden nicht „geprüft". Im Gegensatz zu Videokameras in unseren Häusern ist CSS eher mit einem Rauchmelder vergleichbar, der eine wichtige Sicherheitsfunktion erfüllt, ohne unsere normalen Aktivitäten zu beeinträchtigen.

„Der einzige Einsatz von CSS in der freien Welt war von Apple im Jahr 2021, was nach eigenen Angaben dem neuesten Stand der Technik entspricht.“

Technologieunternehmen setzen CSS-Techniken schon seit Jahren in verschiedenen Zusammenhängen ein. Die Messaging-App Signal verwendet eine Technik, die verkürzte Hashes der Telefonnummern in Ihrem lokalen Adressbuch verwendet, um Kontakte zu finden; siehe https://signal.org/blog/private-contact-discovery/. Apple setzt CSS derzeit ein, um sensible Fotos zu erkennen; siehe https://support.apple.com/en-us/HT212850.

„Diese Bemühungen wurden nach weniger als zwei Wochen aufgrund von Datenschutzbedenken und der Tatsache, dass das System bereits gekapert und manipuliert worden war, wieder eingestellt.“

Solange dies nicht mit einem Zitat belegt ist, handelt es sich um eine reine Spekulation über die Motive von Apple.

„Wenn CSS auf dem Gerät einer Person installiert wird, verhält es sich wie Spyware und ermöglicht es Gegnern, einfachen Zugang zu diesem Gerät zu erhalten.“

Völlig falsch und irreführend. Es gibt keinen technischen Zusammenhang zwischen clientseitigem Scannen und der Möglichkeit, sich Zugang zu einem Gerät zu verschaffen.

„Jedes Gesetz, das CSS oder eine andere Technologie für den Zugriff, die Analyse oder die Weitergabe von Kommunikationsinhalten vorschreibt, untergräbt zweifellos die Verschlüsselung.“

Auch das ist unwahr; es gibt keine technische Verbindung zwischen CSS und Verschlüsselung.

„Selbst, wenn ein solches CSS-System denkbar wäre, besteht ein extrem hohes Risiko, dass es missbraucht wird. Wir gehen davon aus, dass erheblicher Druck auf die politischen Entscheidungsträger ausgeübt werden wird, den Anwendungsbereich auszuweiten, um zunächst die Anwerbung von Terroristen, dann andere kriminelle Aktivitäten und schließlich die Äußerung von Dissidenten zu erfassen.“

Auf welcher Grundlage schätzen sie das „extrem hohe“ Missbrauchsrisiko und die Bereitschaft der Regierungen ein, CSS zu kompromittieren? Das gleiche Argument wurde vor 15 Jahren gegen das serverseitige PhotoDNA-Scannen vorgebracht; was ist aus diesen Vorwürfen geworden?

„Wenn ein solcher Mechanismus implementiert würde, müsste er zum Teil durch Security by Obscurity (Sicherheit durch Unklarheit) erfolgen, da es sonst für die Benutzer leicht wäre, die Erkennungsmechanismen zu umgehen, indem sie zum Beispiel die Datenbank von Hash-Werten leeren oder einige Überprüfungen umgehen.“

Auch das stimmt nicht; viele Sicherheitsmechanismen (DVD- und Blu-ray-Player, Schutz von HDMI-Inhalten, SIM-Sperren für Mobiltelefone, Beschränkungen bei Firmware-Updates, Sicherheit von Mobilgeräten, Verschlüsselung von Computerfestplatten) beruhen auf geräteinterner Sicherheit, ohne dass diese verschleiert wird. Die unzulässige Veränderung der Software eines aktualisierten iPhones ist zum Beispiel gar nicht so einfach. Diese Maßnahmen sind nicht perfekt, aber die Unternehmen investieren weiterhin in sie, weil sie funktionieren.

„Wir haben ernsthafte Bedenken, ob die von der Verordnung auferlegten Technologien wirksam wären: Die Täter wären sich dieser Technologien bewusst und würden auf neue Techniken, Dienste und Plattformen ausweichen, um CSAM-Informationen auszutauschen und sich der Entdeckung zu entziehen."

Dies ist ein fatalistisches Argument, das von Gruppen verwendet wird, die gegen die Einführung von Regeln oder Maßnahmen in einem bestimmten Bereich sind. Würden dieselben Wissenschaftler vorschlagen, dass die Gesellschaft den leichten Zugang zu Schusswaffen nicht einschränken sollte, weil böswillige Personen neue Wege finden könnten, um diese Gegenstände zu erwerben? Mechanismen zur Unterdrückung illegaler Verhaltensweisen sind nie perfekt, und sie haben auch keinen definierten Endpunkt. Sie entwickeln sich ständig weiter und sind so konzipiert, Gefährdungspotenziale durch Schwierigkeiten und Hindernisse innerhalb der Systeme einzuschränken.

Ich spreche aus eigener Erfahrung: Wir erkennen und entfernen täglich Zehntausende Missbrauchsdarstellungen mit Hilfe von Hashing-Technologien. Wir stellen unsere Bemühungen nicht ein, weil die Täter versuchen, sich der Erkennung zu entziehen - wir passen uns der veränderten Bedrohung an und entwickeln Innovationen, so wie es von Technologieunternehmen erwartet werden sollte.

„Es sind eher die Beschwerden der Nutzer als die KI, die in der Praxis zur Erkennung von neuem Missbrauchsmaterial führen."

Nach Angaben von Google verarbeitet die KI-basierte Content Safety API jedes Jahr Millionen von Bildern: https:// protectingchildren.google/tools-for-partners/. Vermutlich verwenden sie und ihre Partner dieses Tool, weil es tatsächlich bisher unbekannte Bilder des sexuellen Missbrauchs aufspürt.

Über den Canadian Centre for Child Protection: Der Canadian Centre for Child Protection (C3P) ist eine nationale Wohltätigkeitsorganisation, die sich für den Schutz aller Kinder einsetzt. Das Ziel der Organisation ist es, den sexuellen Missbrauch und die Ausbeutung von Kindern durch Programme, Dienstleistungen und Ressourcen für kanadische Familien, Erzieher, Kinderhilfsorganisationen, Strafverfolgungsbehörden und andere Parteien zu reduzieren. Die C3P betreibt auch Cybertip.ca, Kanadas Hotline zur Meldung von sexuellem Missbrauch und sexueller Ausbeutung von Kindern im Internet, sowie das Projekt Arachnid, eine Webplattform, die dazu dient, bekannte Bilder von sexuellem Kindesmissbrauch (CSAM) im freien und dunklen Internet aufzuspüren und der Industrie Hinweise zur Entfernung zu geben.