Escaneo del lado del cliente

Qué es y por qué amenaza las comunicaciones privadas confiables

Agosto de 2022

El cifrado es una tecnología diseñada para ayudar a los usuarios de Internet a mantener su información y sus comunicaciones privadas y seguras. El proceso de cifrado codifica la información para que solo pueda ser leída por alguien con la "clave" que descifra la información. El cifrado protege las actividades cotidianas como la banca en línea y las compras. También evita que se roben datos en los casos de infracciones de datos, y garantiza que los mensajes privados conserven su carácter privado. El cifrado también es crucial para proteger las comunicaciones de las autoridades del orden público, el personal militar y, cada vez más, el personal de respuesta ante emergencias.

El cifrado de extremo a extremo (E2E), donde las claves necesarias para descifrar una comunicación cifrada residen solo en los dispositivos que se comunican, proporciona el nivel más alto de seguridad y confianza. Por diseño, solo el destinatario previsto tiene la clave para descifrar el mensaje. El cifrado E2E es una herramienta esencial para garantizar comunicaciones seguras y confidenciales. Agregar el escaneo de mensajes, incluso si es "del lado del cliente", rompe el modelo de cifrado E2E y viola fundamentalmente la confidencialidad que los usuarios esperan.

¿Qué es el escaneo del lado del cliente?

Escaneo del lado del cliente (CSS, Client-side scanning) se refiere en general a los sistemas que escanean el contenido del mensaje, es decir, texto, imágenes, videos y archivos, para encontrar coincidencias o similitudes con una base de datos de contenido objetable antes de que el mensaje se envíe al destinatario previsto. Por ejemplo, su software antivirus puede hacer esto para encontrar programas malignos en su computadora y deshabilitarlos.

Con el avance de los principales proveedores de plataformas hacia la implementación de más cifrado E2E, y el llamado de algunos agentes del orden público a facilitar el acceso al contenido de los mensajes con el fin de ayudar a identificar y evitar el intercambio de contenido objetable,[1]

1 https://www.newamerica.org/oti/press-releases/open-letter-law-enforcement-us-uk-and-australia-weak-encryption- puts-billions-internet-users-risk/

existe la posibilidad de que el escaneo del lado del cliente surja como el mecanismo preferido para abordar el contenido objetable compartido en servicios con cifrado E2E sin romper el cifrado.

Sin embargo, el escaneo del lado del cliente podría comprometer la privacidad y seguridad que los usuarios asumen y de las que se fían. Al hacer que el contenido de los mensajes ya no sea privado entre el remitente y el destinatario, el escaneo del lado del cliente rompe el modelo de confianza E2E. La complejidad que agrega también podría limitar la fiabilidad de un sistema de comunicaciones y posiblemente impedir que los mensajes legítimos lleguen a sus destinos previstos.

Escaneo del lado del cliente para evitar que se comparta contenido objetable

Cuando su intención es evitar que las personas intercambien contenido objetable, el escaneo del lado del cliente generalmente se refiere a una forma en la que el software en los dispositivos del usuario (a menudo llamados "clientes" e inclusive cosas como teléfonos inteligentes, tabletas o computadoras) crea "huellas digitales"[2] funcionalmente únicas del contenido del usuario (llamadas "hashes"). Luego los compara con una base de datos de huellas digitales de contenido cuestionable conocido, tal como software malicioso (malware), imágenes, videos o gráficos.3 Si se encuentra una coincidencia, el software puede evitar que ese archivo se envíe, o notificar a un tercero sobre el intento de envío, a menudo sin que el usuario lo sepa. Los enfoques más nuevos para el escaneo del lado del cliente también buscan contenido objetable nuevo utilizando algoritmos más sofisticados. Esto es difícil y hace que la posibilidad de falsos positivos sea aún más probable.

Cómo funciona el escaneo del lado del cliente

Existen dos métodos básicos de escaneo del lado del cliente en busca de contenido objetable en un servicio de comunicaciones cifradas E2E. Uno realiza la comparación de huellas digitales en el dispositivo del usuario y el otro realiza la comparación en un servidor remoto (el contenido permanece en el dispositivo).

2 Se podría desarrollar un sistema en el que las huellas digitales sean menos únicas, lo que daría lugar a que más elementos de contenido utilicen la misma huella digital. Sin embargo, cuando los falsos positivos pueden dar lugar al uso de recursos serios (como una redada policial), los diseñadores de sistemas de escaneo del lado del cliente se ven incentivados a que las huellas digitales sean lo más únicas posible.

3 El escaneo del lado del cliente es solo una de las formas propuestas para que las autoridades del orden público o las agencias de seguridad tengan acceso a las comunicaciones cifradas de los usuarios. Para obtener más información consulte: https://www.internetsociety.org/resources/doc/2018/encryption-brief/

1. **Comparación que se realiza en el dispositivo del usuario (coincidencia local de huellas digitales)**
   La aplicación en el dispositivo de un usuario (teléfono, tableta o computadora) tiene una base de datos completa actualizada de huellas digitales funcionalmente únicas de contenido de interés conocido. El contenido que el usuario va a cifrar y enviar en un mensaje se convierte en una huella digital usando las mismas técnicas aplicadas a las huellas digitales en la base de datos completa. Si se encuentra una coincidencia, o un algoritmo clasifica el contenido como probablemente objetable, es posible que no se envíe el mensaje y que se notifique a un tercero designado (como las autoridades policiales, las agencias de seguridad nacional o el proveedor de los servicios de filtrado).
2. **Comparación realizada en un servidor remoto**
   Puede haber desafíos importantes con el mantenimiento de una base de datos completa y algoritmos sofisticados que realizan análisis en tiempo real en el dispositivo de un usuario. La alternativa es transmitir las huellas digitales del contenido de un usuario a un servidor donde se realiza una comparación con una base de datos central.

Problemas con el escaneo del lado del cliente para detectar contenido objetable

Cuando la comparación de huellas digitales se realiza en un servidor remoto, este podría permitir que el proveedor de servicios, y cualquier otra persona con la que elija compartir la información, supervise y filtre el contenido que un usuario desea enviar. Se aplican las mismas consideraciones cuando la comparación se lleva a cabo en el dispositivo del usuario, si se notifica a terceros sobre cualquier contenido objetable encontrado. Esto básicamente frustra la finalidad del cifrado E2E. Las comunicaciones con cifrado E2E privadas y seguras entre dos partes, o entre un grupo, deben mantenerse privadas. Si las personas sospechan que se está escaneando su contenido, pueden autocensurarse, cambiar a otro servicio sin escaneo del lado del cliente o usar otro medio de comunicación.

Crea vulnerabilidades que los delincuentes pueden aprovechar: Agregar la funcionalidad de escaneo del lado del cliente aumenta la superficie de ataque al crear formas adicionales de interferir con las comunicaciones manipulando la base de datos de contenido objetable. Los adversarios con la capacidad de agregar huellas digitales a la base de datos y recibir notificaciones cuando se encuentran coincidencias con dichas huellas digitales tendrían una manera de monitorear el contenido seleccionado del usuario antes de que se cifre y envíe. Esto les permitiría rastrear a quién, cuándo y dónde se comunicó cierto contenido. Estas huellas digitales podrían incluir contraseñas comúnmente utilizadas u otra información para permitir ataques tales como los de ingeniería social, extorsión o chantaje. Al aprovechar las funciones de bloqueo de un sistema,

los delincuentes podrían incluso optar por impedir que los usuarios envíen contenido específico. Esto podría estar dirigido a impactar usos legítimos, lo que podría impedir las comunicaciones de las autoridades del orden público, la respuesta de emergencia y el personal de seguridad nacional.

Crea nuevos desafíos técnicos y de proceso: Si las comparaciones se realizan en el dispositivo del usuario, mantener una versión actualizada de la base de datos de referencia completa y los algoritmos en cada dispositivo presenta su propio conjunto de desafíos. Estos incluyen posibles restricciones en materia de procesos (por ejemplo, el proceso de agregar o quitar huellas digitales de contenido a la base de datos, y a quien tiene control o acceso a ella), el ancho de banda necesario para transmitir versiones actualizadas de la base de datos, y la potencia de procesamiento requerida en los dispositivos para realizar la comparación en tiempo real. Otras consideraciones incluyen la posible exposición de la base de datos de referencia al instalarla en el dispositivo del cliente, lo que podría proporcionar a los delincuentes información sobre el sistema de escaneo. Si las comparaciones se realizan en un servidor central, la huella digital del contenido que el usuario intenta enviar será accesible a cualquiera que controle ese servidor central, con independencia de si califica como "objetable" a los ojos de la parte supervisora. Esto da lugar a un nuevo conjunto de problemas en torno a la seguridad y la privacidad de los usuarios, ya que se presenta la posibilidad de exponer detalles de su actividad a cualquier persona con acceso al servidor.

Irregularidad de funciones; esta podría usarse para otras cosas: los mismos métodos implementados con la esperanza de combatir lo peor de lo peor (por ejemplo, explotación infantil o contenido terrorista, los dos propósitos citados con más frecuencia para justificar su uso) también pueden usarse para vigilancia masiva y fines represivos. Un documento de 2021 sobre los riesgos del escaneo del lado del cliente señaló que un sistema CSS podría construirse de manera que le brinde a una agencia la capacidad de escanear de manera preventiva cualquier tipo de contenido en cualquier dispositivo, para cualquier propósito, sin una orden judicial o sospecha. Del mismo modo, las mismas técnicas para prevenir la distribución de material de abuso sexual infantil (CSAM) se pueden usar para hacer cumplir políticas como la censura y la supresión de la disidencia política al evitar que se comparta contenido legítimo o bloquear las comunicaciones entre usuarios (como opositores políticos). Es difícil limitar la base de datos de modo que incluya únicamente huellas digitales de imágenes, videos o URL relacionados con actividades ilícitas (como algunos proponen). Al crear huellas digitales de más contenido para compararlas con las huellas digitales del contenido del usuario o al ampliar el alcance de un algoritmo para clasificar tipos adicionales de contenido del usuario como objetables, quienquiera que controle el sistema puede detectar cualquier contenido de interés. Un sistema de escaneo del lado del cliente podría extenderse para monitorear el contenido de texto de los mensajes que se envían, con implicaciones claras y devastadoras para la libertad de expresión.

Falta de efectividad: los sistemas de comunicaciones encriptadas E2E existen fuera de la jurisdicción de cualquier gobierno. Para evitar que lo atrapen, un delincuente verdaderamente determinado podría dejar de usar servicios que utilizan el escaneo del lado del cliente y comenzar a utilizar otros. Es técnicamente simple para los delincuentes realizar modificaciones en el contenido objetable, cambiando así la huella digital y evitando la detección por parte del sistema de escaneo del lado del cliente.

Conclusión

Detener la difusión de material terrorista y de explotación infantil es una causa importante. Sin embargo, no se puede lograr debilitando la seguridad de las comunicaciones de los usuarios para monitorear potencialmente lo que las personas se dicen entre sí. El escaneo del lado del cliente reduce la seguridad y la privacidad generales para los usuarios respetuosos de la ley, al tiempo que corre el riesgo de no cumplir con su objetivo de cumplimiento de la ley declarado. El cifrado E2E garantiza que miles de millones de usuarios de todo el mundo puedan comunicarse de forma segura y confidencial.[4] Las plataformas principales continúan avanzando hacia su adopción como una forma de conseguir confiabilidad en sus plataformas y servicios.[5] El escaneo del lado del cliente en los servicios de comunicaciones con cifrado E2E no es una solución para filtrar contenido objetable. Tampoco lo es ningún otro método que debilita el núcleo de las comunicaciones privadas y fiables de las que todos dependemos.

Referencias

Internet Society, junio de 2018. Encryption Brief (Resumen sobre cifrado). https://www.internetsociety.org/resources/doc/2018/encryption- brief/

Matthew Green, diciembre de 2019. ¿Pueden los sistemas con cifrado de extremo a extremo detectar imágenes de abuso sexual infantil?

https://blog.cryptographyengineering.com/2019/12/08/on-client-side-media-scanning/

Electronic Frontier Foundation, noviembre de 2019. Why Adding Client-Side Scanning Breaks End-To-End Encryption (Por qué al agregar el escaneo del lado del cliente se rompe el cifrado de extremo a extremo). https://www.eff.org/deeplinks/2019/11/why-adding-client-side-scanning-breaks-end-end- encryption

4 https://telegram.org/blog/200-million y https://www.newsweek.com/whatsapp-facebook-passes-two-billion-users-pledges-encryption-support-1486993

5 https://www.facebook.com/notes/2420600258234172/

Centro para la democracia y la tecnología (CDT), 2021. Moderación de contenido en sistemas cifrados: https://cdt.org/insights/outside-looking-in-approaches-to-content-moderation-in-end-to-end-encrypted-systems/

Hal Abelson, Ross Anderson, Steven M. Bellovin, Josh Benaloh, Matt Blaze, Jon Callas, Whitfield Diffie, Susan Landau, Peter G. Neumann, Ronald L. Rivest, Jeffrey I. Schiller, Bruce Schneier, Vanessa Teague, Carmela Troncoso, octubre de 2021. Errores en nuestros bolsillos: los riesgos del escaneo del lado del cliente. https://www.cs.columbia.edu/~smb/papers/bugs21.pdf