Un nuevo análisis de los ataques de huellas dactilares de sitios web (WF) dirigidos al navegador web Tor, reveló que es posible que un adversario obtenga un sitio web frecuentado por una víctima, pero solo en escenarios en los que el actor de la amenaza está interesado en un subconjunto específico de los sitios web visitados por los usuarios.

> *«Aunque los ataques pueden superar el 95% de precisión cuando se monitorea un pequeño conjunto de cinco sitios web populares, los ataques indiscriminados (no dirigidos) contra conjuntos de 25 y 100 sitios web no superan una precisión del 80% y 60%, respectivamente»,* dijeron los investigadores Giovanni Cherubin, Rob Jansen y Carmela Troncoso.

El navegador Tor ofrece *«comunicación no vinculable»* a sus usuarios al enrutar el tráfico de Internet a través de una red superpuesta, que consta de más de seis mil relés, con el objetivo de anonimizar la ubicación de origen y el uso de terceros que realizan vigilancia de red o análisis de tráfico. Lo logra construyendo un circuito que atraviesa a través de un relé de entrada, medio y salida, antes de reenviar las solicitudes a las direcciones IP de destino.

Además de eso, las solicitudes se cifran una vez para cada relé para dificultar aún más el análisis y evitar la fuga de información. Si bien los propios clientes Tor no son anónimos con respecto a sus relés de entrada, debido a que el tráfico está encriptado y las solicitudes saltan a través de múltiples saltos, los relés de entrada no pueden identificar el destino de los clientes, al igual que los nodos de salida no pueden discernir un cliente para la misma razón.

Los ataques de huellas dactilares de sitios web en Tor tienen como objetivo romper estas protecciones de anonimato y permitir que un adversario observe los patrones de tráfico encriptados entre una víctima y la red Tor para predecir el sitio web visitado por la víctima. El modelo de amenaza ideado por los académicos presupone que un atacante ejecuta un nodo de salida, para capturar la diversidad del tráfico generado por los usuarios reales, que luego se utiliza como fuente para recopilar los rastros de tráfico de Tor y diseñar un modelo de clasificación basado en el aprendizaje automático encima de la información recopilada para

inferir las visitas al sitio web de los usuarios.

El modelo del adversario implica una *«fase de entretenimiento en línea que utiliza observaciones de tráfico Tor genuino recopiladas de un relé de salida para actualizar continuamente el modelo de clasificación a lo largo del tiempo»,* explicaron los investigadores, que ejecutaron los relés de entrada y salida durante una semana en julio de 2020 usando una versión personalizada de Tor v0.4.3.5 para extraer la información de salida relevante.

Para mitigar cualquier inquietud ética y de privacidad que surja del estudio, los autores del artículo enfatizaron las precauciones de seguridad incorporadas para evitar la filtración de sitios web sensibles que los usuarios pueden visitar a través del navegador Tor.

> *«Los resultados de nuestra evaluación del mundo real demuestran que los ataques WF solo pueden tener éxito en la naturaleza si el adversario tiene como objetivo identificar sitios web dentro de un grupo pequeño. En otras palabras, los adversarios que tengan como objetivo monitorear generalmente las visitas al sitio web de los usuarios fallarán, pero los adversarios enfocados que se dirijan a una configuración de cliente y un sitio web en particular pueden tener éxito»,* concluyeron los investigadores.