Investigadores de seguridad cibernética encontraron una nueva técnica que hace que un ataque de canal lateral basado en el tiempo remoto sea más efectivo, independientemente de la congestión de red entre el adversario y el servidor objetivo.

Los ataques de temporización remota que funcionan por medio de una conexión de red se ven afectados principalmente por variaciones en el tiempo de transmisión de la red, que a su vez, depende de la carga de la conexión de red en un determinado momento.

Pero debido a que medir el tiempo necesario para ejecutar algoritmos criptográficos es crucial para llevar a cabo un ataque de temporización, y en consecuencia, filtrar información, la fluctuación en la ruta de red desde el atacante al servidor puede hacer que sea poco práctico explotar exitosamente los canales laterales de temporización que dependen de una pequeña diferencia en el tiempo de ejecución.

El nuevo método, llamado Timeless Timing Attacks (TTAs), por investigadores de DistriNet Research Group y la Universidad de Nueva York, Abu Dhabi, aprovecha la multiplexación de protocolos de red y la ejecución concurrente por aplicaciones, haciendo que los ataques sean inmunes a las condiciones de la red.

> *«Estos ataques de sincronización basados en la concurrencia infieren una diferencia de sincronización relativa al analizar el orden en que se devuelven las respuestas y, por lo tanto, no se basan en ninguna información de sincronización absoluta»,* dijeron los investigadores.

Multiplexación de solicitudes de HTTP/2 para reducir la fluctuación de fase

A diferencia de los típicos ataques basados en el tiempo, en los que los tiempos de ejecución se miden de forma independiente y secuencial, esta técnica intenta extraer información del orden y la diferencia del tiempo relativa entre dos solicitudes ejecutadas simultáneamente

sin depender de ninguna información de tiempo.

Para hacerlo, un atacante inicia un par de solicitudes HTTP/2 al servidor de la víctima directamente o utilizando un sitio cruzado, como un anuncio malicioso o engañando a la víctima para que visite una página web controlada por el atacante, para lanzar solicitudes al servidor a través de código JavaScript.

El servidor devuelve un resultado que contiene la diferencia en el tiempo de respuesta entre la segunda solicitud y la primera. El TTA funciona teniendo en cuenta si la diferencia es positiva o negativa, donde el positivo indica que el tiempo de procesamiento de la primera solicitud lleva menos tiempo que el procesamiento de la segunda solicitud.

> *«En los servidores web alojados por medio de HTTP/2, encontramos que una diferencia de tiempo tan pequeña como 100 ns se puede inferir con precisión del orden de respuesta de aproximadamente 40,000 pares de solicitudes»,* dijeron los investigadores.
>
> *«La diferencia de tiempo más pequeña que pudimos observar en un ataque de tiempo tradicional en Internet, fue de 10 µs, 100 veces mayor que nuestro ataque basado en la concurrencia».*

Una limitación de este enfoque es que los ataques dirigidos a servidores que usan HTTP/1.1 no pueden explotar el protocolo para unir múltiples solicitudes en un solo paquete de red, lo que requiere que se realice un ataque de sincronización concurrente utilizando múltiples conexiones en lugar de enviar todas las solicitudes por medio de la misma conexión.

Esto se debe al uso del bloqueo de cabecera de línea (HOL) de HTTP/1.1, que hace que todas las solicitudes a través de la misma conexión se manejen de forma secuencial, mientras que

HTTP/2 resuleve el problema mediante la multiplexación de solicitudes.

En la actualidad, el 37.46% de todos los sitios web de escritorio se sirven a través de HTTP/2, número que aumenta al 54.04% para los sitios que admiten HTTPS. Aunque esto significa que se trata de una gran cantidad de sitios web susceptibles a los TTA, los investigadores destacan que muchos de los sitios dependen de las redes de entrega de contenido (CDN), como Cloudflare, que aún utiliza HTTP/1.1 para las conexions entre el CDN y el sitio de origen.

Servicios Tor Onion y Wi-Fi EAP-PWD vulnerables

Los investigadores descubrieron que los ataques de temporización basados en la concurrencia también se pueden implementar contra los servicios Onion Tor, incluidos lo que solo admiten HTTP/1.1, lo que permitiría a un atacante crear dos conexiones Tor a un servicio onion en particular, y luego enviar simultáneamente una solicitud en cada una de las conexiones para medir una diferencia de tiempo de un micro segundo.

Además, el método de autenticación EAP-PWD, que utiliza una contraseña compartida entre el servidor y el solicitante cuando se conecta a redes WiFi, se vuelve vulnerable a los ataques de diccionario al explotar una fuga de tiempo en el protocolo de enlace Dragonfly para revelar la información de la contraseña.

Aunque los ataques de tiempo pueden ser contrarrestados al asegurar la ejecución en tiempo constante, no es tan fácil como suena, especialmente para aplicaciones que dependen de componentes de terceros. De forma alternativa, los investigadores sugieren agregar un retraso aleatorio a las solicitudes entrantes y asegurarse de que las diferentes solicitudes no se combinen en un solo paquete.

Esta no es la primera vez que se utilizan ataques de temporización remota para filtrar información confidencial. Los investigadores han demostrado que es posible explotar los canales laterales de la caché para detectar las contraseñas SSH de la caché de la CPU Intel (NetCAT) e incluso, lograr una ejecución especulativa similar a Spectre a través de una

conexión de red (NetSpectre).

> *«Dado que los ataques de NetSpectre apuntan a aplicaciones por encima de la capa de red, un atacante podría, en teoría, aprovechar nuestros ataques de tiempo basados en la concurrencia para mejorar la precisión del tiempo»,* dijeron los investigadores.

Estos hallazgos se presentarán en el Simposio de Seguridad USENIX a finales de 2020. Los investigadores también publicaron una herramienta basada en Python para probar los servidores HTTP/2 en busca de vulnerabilidades de TTA.