Google informó haber identificado una vulnerabilidad de día cero en el motor de base de datos de código abierto SQLite, empleando su marco de trabajo asistido por un modelo de lenguaje extenso (LLM) denominado Big Sleep (anteriormente conocido como Project Naptime).

El gigante tecnológico describió este avance como la *«primera vulnerabilidad en el mundo real»* detectada mediante el uso de un agente de inteligencia artificial (IA).

> *«Creemos que este es el primer ejemplo público de un agente de IA que descubre un problema de seguridad de memoria explotable, hasta ahora desconocido, en un software de uso extendido en el mundo real»,* afirmó el equipo de Big Sleep en una entrada de blog.

La vulnerabilidad en cuestión es un desbordamiento de pila en SQLite, que ocurre cuando un software hace referencia a una ubicación de memoria antes del inicio del búfer, causando un fallo o permitiendo la ejecución de código arbitrario.

> *«Este problema suele surgir cuando un puntero o su índice se reduce a una posición anterior al búfer, cuando las operaciones de punteros resultan en una posición fuera de los límites válidos de la memoria, o cuando se utiliza un índice negativo»*, de acuerdo con la descripción de la clasificación de errores del Common Weakness Enumeration (CWE).

Tras una divulgación responsable, la vulnerabilidad fue solucionada a principios de octubre de 2024. Cabe destacar que esta falla fue descubierta en una rama de desarrollo de la biblioteca, lo que significa que fue identificada antes de formar parte de una versión oficial.

Project Naptime fue detallado inicialmente por Google en junio de 2024 como un marco técnico para mejorar los métodos de descubrimiento automatizado de vulnerabilidades.

Desde entonces, ha evolucionado en Big Sleep como parte de una colaboración más amplia entre Google Project Zero y Google DeepMind.

El propósito de Big Sleep es utilizar un agente de IA que emule el comportamiento humano al identificar y demostrar vulnerabilidades de seguridad, aprovechando las capacidades de comprensión de código y razonamiento de un LLM.

Esto implica el uso de un conjunto de herramientas especializadas que permite al agente explorar el código, ejecutar scripts de Python en un entorno aislado para generar entradas de fuzzing, depurar el programa y observar los resultados.

> *«Pensamos que este trabajo tiene un enorme potencial defensivo. Identificar vulnerabilidades en el software antes de su lanzamiento significa que los atacantes no tienen oportunidad de aprovecharlas: las vulnerabilidades se corrigen antes de que los atacantes puedan explotarlas»,* señaló Google.

No obstante, la empresa también subrayó que estos resultados siguen siendo experimentales, y agregó que *«la posición del equipo de Big Sleep es que, en la actualidad, es probable que un fuzzer específico para el objetivo sea igual o más efectivo (en la detección de vulnerabilidades)».*