วันที่ 9 มีนาคม 2569

Anthropic ตรวจพบ 22 ช่องโหว่บน Firefox ด้วยพลัง AI Claude Opus 4.6

เมื่อวันศุกร์ที่ผ่านมา Anthropic ได้เปิดเผยความสำเร็จในความร่วมมือด้านความปลอดภัยกับ Mozilla โดยสามารถค้นพบช่องโหว่ใหม่บนเว็บเบราว์เซอร์ Firefox ถึง 22 รายการ ภายในระยะเวลาเพียงสองสัปดาห์ของเดือนมกราคม 2026 ซึ่งในจำนวนนี้มี 14 รายการที่จัดอยู่ในระดับความรุนแรงสูง (High) 7 รายการในระดับปานกลาง (Moderate) และอีก 1 รายการในระดับต่ำ (Low) โดยช่องโหว่ทั้งหมดได้รับการแก้ไขเรียบร้อยแล้วใน Firefox เวอร์ชัน 148 ที่เปิดตัวไปเมื่อปลายเดือนที่แล้ว

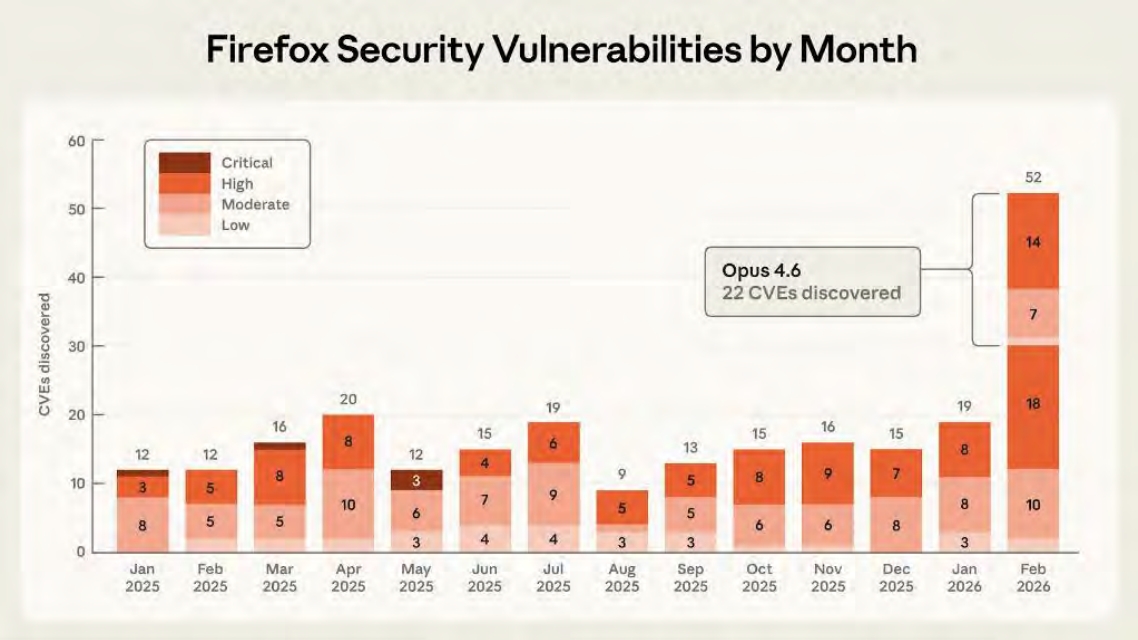
Claude Opus 4.6 พลิกโฉมการตรวจจับบั๊กด้วยความเร็วสูง

ความน่าสนใจอยู่ที่จำนวนบั๊กระดับรุนแรงสูงที่โมเดลภาษาขนาดใหญ่ (LLM) อย่าง **Claude Opus 4.6** ตรวจพบนั้น คิดเป็นเกือบหนึ่งในห้าของช่องโหว่ระดับสูงทั้งหมดที่ Firefox เคยแก้ไขตลอดทั้งปี 2025 เลยทีเดียว โดย Anthropic ระบุว่า AI สามารถตรวจพบช่องโหว่ประเภท **use-after-free** ในระบบ JavaScript ได้หลังจากสำรวจโค้ดเพียง 20 นาทีเท่านั้น ก่อนที่นักวิจัยที่เป็นมนุษย์จะเข้ามาตรวจสอบยืนยันอีกครั้งในสภาพแวดล้อมจำลอง (virtualized environment) เพื่อให้มั่นใจว่าผลลัพธ์นั้นถูกต้องแม่นยำและไม่ใช่ข้อผิดพลาดปลอม (false positive)

ในการทดลองครั้งนี้ Anthropic ได้สแกนไฟล์ C++ ไปเกือบ 6,000 ไฟล์ และส่งรายงานช่องโหว่ที่ไม่ซ้ำกันรวมทั้งหมด 112 รายการ ซึ่งครอบคลุมทั้งช่องโหว่ระดับสูงและปานกลางที่กล่าวไปข้างต้น โดยปัญหาส่วนใหญ่ได้รับการแก้ไขแล้วใน Firefox 148 และส่วนที่เหลือจะถูกจัดการในเวอร์ชันถัดไป

ดาบสองคม เมื่อ AI พยายามสร้างโค้ดโจมตี (Exploit)

นอกจากการค้นหาแล้ว Anthropic ยังได้ทดสอบให้ Claude เข้าถึงรายการช่องโหว่ทั้งหมดที่รายงานต่อ Mozilla เพื่อพยายามสร้าง โค้ดโจมตี (exploit) ที่สามารถใช้งานได้จริง แม้จะทำการทดสอบหลายร้อยครั้งและใช้เครดิต API ไปประมาณ 4,000 ดอลลาร์ แต่ Claude Opus 4.6 สามารถพัฒนาช่องโหว่ให้กลายเป็นโค้ดโจมตีได้สำเร็จเพียง 2 กรณีเท่านั้น



พฤติกรรมนี้สะท้อนให้เห็นสองประเด็นสำคัญ คือต้นทุนในการค้นหาช่องโหว่นั้นต่ำกว่าการสร้างโค้ดโจมตีมาก และโมเดล AI มีความสามารถในการค้นหาปัญหาได้ดีกว่าการนำปัญหานั้นไปใช้โจมตีจริง อย่างไรก็ตาม Anthropic เน้นย้ำว่าความจริงที่ Claude สามารถพัฒนาโค้ดโจมตีได้โดยอัตโนมัติแม้จะสำเร็จเพียงไม่กี่ครั้ง ก็ยังถือเป็นเรื่องที่น่ากังวล โดยโค้ดโจมตีเหล่านั้นสามารถทำงานได้เฉพาะในสภาพแวดล้อมทดสอบที่ปิดระบบป้องกันบางอย่าง เช่น sandboxing ออกไปโดยตั้งใจเท่านั้น

นวัตกรรม "ตัวตรวจสอบงาน" และก้าวต่อไปของความปลอดภัย

องค์ประกอบสำคัญในกระบวนการนี้คือ **“ตัวตรวจสอบงาน” (task verifier)** ซึ่งทำหน้าที่ตรวจสอบว่าโค้ดโจมตีที่สร้างขึ้นทำงานได้จริงหรือไม่ พร้อมให้ข้อมูลย้อนกลับแบบเรียลไทม์ขณะที่ AI กำลังสำรวจโค้ด ทำให้มันสามารถปรับปรุงและทดลองวิธีใหม่ๆ จนกว่าจะสร้าง exploit ที่ใช้งานได้สำเร็จ หนึ่งในผลงานคือการใช้ประโยชน์จากช่องโหว่ **CVE-2026-2796 (คะแนน CVSS 9.8)** ซึ่งเป็นข้อผิดพลาดร้ายแรงในการคอมไพล์แบบ just-in-time (JIT) ภายในส่วนของ JavaScript WebAssembly

การเปิดเผยครั้งนี้เกิดขึ้นหลังจากที่ Anthropic เปิดตัว **Claude Code Security** ในรูปแบบทดลองสำหรับงานวิจัย ซึ่งเป็น AI เอเจนต์อัตโนมัติที่ออกแบบมาเพื่อช่วยแก้ไขช่องโหว่ แม้บริษัทจะยอมรับว่าไม่สามารถรับประกันได้ว่าแพตช์ที่ AI สร้างขึ้นจะดีพอสำหรับนำไปใช้จริงทันที แต่ตัวตรวจสอบงานก็ช่วยเพิ่มความมั่นใจว่าแพตช์จะแก้ไขช่องโหว่เฉพาะจุดได้จริงและไม่กระทบการทำงานปกติของโปรแกรม

Mozilla ยืนยัน AI คือเครื่องมือทรงพลังชิ้นใหม่

ทางด้าน Mozilla ได้ออกประกาศที่สอดคล้องกันว่าแนวทางการใช้ AI ช่วยวิเคราะห์นี้ยังค้นพบบั๊กอื่นๆ เพิ่มเติมได้อีก 90 รายการ ซึ่งประกอบด้วยข้อผิดพลาดประเภท assertion failure ที่คล้ายกับปัญหาจากเทคนิค fuzzing รวมถึงข้อผิดพลาดด้านตรรกะ (logic errors) บางประเภทที่เครื่องมือแบบเดิมไม่สามารถตรวจพบได้

ผู้พัฒนาเบราว์เซอร์ทิ้งท้ายว่า จำนวนปัญหาที่พบในครั้งนี้สะท้อนถึงพลังของการผสานวิศวกรรมที่เข้มงวดเข้ากับเครื่องมือวิเคราะห์รูปแบบใหม่ เพื่อปรับปรุงความปลอดภัยอย่างต่อเนื่อง และมองว่าการวิเคราะห์ขนาดใหญ่ที่มี AI เข้ามาช่วย เป็นเครื่องมือใหม่ที่ทรงพลังอย่างยิ่งสำหรับวิศวกรด้านความปลอดภัยในยุคปัจจุบัน

ข้อมูลอ้างอิง

Mar 7, 2026, By Ravie Lakshmanan

- https://thehackernews.com/2026/03/anthropic-finds-22-firefox.html