วันที่ 10 สิงหาคม 2569

Claude Code และ Gemini CLI มีช่องโหว่ที่ทำให้ GitHub Issue สามารถเข้าถึงความลับของ CI Workflow ได้

GitHub Issue ที่ถูกเปิดโดยบัญชีซึ่งไม่มีสิทธิ์ใดๆ ใน Repository ก็เพียงพอที่จะทำให้สามารถสั่งรันโค้ดบน CI Runner ที่อยู่เบื้องหลัง Repository ของ Coding Agent ของ Anthropic และ Google ได้ ส่วนในระบบของ OpenAI ช่องโหว่นี้เพียงพอที่จะเข้าควบคุมการทำงานของ Agent ในรอบถัดไปได้

Novee Security ทดสอบการโจมตีดังกล่าวกับ Agent ของผู้ให้บริการแต่ละราย โดยใช้การตั้งค่าตามค่าเริ่มต้นที่ผู้ให้บริการจัดส่งมาให้ และได้นำเสนอผลงานดังกล่าวที่งาน Black Hat USA เมื่อวันที่ 5 สิงหาคม โดยพบช่องโหว่ที่ได้รับหมายเลข CVE จำนวน 2 รายการ และทั้งสองรายการได้รับการแก้ไขแล้ว

Gemini CLI พบช่องโหว่ที่มีความรุนแรงสูงที่สุดในสองรายการ โดย CVE-2026-12537 (CVSS v4: 10.0) เป็นช่องโหว่ OS Command Injection ในส่วนที่ใช้เปิดใช้งาน Container ซึ่งสามารถถูกโจมตีผ่านไฟล์ .gemini/.env ที่ถูกสร้างขึ้นเป็นพิเศษ ทำให้ผู้โจมตีที่ไม่มีสิทธิ์สามารถสั่งรันโค้ดบน Host ของแพลตฟอร์ม CI แบบ Headless ได้ ก่อนที่ Sandbox จะเริ่มทำงาน ช่องโหว่นี้ได้รับการแก้ไขแล้วใน Gemini CLI 0.39.1 และ run-gemini-cli 0.1.22

ใน Claude Code ช่องโหว่ CVE-2026-54316 ทำให้ระบบนับจำนวนการดาวน์โหลดแบบสาธารณะของ Hugging Face ถูกนำมาใช้เป็นช่องทางขโมยข้อมูล โดยสามารถส่ง API Key ออกไปทีละตัวอักษร และได้รับการแก้ไขแล้วในเวอร์ชัน 2.1.163 Claude Code ทุกเวอร์ชันตั้งแต่ 0.2.54 จนถึง 2.1.163 ได้รับผลกระทบ โดย Anthropic ระบุว่าการโจมตีจำเป็นต้องทำให้ข้อมูลที่ไม่น่าเชื่อถือถูกนำเข้ามาอยู่ใน Context ของ Claude Code

ส่วนช่องโหว่ที่พบใน Codex ไม่ได้นำไปสู่การออกแพตช์ในระดับเวอร์ชันของผลิตภัณฑ์ และไม่มีการกำหนดหมายเลข CVE โดย Novee ระบุว่า OpenAI มองว่า Sandbox ของตนทำงานตรงตามที่มีการระบุไว้ในเอกสาร ดังนั้นผู้ใช้งานควรอัปเดต Gemini CLI เป็น 0.39.1, run-gemini-cli เป็น 0.1.22 และ Claude Code เป็น 2.1.163 รวมถึงตรวจสอบ Workflow ที่ผู้ใช้งานภายนอกสามารถเรียกใช้งานได้

ช่องโหว่ที่ทำให้สามารถรันโค้ดบน Host ของ Gemini ไม่จำเป็นต้องอาศัยการหลอกให้โมเดลทำบางสิ่งโดยเฉพาะ จุดผิดพลาดที่พบในทั้งสามระบบมีลักษณะคล้ายกัน โดยเกิดขึ้นในส่วนที่เรียกว่า Harness ซึ่งเป็นโค้ดที่ทำหน้าที่เชื่อมระหว่างโมเดลกับระบบจริง และเป็นผู้ตัดสินใจว่าอะไรจะถูกนำไปประมวลผลจริง โดยส่วนหนึ่งของระบบกำหนดค่าหนึ่งว่าเป็นค่าที่ปลอดภัย แต่ส่วนถัดมากลับนำค่านั้นไปใช้งานด้วยสิทธิ์ที่สูงกว่า

Novee พบว่า Command Validator ของ Claude Code จะตัดข้อความที่อยู่ภายใน Single Quote ออกก่อนที่จะดำเนินการตรวจสอบทั้ง 23 รายการ ซึ่งถือเป็นพฤติกรรมที่ถูกต้องสำหรับ Bash แต่ทำให้ Payload ที่ถูกใส่ไว้ในค่าของ git push --receive-pack ซึ่งเป็น Flag ที่ Git ใช้ สามารถส่งไปถึง Runner ได้โดยไม่ถูกแก้ไข การโจมตีในรูปแบบนี้ไม่มีหมายเลข CVE และยังไม่มีการระบุเวอร์ชันที่แก้ไขอย่างเป็นทางการ

Gemini CLI จะตรวจสอบรายการเครื่องมือที่ได้รับอนุญาต หรือ Tool Allowlist เฉพาะในตอนที่มีการลงทะเบียนเครื่องมือเท่านั้น แต่เมื่อถึงเวลาที่มีการใช้งานจริงกลับไม่มีการตรวจสอบเพิ่มเติม และเมื่อใช้ตัวเลือก --yolo ทุกคำสั่งที่โมเดลร้องขอจะได้รับการอนุมัติโดยอัตโนมัติ Google ได้แก้ไขทั้งปัญหานี้และช่องโหว่ในส่วนเปิดใช้งาน Container ภายใต้คำแนะนำด้านความปลอดภัยฉบับเดียวกัน โดยระบุว่าแพตช์ดังกล่าว "affects all Gemini CLI GitHub Actions."

คำแนะนำดังกล่าวยังไม่มีการระบุหมายเลข CVE เช่นกัน แต่ Google Cloud ซึ่งเป็นหน่วยงาน CNA ได้เผยแพร่หมายเลขดังกล่าวแยกต่างหากและเชื่อมโยงกลับมายังคำแนะนำฉบับนี้ โดย Anthropic ประเมินความรุนแรงของช่องโหว่ใน Claude Code อยู่ที่ระดับ Moderate หรือ CVSS v4 6.0 ขณะที่ NVD ให้คะแนน CVSS v3.1 อยู่ที่ 9.1 ทั้งนี้ NVD ยังไม่ได้ให้คะแนนช่องโหว่นี้ตาม CVSS v4 ดังนั้นคะแนนทั้งสองจึงไม่สามารถนำมาเปรียบเทียบกันโดยตรงได้

ช่องโหว่ของ Codex เป็นกรณีที่ไม่มีเวอร์ชันใหม่ให้ติดตั้งเพื่อแก้ไข Novee พบว่า Repository openai/codex เรียกใช้ Codex สองรอบภายใน Job เดียวกัน โดยทั้งสองรอบใช้ Checkout เดียวกัน ทำให้ Codex รอบแรกสามารถเขียนไฟล์ AGENTS.md ซึ่งเป็นไฟล์ที่ Codex รอบที่สองจะโหลดขึ้นมาใช้เป็นคำสั่งของตัวเองได้ การตรวจสอบ JSON ระหว่างการทำงานทั้งสองรอบที่ล้มเหลว คือสิ่งที่ทำให้ Codex รอบที่สองถูกเรียกใช้งาน

Workflow ปัจจุบันของ OpenAI แยกการทำงานทั้งสองรอบออกเป็นคนละ Job และเรียกใช้ Codex ด้วยตัวเลือก drop-sudo พร้อม Sandbox แบบ Read-only ขณะเดียวกันคำแนะนำของ OpenAI ในปัจจุบันระบุว่าไฟล์คำสั่งภายใน Repository ควรถูกมองว่าเป็นส่วนหนึ่งของ "untrusted input surface" หรือข้อมูลที่ไม่ควรเชื่อถือ และแนะนำให้เรียกใช้ Codex เป็นขั้นตอนสุดท้ายของ Job พร้อมเตือนว่า หากไม่ทำเช่นนั้น Codex อาจทิ้งไฟล์บางอย่างไว้ให้ขั้นตอนที่มีสิทธิ์สูงกว่าซึ่งทำงานต่อจากนั้นนำไปใช้งาน

อย่างไรก็ตาม การเปลี่ยนแปลงเหล่านี้ไม่ได้แสดงให้เห็นว่า Codex เองถูกปรับให้จัดการกับไฟล์คำสั่งที่สามารถเขียนแก้ไขได้แตกต่างจากเดิม สิ่งที่แหล่งข้อมูลยืนยันได้คือ OpenAI ได้แก้ไข Workflow ในระดับ Repository และปรับปรุงเอกสารคำแนะนำ

ข้อมูลของ CISA สำหรับ CVE ทั้งสองรายการของ Gemini และ Claude Code ระบุว่ายังไม่พบการนำช่องโหว่ไปใช้โจมตีจริง และ The Hacker News ยืนยันเมื่อวันที่ 7 สิงหาคมว่า ทั้งสองรายการยังไม่ปรากฏอยู่ใน Known Exploited Vulnerabilities Catalog ของหน่วยงาน นอกจากนี้ยังพบ Repository สาธารณะบน GitHub ที่ระบุว่าตนเองเป็น Lab สำหรับจำลองช่องโหว่ของ Claude Code ซึ่งถูกเผยแพร่มาตั้งแต่วันที่ 18 มิถุนายน อย่างไรก็ตาม จากแหล่งข้อมูลที่ตรวจสอบ ยังไม่พบหลักฐานว่าการโจมตีรูปแบบใดรูปแบบหนึ่งถูกนำไปใช้กับเป้าหมายจริง

เหตุการณ์ดังกล่าวเกิดขึ้นในช่วงเวลาเดียวกับที่ Pillar Security รายงานเมื่อวันที่ 4 สิงหาคมว่า ผู้ดำเนินการของ Worm ชื่อ ChainDrop ที่แพร่กระจายผ่าน npm ได้ฝัง Hook ชื่อ Claude Code SessionStart และ Task ชื่อ VS Code folderOpen ไว้ใน Repository ที่ถูกบุกรุก โดยโค้ดดังกล่าวจะทำงานทันทีเมื่อผู้พัฒนาเปิด Workspace แทนที่จะรอให้มีการติดตั้งแพ็กเกจ

ข้อมูลอ้างอิง

Aug 07, 2026, By Ravie Lakshmanan

- https://thehackernews.com/2026/08/claude-code-and-gemini-cli-flaws-let.html