reversing everything with claude code

Zane St. John

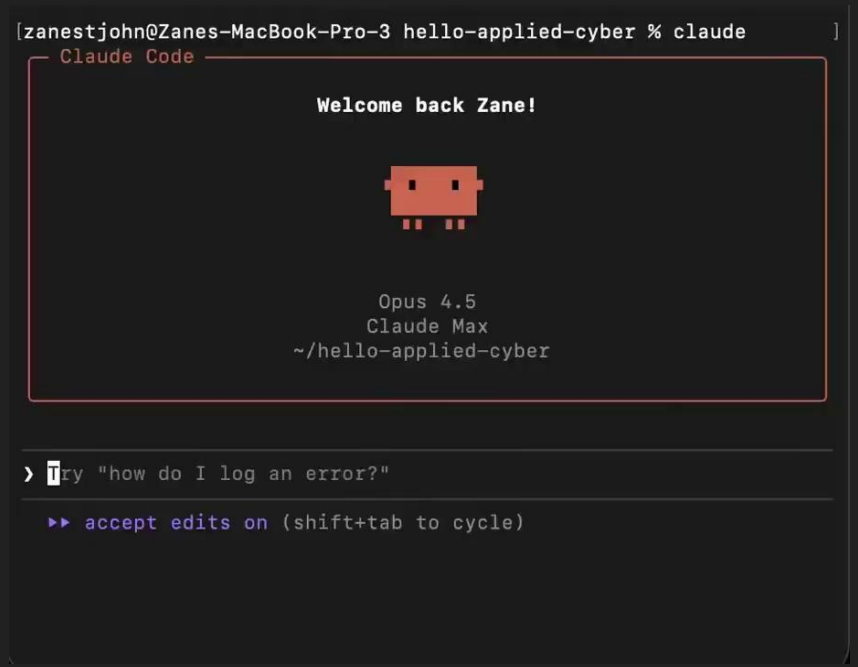
about me

- tl;dr I like breaking things (responsibly)
- Class of ’27, SymSys
- Creative technologist with an interest in cybersecurity
- Worked in a variety of spaces including web3, AI/RAG for gov’t, experiential design
- I also know some Applied Cyber people
    - Former participant in Stanford Bug Bounty (iykyk)
    - Love you all fr

claude code

- Powerful agentic orchestration layer built on top of Claude models
- Built-in prompts and tools expose command-line capabilities, web search, data connectors (MCP)
- Highly capable beyond standard SWE: writing & analysis, data processing, **reverse engineering?**

*I will sometimes abbreviate it as **CC**.*



why?

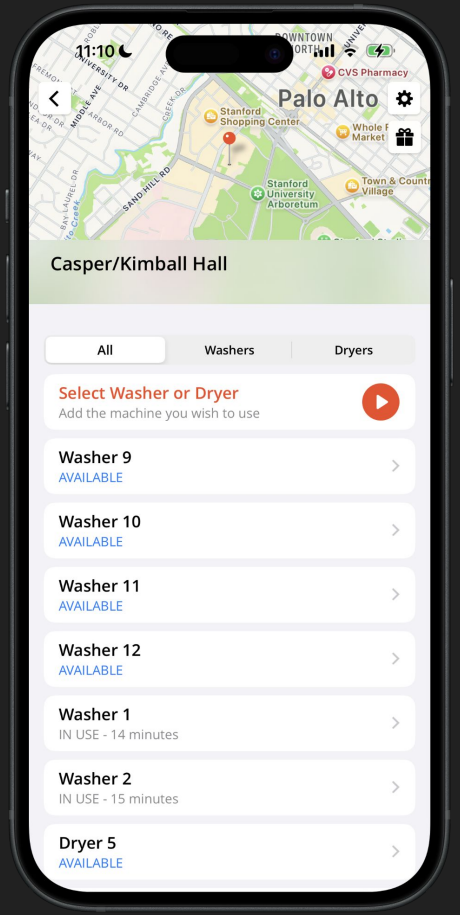
why?

- Reduce RE fatigue → LLMs do the “dirty work”
  - Static analysis without the stress?
- Agentic tooling → test protocols and APIs without human intervention
  - Sub-agents → potentially test multiple strategies in parallel
- Reduce man-hours → save time and money
  - (CC starts at $20/mo, scaling up to $100 or $200 plans for more extensive use)

warmup: a washing machine

motivation

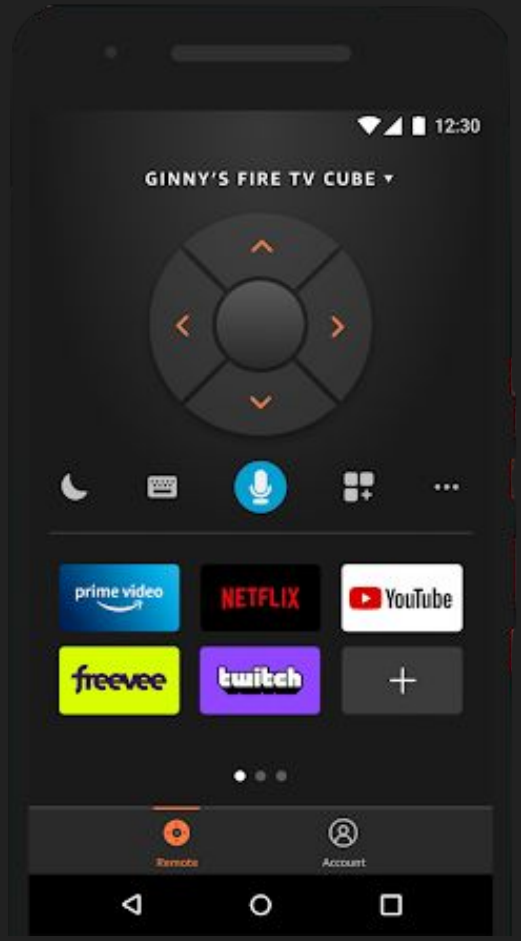
- Washing machines around campus can be monitored via **Speed Queen** app
- However, app does not show how long a machine has been done
- In a crowded dorm like mine (Kimball), people throw people's laundry out so theirs can go in
- **Goal: Find out which machines have been sitting the longest**



easy: a fire tv

motivation

- Fire TV allows local control over network via proprietary mobile app
- API is undocumented but extremely useful to open-source community
- Potential use in smart home automation, control systems, and FOSS apps



process

- Open Fire TV app on mitmproxy-proxied phone
- Click around in the Fire TV remote to record various request types
- Save the output of this session to a `.mitm` dump file
- **Ask Claude Code to:**
  - **Read this dump, infer API, and write documentation**
  - **Write a library & demo program in Python**

outcome

- **Protocol documentation** → API docs for an otherwise undocumented API
- **Working library and demo** → control Fire TV programmatically

- My application: Used Claude-generated protocol docs to develop Fire TV driver for home theater
    - (Yes, even the driver was written by Claude Code!)

hard: a printer

motivation

- Kodak STEP printer and similar printers available for (relatively) low price
- Most of these printers require proprietary app to send photos over Bluetooth
- No documented API, few efforts to reverse-engineer
- **Goal: Print programmatically and without using the app**

small problem

- My laptop runs macOS, which has poor support for RFCOMM, the Bluetooth protocol used to communicate with the printer
- This creates two challenges:
  1. CC can't write code that allows my computer to print directly
  2. CC agent can't communicate directly with the printer to investigate its capabilities
- **Solution: ESP32 as Proxy**

hard(er?): a projector

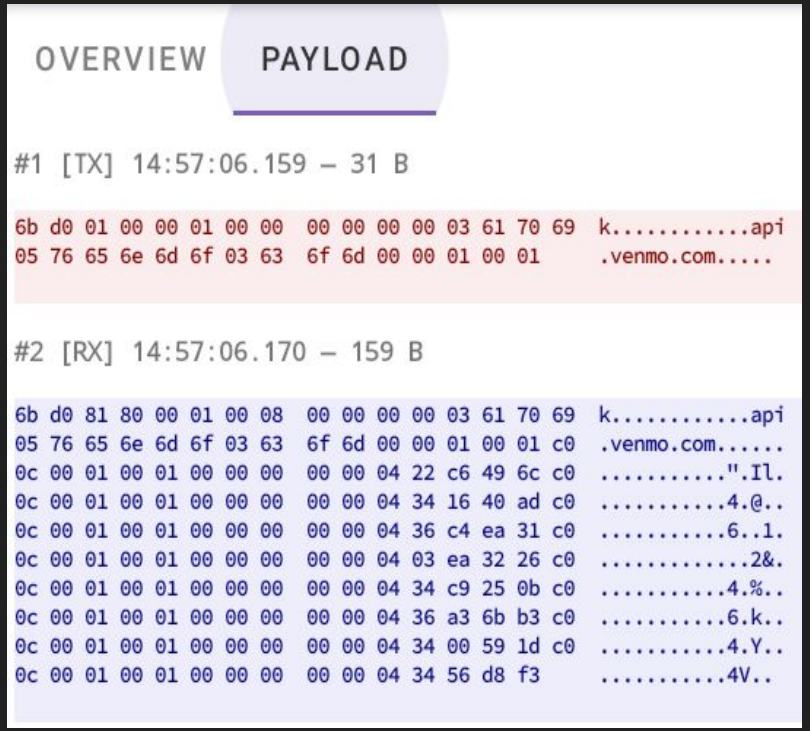
motivation

- Proliferation of inexpensive, Android-based projector on AliExpress, Temu, TikTok Shop
- Cheap hardware and misleading claims
- Rumors of surreptitiously malicious behavior

→ **Can we use CC to find the malware?**
→ **Can CC determine how the malware works?**

what we found

package	purpose
com.hotack.silentsdk	Remote Access Trojan (RAT)
com.htc.eventuploadservice	Detailed event logging
com.htc.expandsdk	Ad injection + potential malware persistence
com.htc.htcotaupdate	OTA updates over HTTP



what did we learn?

takeaways

- Keep scope narrow and targeted
- Outline specific goals
- Expose devices, protocols, and APIs with MCPs & CLIs
- Encourage to-do list generation, multi-agent use, `ultrathink`

reminders

- Act responsibly when conducting good-faith security research
- Responsible disclosure – report vulnerabilities as they are found
- Do not break the law!!!

future steps

- Expanded RE-ing of binaries with LLM tooling
  - ghidraMCP, Frida MCP, etc.
- Explore Claude's safeguards and their limits
- Experiment with computer use for greater agent autonomy
- *For me: try RE-ing with other agents*
  - *(Codex?)*