



2023 Privacy and IoT Research Exploration

Team Comeback

Tedman Nguyen and Marina Wild
Orange Coast College and Saddleback College



Introduction

Is your smartspeaker snooping on you?

- 25% of Americans own a smartspeaker (NPR, 2022) but 91% are worried speakers are eavesdropping (Hub Entertainment Research, 2019).
- We engineered a privacy focused virtual assistant using a Raspberry Pi computer and Mycroft, an open source virtual assistant, and through this we compared network traffic to Amazon's Alexa.
- We found that Alexa sent data to advertising and sales companies, while Mycroft did not.

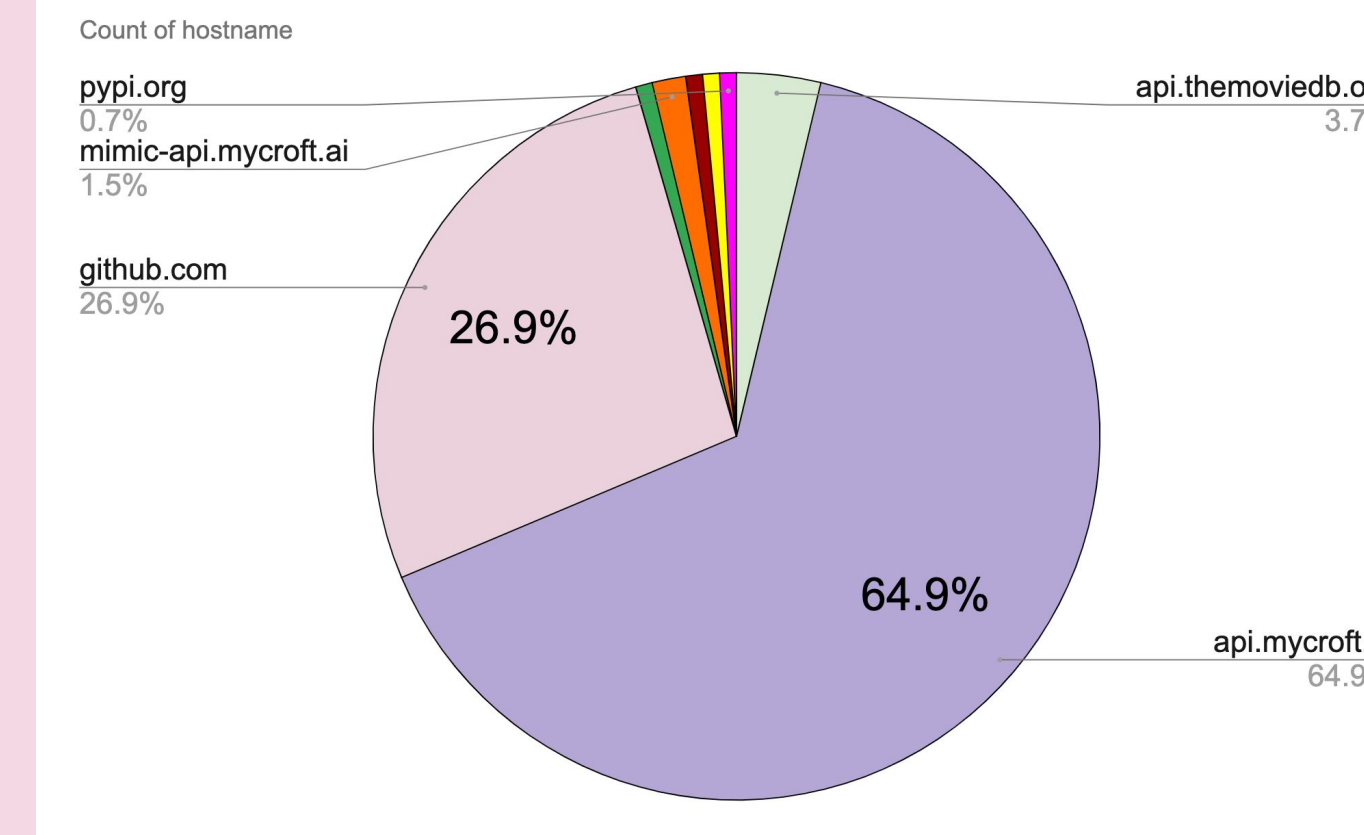
Did you know your online identity is at risk, even with "incognito"?

- Our team discovered the multitude of ways in which websites and applications (like TikTok and Amazon) collect user information, with or without cookies.
- Using Chromium's Developer Tools, we scanned outgoing network traffic on TikTok and analyzed the data using a python script, to see how our "virtual fingerprints" were created.

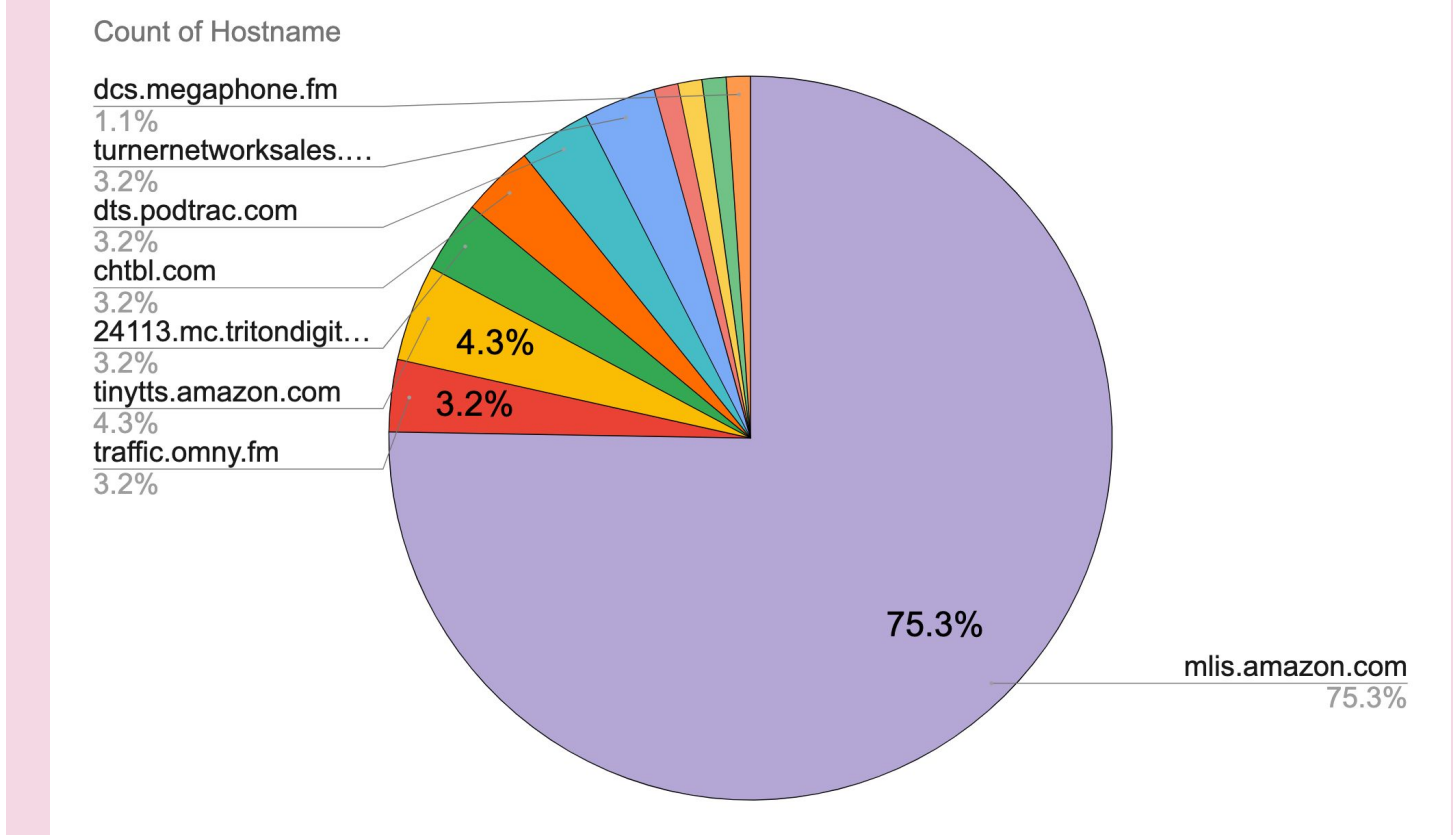
Mycroft vs. Alexa

Why Mycroft and not Alexa?

- After analyzing the network traffic from Alexa, we found that 7.5% of the network traffic was being sent to known marketing and advertising companies. For Mycroft, 0% was being sent to advertising companies.
- We can conclude that Mycroft does not track data like other traditional assistants. You can enjoy the benefits of a voice assistant without risking your privacy and your personal data.
- Mycroft can be customized to fulfill a certain need. Users can install special open source skills from Github. Whether you want help with your organization or maybe you just want an assistant that knows cool facts, Mycroft is a wonderful alternative to Alexa.

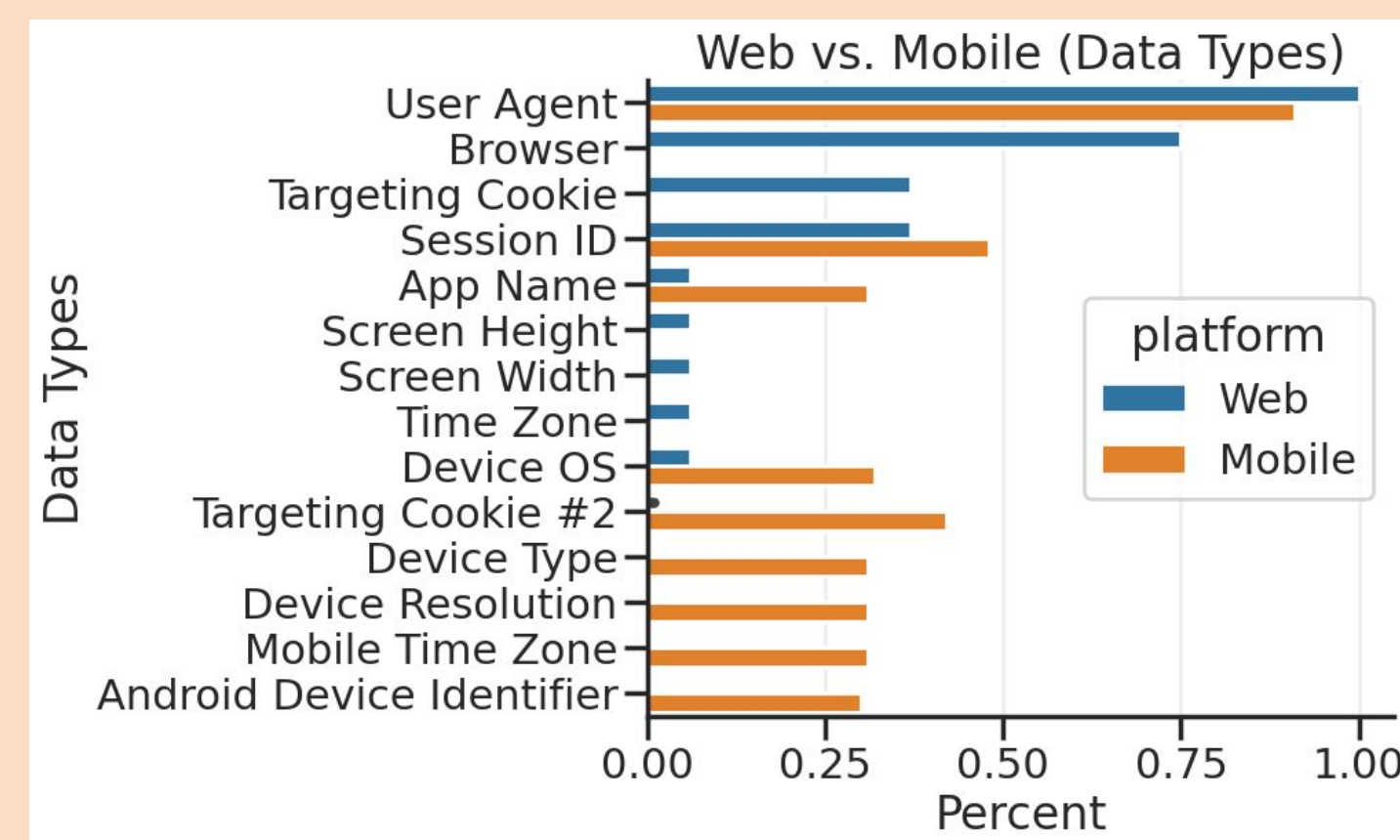


Pie Chart 1. Mycroft Network Traffic. Hostnames included Mycroft, Github, and MovieDB (which was used for our Shrek questions). Advertising / marketing hosts were non-existent.



Pie Chart 2. Amazon Alexa Network Traffic. Host names included Amazon, but also advertising companies like Triton Digital and Turner Network Sales, which sells entertainment products

TikTok Follows You



- TikTok collects user information through user actions. We explored "reposting" vs "liking" and "following" the content creator. In general, we discovered that "liking" and "following" was twice as effective at changing user curated content, when compared to "reposting" (14% vs 7% of new content shown on the "For You" page).
- TikTok also collects information from the user's device, including but not limited to timezone, language, screen resolution, operating system, browser name and version.
- On mobile, TikTok even collects a unique Android device identifier. In general, TikTok mobile uses more cookies to collect information than on its web application.

Building a Private Assistant



Procedure

- Installed a RaspBerry Pi OS by flashing the OS onto an SD card and setting up login, password, Wi-Fi.
- Downloaded Mycroft AI onto Raspberry Pi.
- Configured the device to utilize the external microphone and speaker.
- Installed Mycroft Skills, using GitHub, such as moviemaster, pokemon, learning, etc. to enable Mycroft to be more interactive and fun.
- Created a custom skill using GitHub and installed skill on our Mycroft (a Shrek skill that provides character information and writes an ode to Shrek)
- Tracked Mycroft to see if there was any tracking occurring while using the voice assistant (there was none!)

Challenges & Future

Challenges

- Our Mycroft would crash if we installed too many skills at once or did not give Mycroft time to process input. Our solution was to practice patience, reboot Mycroft to fix the issue, or even reflash the SD card, if needed.
- We both are relatively new to the topic of privacy and data as well as things like command line and Python, but we worked together to learn as much as we could and we have grown as individuals within Computer Science.

Future

- As we continue on to complete our technical degrees, we will now be taking privacy into consideration throughout any project we choose to conduct. We can help build applications that are more transparent about the way they collect data and ones that do not send information to advertisers.
- We can help inform people around us how smart devices and browsers collect data, even when they are clearing their browser history.
- We understand that we have a huge responsibility to protect people's privacy and data no matter what field we enter into.

Acknowledgements

We want to say thank you to the whole team at ProperData for providing us with the opportunity to expand our knowledge and learn the importance of privacy and security. We want to further thank Marilyne, Ernest, Hieu, Rahmadi, Athina, UCI Catering, and Alison for making this experience, not only valuable, but very fun!
Thank you!

References

- NPR and Edison Research. The Smart Audio Report. 2022 <https://www.nationalpublicmedia.com/insights/reports/smart-audio-report>
- Hub Entertainment Research. The Case for Voice Control. December 31, 2019 <https://hubresearchllc.com/>

