Improving Measurement Vantage Points Within Smart Home Networks to Identify and Mitigate IoT Security and Privacy Risks

Danny Yuxing Huang (New York University) and Deepak Kumar (Stanford University)

Critical research questions: Consumer home networks are increasingly filled with insecure "smart home" IoT devices. Given that most of these devices are hidden behind NATs, how can researchers improve current measurement vantage points into real smart home networks to assess and mitigate the security and privacy threats of IoT devices?

Reasons for these research questions: We believe these questions are important for the following reasons:

Devices on home networks are rife with diverse security vulnerabilities. IoT devices are gaining in popularity in home networks. This ecosystem is loosely regulated and heavily fragmented, with a myriad of OEMs and software developers involved. IoT vulnerabilities thus take many forms. Our preliminary evidence shows that many devices—both IoT and non-IoT—naively trust the local area network (e.g., opening ports and not checking for single origins [1]) and not defend against even the most basic attacks (e.g., ARP spoofing and man-in-the-middling [2]). Furthermore, the vulnerability profile of IoT devices varies across geographic region and device type [3]. Fine-grained measurements in real home networks are a critical first step toward identifying the diverse and often previously unseen security and privacy risks before we can mitigate them.

Current measurement techniques are insufficient for large-scale analysis. Most IoT measurement research currently takes one of three forms: in-lab experimentation, proprietary data, or crowdsourcing. Although in-lab experiments typically help towards better device identification and deeper case studies on specific device vulnerabilities, they do not offer insight into how devices are used in practice. In contrast, proprietary data can provide large-scale access to home networks around the world [3], but they lack reproducibility and restrict access to only those with sufficient connections to industry, thus limiting potential research. A third, smaller category of research focuses on crowdsourced efforts by academics (e.g., IoT Inspector and BISmark [2,4]) which provide deeper access but have thus far been limited in scale (both geographic and a wide number of participants) and study duration (participants running measurement software and hardware for short periods of time).

Areas for NSF support: We believe that crowdsourcing will allow researchers to conduct long-running in-situ home measurements in a way that is flexible (i.e., tailoring to researchers' needs because researchers can develop the platform themselves), privacy-preserving (i.e., protecting the volunteers that contribute the data), and reproducible (because researchers have access to the data), if NSF could provide support in the following areas:

Developing self-sustaining measurement infrastructure at scale. Current crowdsourced techniques are mired with challenges: Although 5,000+ non-paid volunteers have installed IoT Inspector, volunteer users are primarily located in the US, contain few IoT devices in their homes, only record 45 minutes of data in the median case, and require researchers to constantly attract new volunteers [2]. To encourage a wide breadth of voluntary, non-paid participants over a longer duration, researchers need to improve the usability of their crowdsourcing tools to incentivize adoption. NSF could offer support in HCI (especially usability) research, for example, by developing and improving UI/UX designs (e.g., more friendly and for non-technical users with new software features) for crowdsourcing systems to attract a significantly larger user base (e.g., hundreds of thousands of users) from diverse geographical regions and maintain user engagement for a long period of time (e.g., boosting daily active users). These improvements will help researchers gather real-world data from a diverse sample and conduct longitudinal studies without significant overhead in recruiting and keeping users.

Curating, sharing, and analyzing data. While such a crowdsourcing-based system will likely provide researchers with vantage points from within real home networks, crowdsourcing introduces two main challenges: (i) data quality, because user inputs could be missing, inconsistent, or inaccurate; and (ii) user privacy, as we are potentially collecting sensitive data from real users. NSF could offer support in machine learning research that first ensures the quality of crowdsourced data, e.g., by developing NLP and anomaly detection techniques to validate user labeled network data. Moreover, NSF could support research on crowdsourcing and privacy, e.g., by developing techniques in federated learning and differential privacy, that would balance the tension between protecting user privacy and the data access needed for research.

Making knowledge useful for consumers and regulators. The datasets and models from crowdsourcing will provide consumers and regulators with transparency into what would otherwise be black-box IoT devices, if researchers present their findings in a usable and non-technical form. NSF could offer support in HCI research, e.g., user studies to understand mental models of security and privacy risks *in the real world* (in contrast to existing work in the lab [5,6]), as well as UI/UX research to visualize findings from the massive dataset through public, consumer- and regulator-facing websites or dashboards (no existing work known). This effort will help us make a broad impact, ranging from influencing individual consumers' purchase behaviors toward more secure IoT products, to shaping regulations in IoT security and privacy.

References

[1] “Fast Web-based Attacks to Discover and Control IoT Devices.” Gunes Acar, Danny Yuxing Huang, Frank Li, Arvind Narayanan, Nick Feamster. *ACM SIGCOMM Workshop on IoT Security and Privacy (IoT S&P)*. 2018.

[2] “IoT Inspector: Crowdsourcing Labeled Network Traffic from Smart Home Devices at Scale.” Danny Yuxing Huang, Noah Apthorpe, Gunes Acar, Frank Li, Nick Feamster. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT / Ubicomp)*. 2020.

[3] “All Things Considered: An Analysis of IoT Devices on Home Networks.” Deepak Kumar, Kelly Shen, Benton Case, Deepali Garg, Galina Alperovich, Dmitry Kuznetsov, Rajarshi Gupta, Zakir Durumeric. *USENIX Security Symposium.* 2019.

[4] “BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks.” Srikanth Sundaresan, Sam Burnett, Nick Feamster, Walter de Donato. *USENIX Annual Technical Conference*. 2014.

[5] “Ask the Experts: What Should Be on an IoT Privacy and Security Label?” Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Cranor, and Hanan Hibshi. *IEEE S&P*. 2020.

[6] “Exploring How Privacy and Security Factor into IoT Device Purchase Behavior.” Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Cranor, and Henry Dixon. *CHI*. 2019.