Weijia He's Research Statement

The proliferation of IoT devices in recent years has raised significant security and privacy concerns [1]. The pervasiveness and fragmentation of IoT devices necessitate a new threat model distinct from traditional computing devices such as computers or smartphones. However, many security practices and mindsets have not changed. Access control in smart homes, for example, retains the admin-guests model that we used to have on computers, despite the fact that the social relationships between users are much more complicated [2]. IoT systems are also sensitive to environmental changes, attackers can alter the system's behaviors by changing the environment. However, many security mechanisms fail to recognize threats from the analog world [3].

My research aims to bridge the gap between existing security systems and the challenges that real-world users face. I design usable, robust, and secure IoT sensing systems for ubiquitous computing, with techniques from security, human-computer interaction (HCI), and data-driven system building. I've worked on various aspects of an IoT system, such as access control [2, 3, 4], context sensing [3], network security (under submission), and trigger-action programs [5, 6], to make the system more secure, private, and usable. Throughout my Ph.D., I have published at venues like USENIX Security [2], IMWUT (UbiComp) [6], ICSE [5], EuroS&P [3], and CHI [7]. I was also named a 2022 Siebel Scholar for my efforts. Below, I will discuss my previous research, research philosophy, and future plans.

Prior Research

Rethinking Robust Access Control in Home IoT. While one-third of US households already own at least one IoT device [8], bringing additional IoT devices into homes will exacerbate challenges to security practices like access control and authentication [9, 10]. I'm curious how a household would handle access control policies involving multiple stakeholders (e.g., spouse, children, domestic workers), and whether current access control systems would suffice.

My first research project in my PhD sought to understand how desired access control policies change over different device capabilities, contexts, and social relationships. For this project, I primarily employ HCI techniques. I conducted a 425-person online survey to determine people's default attitudes toward various potential users and how those attitudes change depending on contexts such as time, location, people around them, and so on. We found that access control in a home IoT environment involves a variety of users who have complex social relationships, which changes people's desired default policies. A simple admin-guest user model is no longer adequate. Furthermore, the desired policy is also dependent on the variable contexts in a home, as shown in Figure 1. For example, we found that children are one of the least trusted members of a family, but given specific circumstances, such as having an adult around to supervise them, they could be given more access. Similarly, domestic workers or visiting family members may have greater control over the devices, particularly when the owner is not present. These findings shed light on people's complex mental models about home IoT access control, as well as how inadequate current access control is.

Knowing how contexts can change one's desired access control policy, I wondered why a more suitable, context-based access control isn't available yet. The question prompts me to do a literature review on sensing. All the papers offer promising approaches for making a system more context-aware. Unfortunately, as a security researcher, I quickly realized that the majority of these works are not impenetrable to adversarial parties. Furthermore, when I looked through the security literature, I noticed that many papers ignore the possibility of an internal attacker with physical access to the environment [11, 12].

I thus led a systematization of knowledge (SoK) study on both sensing and security papers. I carefully selected and synthesized previous literature to create a framework for evaluating sensing methods in terms of security, privacy, and usability, and each sensing method is also mapped to a context discovered in the previous study. As a result, my work can assist future researchers or smart home designers in selecting the most appropriate sensing method for their needs. A camera, for example, is ideal for monitoring outdoor activities because they are widely available and simple to install, and outdoor areas are typically regarded as less privacy-sensitive. Cameras may not be a good option for monitoring indoor activities that are more privacy-sensitive, and other sensing methods such as RF sensors or even WiFi signals can be used instead [13, 14]. The framework can be used by any smart home producers to think about the sensing options they can have in their products. It also points out potential security risks that they should take precautions about, which eventually makes their design more secure.

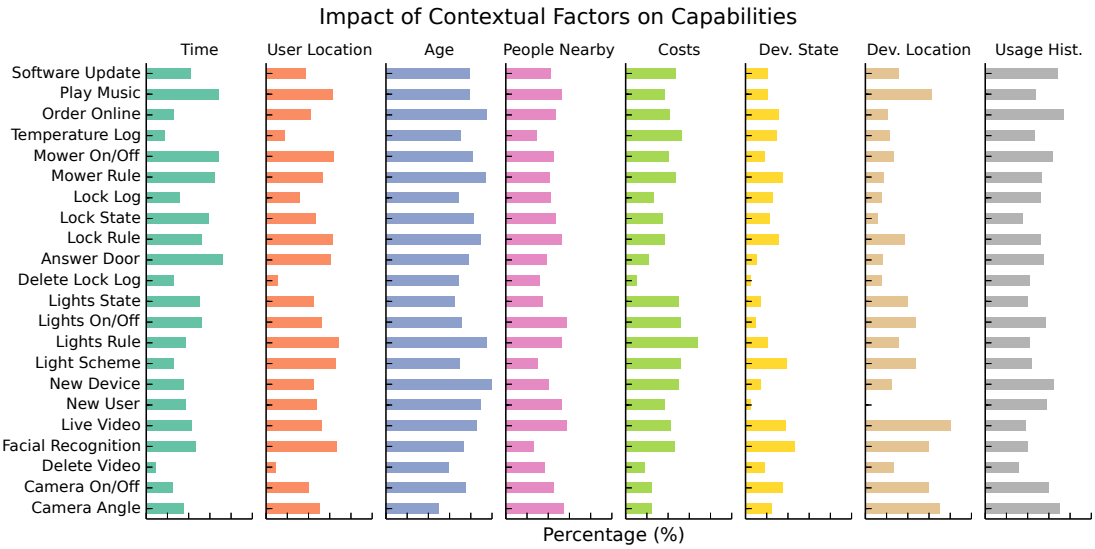

Figure 1: Sometimes access must depend on the context. In the study we asked participants for such factors and identified multiple that are very influential (such as the age of the user) and learned how they contribute to the decision make process.

Automatically Generated, Generalizable Network Allowlists for IoT Devices. Another distinguishing feature of IoT devices is that their functionality is typically simpler, which means their network traffic is less chaotic than that of traditional computing devices such as computers or smartphones. We hypothesize that allowlist may be feasible here. According to some research, it is possible to learn allowlists on a single device [15, **?**]. Learning allowlists for every device, on the other hand, can create great burdens on end-users. It would be preferable to have a generalizable allowlist. As a result, I set out to investigate common network behaviors of home IoT devices on a much larger scale, attempting to comprehend the design space of allowlists and the changes an allowlist would bring to these devices.

I used the IoT Inspector dataset to explore the design space of automatically generating allowlists from 5,439 home IoT devices in the wild [16]. The traffic is noisy, but it is more realistic because, unlike in lab studies, the participants are simply using their devices as they normally would. These devices may run on different network settings, the modalities could be different, and there is no guarantee all the functionalities have been used during data collection.

Working with such a large-scale dataset for allowlist creation is challenging. There are numerous design options available when creating an allowlist, and if one makes a wrong design choice, the whole device can be broken. To gain a better understanding of the design space, we generate allowlists for each product based on different host representations (e.g., hostname, domains, etc.), data subsets (e.g., only include data from a particular region), and strictness (e.g., an allowed host must be contacted by at least N devices). Thanks to the large-scale dataset, we discovered several important factors that are rarely seen in lab studies. For example, creating hostname patterns for load balancing servers is essential. However, a poor host naming approach can result in a pattern that is overly trusting, allowing attackers to circumvent the allowlist restriction (e.g., one has to trust everything under a domain to cope with a partially random hostname).

More than just network analysis, I built our own firewall to see how these allowlists generated from real-world devices would work with devices outside the dataset. The experiment is conducted two years after the data collection. Despite of two years of firmware updates, it turns out that these allowlists can still work with most functionalities that do not involve audio or video streaming, demonstrating the stability of allowlists.

Research Philosophy

My research has touched on various aspects of an IoT system over the years, and many different techniques have been used, ranging from user studies to network measurements. Through these studies, I gradually developed my own research philosophy that can be applied to a variety of fields. I begin by reviewing existing research and systems, making questions on common underlying assumptions that are taken for granted. I then introduce real

users into the picture to determine whether these assumptions are valid in the real world. Finally, I use what I've learned to create actionable metrics, improve existing methods, and build more advanced systems.

Understanding Gaps in Existing Approaches and Systems. My first step in the research is always to revisit current approaches to an open problem. A rigorous literature review or actual measurements of existing systems could be used as the method. For example, while conducting a large-scale literature review on the most recent sensing methods, I became aware of the lack of a security mindset. The realization prompts us to create our own framework and metrics for assessing the vulnerability of existing sensing methods in an adversarial environment.

When I worked on the allowlist project, I also analyzed multiple IoT traffic datasets, ranging from in-lab datasets to large-scale datasets from the wild [16], to understand how heterogeneous IoT traffic can be, which led to the realization that lab studies of IoT traffic are frequently not generalizable, particularly when we are discussing restrictive security methods like allowlists. For example, the Alexa endpoints for a single Amazon Echo rarely change, but there are dozens of such endpoints in the wild. Ignoring such facts can render an allowlist useless.

Standing in the User's Shoes. Thinking about how a real user interacts with the system is always helpful in identifying unreasonable assumptions made by previous research or existing systems. System lab studies frequently focus solely on whether they fulfill their purpose, with no consideration for the complexities that may arise in real life. To gain a better understanding of users, I conducted online surveys to learn about the challenges that a non-technical user might face in a real smart home, as well as how the current access control system might fail [2, 17, 4]. In addition, I have conducted interviews and field studies to evaluate the usability of actual systems. I placed various home IoT devices in people's offices, logged their daily activities for two weeks, and then interviewed them at the end of the study. The process helped me understand participants' reasoning behind each interaction and what a machine-logged device activity trace may miss. For example, in machines' logs, turning lights off always comes before closing the door, but users' reasonings are in the reverse order (i.e., they turn the lights off because they are leaving).

New Insights for Future Systems. Unlike other areas of computer science research, security research is always concerned with safeguarding the future with technological advances. Our discoveries should have an impact on how future systems are built and designed. My previous work in access control encourages future researchers to reconsider how one should design control in a household. I also developed new frameworks and metrics to assist future researchers or smart home designers (such as companies who develop smart home products or DIY-styled smart homeowners) in selecting and evaluating sensing methods that are most appropriate for their own use cases. Allowlist research extends beyond studying devices in labs to understanding and recognizing the heterogeneity of devices in the wild, which can change the design of an allowlist.

Future Research Plans

In the future, I hope to expand my work on home IoT to more complex ubiquitous computing systems, such as smart cities or self-driving cars. These systems share characteristics with home IoT, such as their reliance on sensors. They also involve humans in a complex, and sometimes unseen, way, causing great confusion as well as serious security implications. Techniques I learned from previous research help me better understand users' perceptions and expectations, identify potential attacks and how they affect existing systems, and ultimately build more usable and secure systems. I include some future research plans that I am excited to work on below.

Usable Sensing Transparency for Ubiquitous Computing. With all of the new technologies emerging, the world is becoming more sensor-rich, raising security and privacy concerns. There have been numerous reports of hidden network-enabled devices (e.g., cameras, drones, trackers) being used without consent for stalking, theft, or videotaping. The first step toward protecting people from such fear is to provide adequate transparency. People should be well informed about where these sensors are placed, whether in a hotel room or on a public street.

Detection and localization are the two most important methods for making sensors transparent. Both have been studied for many years, but rarely from the point of view of the user. There are three major concerns: (1) how an attacker can avoid current detection and localization tools if they have complete control over the hidden sensor and its surroundings; (2) if the sensor is well-hidden, how a real user will use existing tools to locate it; and (3) how environments affect the result (e.g., in a hotel room, in a shopping mall, or on a public street).

A system that can detect surrounding sensors and let people know where they are would be great. I intend to begin the project with cameras and then move on to other types of sensors, such as microphones, ultrasound sensors, LiDAR, etc., as they can all be used to obtain sensitive information. To answer the preceding questions, we must first understand how existing detection and localization tools can be blocked or become error-prone. Once we determine how attackers can perplex these tools, we plan to conduct a lab study to see how these tools perform with intervention, as well as how participants may unintentionally ignore certain threats (e.g., blind spots during searching). After obtaining the data, we can create a more robust tool to guide people to locate sensors around them in a more efficient manner. We would eventually like to conduct field studies to further investigate obstacles that one might encounter in a more complicated scenario, such as a crowded street.

Human-Vehicle Interaction in Adversarial Scenarios. In recent years, one of the hottest topics has been self-driving cars. Attacks on autonomous vehicles (AV) sensor systems, such as physical adversarial examples to cameras and LiDAR, are on the rise. If these attacks occur in reality, it is up to the drivers to take control and avoid potential accidents. However, these attacks can catch drivers off guard, and the situation can quickly turn disastrous. It is extremely difficult to expect drivers to make correct decisions under pressure in a matter of seconds.

To defend against such attacks, sensor fusion is always one method for detecting anomalies in sensor data, but with more sensors onboard, it's unclear how to effectively present the anomalies to the driver. If we can choose the sensor wisely, we can not only protect the system, but also provide explanations behind the anomaly, and suggest actionable next steps accordingly. I envision developing a system that can detect potential anomalies and present them to non-technical drivers in an understandable manner using techniques from HCI, sensing, and security. The system should also provide effective actionable suggestions to drivers under emergencies.

References

[1] S. Mattu and K. Hill. The house that spied on me, 2018. `https://gizmodo.com/the-house-that-spied-on-me-1822429852`.

[2] **W. He**, M. Golla, R. Padhi, J. Ofek, M. Dürmuth, E. Fernandes, and B. Ur. Rethinking access control and authentication for the home internet of things (iot). In *Proc. USENIX Security*, 2018.

[3] **W. He**, V. Zhao, O. Morkved, S. Siddiqui, E. Fernandes, J. Hester, and B. Ur. SoK: Context sensing for access control in the adversarial home iot. In *Proc. EuroS&P*, 2021.

[4] **W. He**, J. Hainline, R. Padhi, and B. Ur. Clap on, clap off: Usability of authentication methods in the smart home. In *Proceedings of the Interactive Workshop on the Human Aspect of Smarthome Security and Privacy*, 2018.

[5] L. Zhang, **W. He**, J. Martinez, N. Brackenbury, S. Lu, and B. Ur. Autotap: synthesizing and repairing trigger-action programs using LTL properties. In *Proc. ICSE*, 2019.

[6] L. Zhang, **W. He**, O. Morkved, V. Zhao, M. L. Littman, S. Lu, and B. Ur. Trace2TAP: Synthesizing trigger-action programs from traces of behavior. *IMWUT*, 4(3), 2020.

[7] W. Brackenbury, A. Deora, J. Ritchey, J. Vallee, **W. He**, G. Wang, M.L. Littman, and B. Ur. How users interpret bugs in trigger-action programming. In *Proc. CHI*, 2019.

[8] Statista. Do you own smart home devices - i.e. devices that you can control via a smartphone / an internet connection?, Jul 2021. `https://www.statista.com/forecasts/1097129/smart-home-device-ownership-in-selected-countries`.

[9] E. Fernandes, J. Jung, and A. Prakash. Security Analysis of Emerging Smart Home Applications. In *Proc. IEEE SP*, 2016.

[10] Y. J. Jia, Q. A. Chen, S. Wang, A. Rahmati, E. Fernandes, Z. M. Mao, and A. Prakash. ContexIoT: Towards Providing Contextual Integrity to Appified IoT Platforms. In *Proc. NDSS*, 2017.

[11] R. Schuster, V. Shmatikov, and E. Tromer. Situational access control in the internet of things. In *Proc. CCS*, 2018.

[12] Z. B. Celik, G. Tan, and P. D. McDaniel. Iotguard: Dynamic enforcement of security and safety policy in commodity iot. In *Proc. NDSS*, 2019.

[13] S. Tan, L. Zhang, Z. Wang, and J. Yang. Multitrack: Multi-user tracking and activity recognition using commodity wifi. In *Proc. CHI*, 2019.

[14] C. Hsu, R. Hristov, G. Lee, M. Zhao, and D. Katabi. Enabling identification and behavioral sensing in homes using radio reflections. In *Proc. CHI*, 2019.

[15] J. Habibi, D. Midi, A. Mudgerikar, and E. Bertino. Heimdall: Mitigating the internet of insecure things. *IEEE Internet of Things Journal*, 4(4), 2017.

[16] D. Y. Huang, N. J. Apthorpe, F. Li, G. Acar, and N. Feamster. Iot inspector: Crowdsourcing labeled network traffic from smart home devices at scale. *IMWUT*, 4(2):46:1–46:21, 2020.

[17] **W. He**, J. Martinez, R. Padhi, L. Zhang, and B. Ur. When smart devices are stupid: Negative experiences using home smart devices. In *2019 IEEE Security and Privacy Workshops, SP Workshops 2019*, pages 150–155, 2019.