Weijia He's Research Statement

The proliferation of home IoT devices in recent years has raised significant security and privacy concerns. The pervasiveness and fragmentation of home 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 [5]. My research aims to bridge the gap between existing security systems and the challenges that real-world users face. I revisited systems for their underlying assumptions and questioned how real users and attackers might go against these assumptions and break the security promises in the real world, from both security research and human-computer interaction (HCI) perspectives I've worked on various aspects of an IoT system, such as access control [2, 5, 3], context sensing [5], network security (under submission), and trigger-action programs [8, 9], to make the system more secure, private, and usable. Throughout my Ph.D., I presented papers at conferences such as Usenix Security [2], IMWUT (UbiComp) [9], ICSE [8], EuroSP [5], and CHI [1]. I was also named a Siebel Scholar for my efforts. In the following sections, I will discuss my previous research, research philosophy, and future research.

Understanding Access Control in Home IoT

Home IoT is one of the most well-established IoT concepts. One-third of US households own at least one piece of home IoT equipment [7]. However, bringing IoT into homes introduces new challenges to traditional security practices such as access control. Access control in a home IoT environment involves a variety of users who have complex social relationships, such as spouses, parents and children, domestic workers, and so on. Therefore, a simple admin-guest user model is no longer adequate. Furthermore, the desired policy is also dependent on the variable contexts in a home.

My first research project seeks to understand access control policies change over different device capabilities, contexts, and social relationships. 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 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.

I thus conducted 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 radar sensors or even WiFi signals can be used instead. 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.

Generalizing Allowlists across Home IoT Devices (Under Submission)

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. Allowlists, rather than blocklists, can thus be deployed. According to some research, it is possible to learn allowlists on a single device. 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 analyze traffic from 5,439 home IoT devices in the wild [6]. 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, data subsets, and strictness. 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 completely 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. Despite the two-year gap between the IoT Inspector data collection and our own experiment, 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.

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.

Analyzing Existing Research/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 [6], 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 Users' 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, 4, 3]. 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.

Soliciting 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

Usable Sensing Transparency for Ubiquitous Computing

Although my previous research has primarily focused on smart homes, the concept of ubiquitous computing has expanded beyond the home. With all of the emerging technologies, the world is becoming more sensor-rich, which raises security and privacy concerns. Numerous reports have been made about how hidden network-enabled devices (e.g., cameras, drones, trackers) are used without consent for stalking, theft, or videotaping. Transparency is the first step toward protecting people.

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] Will Brackenbury, Abhimanyu Deora, Jillian Ritchey, Jason Vallee, Weijia He, Guan Wang, Michael L. Littman, and Blase Ur. How users interpret bugs in trigger-action programming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI'19*, page 552, 2019.

[2] Weijia He, Maximilian Golla, Roshni Padhi, Jordan Ofek, Markus Dürmuth, Earlence Fernandes, and Blase Ur. Rethinking access control and authentication for the home internet of things (iot). In *27th USENIX Security Symposium, USENIX Security'18*, pages 255–272, 2018.

[3] Weijia He, Juliette Hainline, Roshni Padhi, and Blase 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.

[4] Weijia He, Jesse Martinez, Roshni Padhi, Lefan Zhang, and Blase 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.

[5] Weijia He, Valerie Zhao, Olivia Morkved, Sabeeka Siddiqui, Earlence Fernandes, Josiah Hester, and Blase Ur. Sok: Context sensing for access control in the adversarial home iot. In *IEEE European Symposium on Security and Privacy, EuroS&P'21*, pages 37–53, 2021.

[6] Danny Yuxing Huang, Noah J. Apthorpe, Frank Li, Gunes Acar, and Nick Feamster. Iot inspector: Crowdsourcing labeled network traffic from smart home devices at scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(2):46:1–46:21, 2020.

[7] 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.

[8] Lefan Zhang, Weijia He, Jesse Martinez, Noah Brackenbury, Shan Lu, and Blase Ur. Autotap: synthesizing and repairing trigger-action programs using LTL properties. In *Proceedings of the 41st International Conference on Software Engineering, ICSE'19*, pages 281–291, 2019.

[9] Lefan Zhang, Weijia He, Olivia Morkved, Valerie Zhao, Michael L. Littman, Shan Lu, and Blase Ur. Trace2tap: Synthesizing trigger-action programs from traces of behavior. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(3):104:1–104:26, 2020.