Earlence Fernandes – Research Statement

Science and technology is progressing at an incredible pace, fueled in part by the ubiquity of computing. These computer-enabled technologies bring new benefits, but also new security and privacy risks. As a systems security researcher, I anticipate these risks, and address their root causes before attackers can exploit them widely. In doing so, my work lays a foundation to secure emerging technologies. Although such an approach to security research benefits from being able to quickly adapt to new domains, it is also vital to explore issues deeply. Thus, I tackle security and privacy problems using the following process:

- *Understand security and privacy failures* through principled measurements and analysis of existing and emerging technology. This involves collecting data through the development of measurement tools and user studies, developing threat models for emerging domains, and discovering attacks on systems.
- *Build security and privacy into systems* by extracting insights from the measurements, identifying the right primitives that offer a reasonable trade-off between security, privacy, usability, and performance, and developing practical solutions that optimize this trade-off for various threat models.

This iterative process has been very effective in anticipating the threats of new technologies, and in addressing them at a design level. Below, I will briefly overview a few examples of my work that illustrate this.

Understanding Security and Privacy Issues. A prerequisite to building secure systems is to understand their failure modes, and the extent of those failures. This raises awareness in multiple stakeholders, and ensures that we tackle the right problems. In keeping with this method, I have led studies investigating security and privacy issues in emerging technologies with a focus on cyber-physical systems (CPSs). For example:

- *Systems security issues:* I performed the first security analysis of Samsung SmartThings, an emerging platform for home automation, showing how overprivilege design flaws can lead to physical attacks [18]. I also led the analysis of If-This-Then-That (IFTTT), an end-user programming system for homes, showing that such systems are vulnerable to compromise and pose a long-term security threat [21]. Attacks on these systems take the form of malicious apps or stolen tokens that can be used to open door locks, cause fake fire alarms, and snoop on occupants. My work on the SmartThings analysis earned the ***Distinguished Practical Paper Award at IEEE S&P 2016***.
- *New threats in ML:* With the increasing deployment of machine learning, including deployments in CPSs, I have also focused my efforts on analyzing its security implications. I recently co-led a study that showed how state-of-the-art deep learning models are vulnerable to physical-world manipulations of their inputs, and how this could affect emerging CPSs that use learning components in their control pipeline [10], [11].
- *Privacy issues:* An increasingly computerized world leads to a lot of data that is helpful in extracting actionable insights. I showed how entity recognition systems, which associate semantic labels to text, transmit in-depth app usage data, such as health information, to cloud servers for analysis, leading to possible privacy violations [23].

Security and Privacy by Design. Motivated by the analyses, I build systems that tackle the security and privacy challenges at design time. To mitigate attacks on platforms like SmartThings and IFTTT, I led the design and implementation of systems that provide security at multiple levels—Tyche [25] introduces risk-based permissions, FlowFence [19] introduces a novel mechanism for information flow control, and DTAP [21] introduces strong integrity properties. To tackle the privacy issues in data analysis systems, I designed and built the Appstract framework, a system and set of algorithms that performs entity recognition in a privacy-respecting way [22], [23]. My work on risk-based permissions earned the ***Best Paper Award at IEEE SecDev 2018***.

Anticipating and fixing security issues in emerging technologies often leads to problems at the intersection of systems security and other areas. Thus, I seek out domain experts in related fields. This helped bootstrap a collaboration that led to an NSF grant on CPS security testing. I was also recently part of a team of computer scientists and law researchers that examined the impact of tricking machine learning models, and determined whether such tricking falls under the purview of the Computer Fraud and Abuse Act [7].

Long-Term Outlook. Finally, I believe that if there are ways to benefit society *after* publishing papers, then it is important to facilitate such positive impact. I collaborated with Samsung to provide guidance and support in fixing issues from my security analysis of SmartThings. This has resulted in platform improvements along multiple dimensions: security-related developer guidance, platform-level security enhancements [4], and better app store

vetting processes. Based on this experience, my work on FlowFence and DTAP has been adapted and implemented in upcoming Samsung products. I am also currently exploring collaborations with Lenovo to implement risk-based permissions. The Appstract system resulted in two patent filings for Microsoft. Furthermore, code and data resulting from these projects is open source, with the goal of bootstrapping future work in the community [1]–[3], [5], [6].

I often engage with the press when society at large would benefit from awareness about my results—the SmartThings analysis and physical adversarial perturbations work received widespread coverage, and raised awareness among key stakeholders. For example, I have advised a member of Congress on Internet of Things security, and several US government agencies including the FTC, the National Academies, and the JASONs on security issues related to the deployment of ML systems in critical infrastructure.

CURRENT WORK

I will discuss some of my recent work that focuses on multiple layers of the consumer CPS computing stack: systems software, learning components, usability, and data analysis. These projects highlight my systems approach to securing emerging technologies, with the goal of gaining their benefits without the risks. The interested reader may refer to the bibliography for pointers to my earlier work in smartphone security [8], [15]–[17], [27], [28], [31].

A. *Systems Security Issues in Consumer Cyber-Physical Systems*

Consumer cyber-physical systems are widely deployed in homes, schools, and offices. A consumer CPS is fundamentally a network of embedded devices that work together with the help of middleware. This middleware unifies heterogeneous devices and network protocols into a uniform platform on top of which higher-level functions run. Vulnerabilities in design of the middleware will lead to remote and device-independent attacks. Thus, analyzing the middleware for flaws, and building security primitives in, is vital to ensure long-term security.

Depending on the type of consumer CPS, there is a range of middleware categories. For complex use-cases, there are platforms that professional developers use to provide functionality like recognizing faces from a camera feed, and automatically controlling a door lock. For simpler use-cases, there are platforms that allow end-users themselves to create automation rules such as turning off an oven when the smoke detector goes off. To ensure deep coverage of the security design problems in CPS middleware, my work focuses on both types of platforms.

Analyzing and Fixing CPS Middleware for Complex Apps [18], [19], [25]. I led the first peer-reviewed security analysis of Samsung SmartThings, a mature middleware platform with wide support for devices and third-party apps [18], [20]. SmartThings shares core design principles with other platforms in this category, and therefore, the insights my analysis extracted are broadly applicable to this class of middleware for CPSs. Using black-box fuzzing and custom-built static analysis tools, I found that SmartThings does not adhere to the classic security principle of least-privilege. Particularly, the middleware enables apps to be overprivileged—a situation where software has access to more sensitive resources than it needs to perform its stated function. Based on the results from this analysis, I created proof-of-concept attacks that reprogrammed door locks and caused fake fire alarms. This work earned the ***Distinguished Practical Paper Award at IEEE S&P 2016***.

Although overprivilege is not a new problem, it is a particularly challenging problem to solve, often requiring tailoring a solution to a specific domain. Another challenging aspect is that apps can use their access to cause physical harm—a threat that is not present in classical computer systems.

To effectively tackle this security design flaw, I invented defenses on multiple levels. First, to limit the risk overprivileged apps pose to the physical world and its occupants, I re-visited access control and proposed that permissions for connected devices like ovens and door locks be grouped in terms of risk, rather than function (which is the norm today). Such a risk-based grouping takes advantage of the intuitive asymmetry in device operations. Taking the example of an oven, `oven.on` is a potential fire hazard, and `oven.off` is uncooked food. We designed and built Tyche, a permission system for smart homes that groups device operations in terms of risk. A challenge in building this system was to devise a technique for middleware developers to consistently and scalably estimate user-perceived risk. Apps running on Tyche remain functional, while reducing their access to high-risk operations by 60%. This work earned the ***Best Paper Award at IEEE SecDev 2018*** [25].

Even if apps use operations that are risk-appropriate, they can still use those permissions in ways that are inconsistent with user expectations. Thus, it is important to also secure *how* apps use their access to sensitive data and devices. Towards addressing this problem, I led the design and implementation of FlowFence [9], [19], [26],

a system where information flow control is a first class primitive. FlowFence is inspired by ideas introduced in earlier work on information flow control such as COWL [29], and Hails [12]. A unique aspect in FlowFence is that it requires developers to package code operating on sensitive data into modules whose return values can only be de-referenced inside other such modules. This approach makes information flow explicit, and is a secure building block for CPS app platforms. Ideas from this project have been adapted for use in Samsung products.

Analyzing and Fixing CPS Middleware for End-User Programming [21]. There is a growing set of platforms that allow end-users to program small automations themselves. This class of *Trigger-Action* platforms allows home owners to create rules like "If smoke is detected, then turn off my oven." To ensure that we gain the benefits of this emerging CPS middleware without the risks, I analyzed its security using measurement tools that I built. The main finding is that existing trigger-action platforms place too much trust in the integrity of cloud services. A more realistic threat model is to assume that a trigger-action platform can be compromised, and attackers can access all OAuth tokens.

The second stage of my process is to develop the right security primitives, and incorporate them at design time. To that effect, I introduced *Decentralized Action Integrity* [21]. This principle improves the security properties of OAuth while maintaining similar usability, and ensures that an attacker who controls a compromised trigger-action platform: (1) can only invoke actions and triggers needed for the rules that users have created; (2) can invoke actions only if it can prove to an action service that the corresponding trigger occurred in the past within a reasonable amount of time; and (3) cannot tamper with any trigger data passing through it undetected. Ideas from this project are currently being implemented in Samsung Products.

B. Understanding Security Issues in Systems That Learn

With the ubiquity of deep learning, and the increasing integration of intelligent components in traditional computer systems, new risks can arise. I will discuss a few of my projects that focus on understanding these new risks, and how they change the attack surface of cyber-physical systems.

Physical Adversarial Examples [10], [11]. Deep Neural Networks (DNNs) are being used in the control pipelines of physical systems like cars, UAVs, and robots. However, DNNs are vulnerable to *adversarial examples*—perturbations to their inputs that cause predictable misbehavior [30]. I co-led a project that investigated the *physical effectiveness* of adversarial perturbations. Current work in adversarial machine learning has assumed digital access to input vectors. However, for CPSs like cars, having digital access is a strong assumption. A more likely threat is that an attacker can manipulate the physical world perceived by the car. My work shows how attackers can modify physical objects to cause these attacks.

Fig. 1: The left image shows graffiti on a Stop sign in a city, a relatively common occurrence that most humans would not think is suspicious. The right image shows our example physical perturbations applied to a real Stop sign. We design our perturbations to mimic graffiti, and thus "hide in the human psyche." The right-side image is interpreted as Speed Limit 45 by DNN-based classifiers.

Our initial results indicate that DNN-based *classifiers* can be effectively tricked in the physical world [11]. Figure 1 shows an example attack, where a physical Stop sign is modified with simple black and white stickers, causing a DNN-based road sign classifier to output the label 'Speed Limit 45' instead of the expected label 'Stop.' The main challenge is to overcome physical world sources of noise such as changing angles/distances, and sensor imperfections

Based on this initial work, we extended our results to attack *object detectors*, a richer type of DNN that locates an object in a scene, in addition to classifying it. This kind of model is harder to trick because it uses a lot of contextual information while making predictions, unlike a classifier that only sees a small part of a scene. Our work successfully attacks the state-of-the-art YOLO detector, making Stop signs disappear from its predictions [10].

Legal Implications of Tricking ML Systems [7]. I am a member of a team of computer scientists and law experts that examines whether attacks, such as the above, constitute as 'hacking' under the CFAA—the primary

anti-hacking law in the US. The goal of this effort is to introduce the law and policy community to how machine learning alters the nature of hacking. One of our contributions is that we identified a misalignment between the early understanding of hacking and today's techniques—this creates ambiguity as to where and how the law applies. For example, how would the law govern behavior that endangers safety, such as manipulating the sensed environment of a car, while tolerating reasonable anti-surveillance measures, such as makeup that foils face recognition? These two cases use similar technical principles, but have dissimilar consequences.

C. *Protecting Users from Cyber-Physical Attacks*

Any new technology that impacts humans will introduce new security risks. Protecting *users* from such risks is an important area to build truly secure systems. I will briefly discuss two projects that focus on users.

Access Control And Authentication for Users [13]. This work re-envisions access control policy specification and authentication ***of humans*** accessing smart home devices, because smart home devices are fundamentally different than classical computers. For example, numerous users interact with a single smart home device, such as a household's shared voice assistant or Internet-connected door lock. Widely deployed techniques for specifying access-control policies and authenticating users fall short when multiple users share a device—current mechanisms are role-based, with roles like admin or guest. However, these roles do not capture the breadth of social relationships among people in homes—mischievous children, parents curious about what their teenagers are doing, and abusive romantic partners. Our work finds that home occupants want to express access control rules in terms of relationships, and the *context* under which actions occur. Another unique aspect in homes is that few devices have screens or keyboards, preventing users from just typing passwords. This project conducts a large scale user study of the desired access control policies in homes, contributes a vocabulary for access control, and puts forth a set of default policies to help bootstrap such systems.

Contextual Integrity for Users [14]. This work incorporates Nissenbaum's property of contextual integrity into smart home applications. Consider an app that opens the windows when the internal temperature is greater than a certain value. If an attacker manipulates the app into opening windows, when say, people are sleeping, the context under which the window opening operation occurs has changed. I helped build ContexIoT, a system that detects such changes and prompts users to authorize an action if it occurs in unfamiliar contexts.

D. *Privacy Issues in Data Analysis Systems*

Software systems produce a lot of data, that is often derived from personal activity. In an increasingly computerized world, new data sources and analyses will lead to actionable insights, but can also lead to privacy violations. For example, if a user prefers a certain thermostat temperature setting when at home, a system that can learn this fact will be able to automatically set the temperature when the user is at a different location (e.g., hotel room). We are seeing an emergence of systems that offer data semantics interpretation services to help users perform such kinds of tasks. However, these systems transmit data, which is often personal, to cloud services for analysis. My study also revealed that currently, very few developers attempt to protect sensitive date because they lack the tools.

Motivated by this analysis, I designed and implemented Appstract [22], [23], a system and a set of algorithms that efficiently and accurately extracts the semantics of textual data (e.g., assigning the semantic label `guitarist` to the string 'Joe Satriani') without transmitting that text to a cloud server. I also helped build Heimdall, a system that lets developers use this semantic data to enable new experiences in a privacy-respecting way [24].

FUTURE WORK

Near Term: Towards End-to-End Secure CPS Middleware. My work so far has examined various security and privacy issues in CPS middleware. Leveraging these results, my goal is to articulate end-to-end security properties that range from protecting users to protecting confidentiality of data. I am interested in two broad themes. 1) Building confidentiality into trigger-action platforms: My work on DTAP introduced strong integrity properties for trigger-action platforms. However, a compromised platform can still passively observe data and leak it. I am interested in exploring how trusted computing and advances in homomorphic encryption can help in formulating strong confidentiality. 2) The intersection of ubiquitous computing and authentication: My work has identified desirable access control policies in homes. However, an open question remains about how such policies might be

enforced. A critical piece in enforcement is to identify humans acting on devices. I plan on leveraging results from the ubiquitous computing community to create novel sensing techniques that will identify humans for authentication purposes.

Medium Term: Security and Privacy using Physical Principles. Computer systems might fail for a variety of reasons, including hackers exploiting vulnerabilities. But, all of these activities create measurable phenomena in the physical world. I am interested in understanding how we can use these physical phenomena to build security and privacy guarantees. For example, a smart oven consumes electricity, a garage door opener makes noise, and a space heater increases temperature. A potential defense is to build systems and algorithms that sense these physical phenomena, and then report when that behavior changes from expectations. Another defense technique is to encode plausible physical behavior into the design of systems. For example, if a truck instantaneously appears in the predictions of a DNN while it was not present an instant before, this is indicative of an attack because its not physically plausible for such a large object to appear instantly.

Long Term: Security and Privacy for Emerging Technologies. As computers become ingrained in every aspect of our lives, the potential for security and privacy issues increases. It is therefore crucial that such new and emerging technologies be subjected to adversarial analyses in the hope of identifying security problems and fixing them at a design level before they become widely deployed. My research has contributed fundamental techniques to improve the security of two existing technologies—smartphones, and cyber-physical systems, and in the long term, I plan to use my systems approach to tackle new challenges that arise from emerging technologies. In general, I am interested in all areas of computer science and beyond that intersect with security and privacy, including computational manufacturing, autonomous systems, and large-scale critical infrastructure.

REFERENCES

[1] "ContexIoT attack code," https://sites.google.com/site/iotcontextualintegrity/home.

[2] "FlowFence code," https://github.com/earlence/FlowFence_Release.

[3] "Robust physical perturbations code," https://github.com/evtimovi/robust_physical_perturbations.

[4] "Samsung smartthings: Capability model improvements," https://smartthings.developer.samsung.com/develop/guides/smartapps/working-with-devices.html.

[5] "SmartThings analysis tools," https://github.com/earlence/SmartThingsAnalysisTools.

[6] "UI deception code," https://github.com/earlence/UIDeceptionBinderChannel.

[7] R. Calo, I. Evtimov, **E. Fernandes**, T. Kohno, and D. O'Hair, "Is tricking a robot hacking?" in *Proceedings of WeRobot*, 2018.

[8] M. Conti, B. Crispo, **E. Fernandes**, and Y. Zhauniarovich, "CRePE: A System for Enforcing Fine-Grained Context-Related Policies on Android," *IEEE Transactions on Information Forensics and Security (TIFS)*, 2012.

[9] M. Conti, **E. Fernandes**, J. Paupore, A. Prakash, and D. Simionato, "OASIS: Operational Access Sandboxes for Information Security," in *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM@CCS)*, 2014.

[10] K. Eykholt, I. Evtimov, **E. Fernandes**, B. Li, A. Rahmati, F. Tramer, A. Prakash, T. Kohno, and D. Song, "Physical adversarial examples for object detectors," in *12th USENIX Workshop on Offensive Technologies (WOOT 18)*. Baltimore, MD: USENIX Association, 2018. [Online]. Available: https://www.usenix.org/conference/woot18/presentation/eykholt

[11] K. Eykholt, I. Evtimov, **E. Fernandes**, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust Physical-World Attacks on Deep Learning Visual Classification," in *Computer Vision and Pattern Recognition (CVPR)*, 2018.

[12] D. B. Giffin, A. Levy, D. Stefan, D. Terei, D. Mazières, J. Mitchell, and A. Russo, "Hails: Protecting data privacy in untrusted web applications," in *Symposium on Operating Systems Design and Implementation (OSDI)*. USENIX, October 2012.

[13] 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 *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, 2018, pp. 255–272. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/he

[14] Y. 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 *21st Network and Distributed Security Symposium*, 2017.

[15] **E. Fernandes**, A. Aluri, A. Crowell, and A. Prakash, "Decomposable Trust for Android Applications," in *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2015.

[16] **E. Fernandes**, Q. A. Chen, J. Paupore, G. Essl, J. A. Halderman, Z. M. Mao, and A. Prakash, "Android UI Deception Revisited: Attacks and Defenses," in *Proceedings of the 20th International Conference on Financial Cryptography and Data Security (FC)*, 2016.

[17] **E. Fernandes**, B. Crispo, and M. Conti, "FM 99.9, Radio Virus: Exploiting FM Radio Broadcasts for Malware Deployment," *IEEE Transactions on Information Forensics and Security (TIFS)*, 2013.

[18] **E. Fernandes**, J. Jung, and A. Prakash, "Security Analysis of Emerging Smart Home Applications," in *Proceedings of the 37th IEEE Symposium on Security and Privacy (S&P)*, 2016.

[19] **E. Fernandes**, J. Paupore, A. Rahmati, D. Simionato, M. Conti, and A. Prakash, "FlowFence: Practical Data Protection for Emerging IoT Application Frameworks," in *Proceedings of the 25th USENIX Security Symposium*, 2016.

[20] **E. Fernandes**, A. Rahmati, J. Jung, and A. Prakash, "The Security Implications of Permission Models in Smart Home Application Frameworks," *IEEE Security and Privacy Magazine*, 2017.

[21] **E. Fernandes**, A. Rahmati, J. Jung, and A. Prakash, "Decentralized Action Integrity for Trigger-Action IoT Platforms," in *22nd Network and Distributed Security Symposium (NDSS)*, 2018.
[22] **E. Fernandes**, O. Riva, and S. Nath, "My OS Ought to Know Me Better: In-app Behavioural Analytics as an OS Service," in *15th Workshop on Hot Topics in Operating Systems (HotOS XV)*, 2015.
[23] **E. Fernandes**, O. Riva, and S. Nath, "Appstract: On-The-Fly App Content Semantics with Better Privacy," in *Proceedings of the 22nd ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2016.
[24] A. Rahmati, **E. Fernandes**, K. Eykholt, X. Chen, and A. Prakash, "Heimdall: A Privacy-Respecting Implicit Preference Collection Framework," in *15th ACM International Conference on Mobile Systems, Applications, and Services*, 2017.
[25] A. Rahmati, **E. Fernandes**, K. Eykholt, and A. Prakash, "Tyche: A risk-based permission model for smart homes," in *Proceedings of the 3rd IEEE CyberSecurity Development Conference (SecDev)*, 2018.
[26] A. Rahmati, **E. Fernandes**, and A. Prakash, "Applying the Opacified Computation Model to Enforce Information Flow Policies in IoT Applications," in *Proceedings of the 1st IEEE CyberSecurity Development Conference (SecDev)*, 2016.
[27] G. Russello, M. Conti, B. Crispo, and **E. Fernandes**, "MOSES: Supporting Operation Modes on Smartphones," in *Proceedings of the 17th ACM Symposium on Access Control Models and Technologies (SACMAT)*, 2012.
[28] G. Russello, B. Crispo, **E. Fernandes**, and Y. Zhauniarovich, "YAASE: Yet Another Android Security Extension," in *3rd IEEE Conference on Privacy, Security, Risk and Trust (PASSAT)*, 2011.
[29] D. Stefan, E. Z. Yang, P. Marchenko, A. Russo, D. Herman, B. Karp, and D. Mazières, "Protecting users by confining JavaScript with COWL," in *Symposium on Operating Systems Design and Implementation (OSDI)*. USENIX, October 2014.
[30] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *International Conference on Learning Representations*, 2014. [Online]. Available: http://arxiv.org/abs/1312.6199
[31] Y. Zhauniarovich, G. Russello, M. Conti, B. Crispo, and **E. Fernandes**, "MOSES: Supporting and Enforcing Security Profiles on Smartphones," *IEEE Transactions on Dependable and Secure Computing (TDSC)*, 2014.