In my first year as a PhD student, I have primarily focused into analyzing security issues by way of Internet-wide measurements. I have worked on three projects in this space. One paper is currently in submission, but the other two were accepted at WWW'17 (17% acceptance rate, first-author) and USENIX Security '17 (16% acceptance rate, second-author). Moving forward, I am especially excited about the perspective that Internet-wide measurements can provide in tackling problems that cause societal harm. During the remainder of my PhD, I plan to focus on mitigating the effects of large-scale disinformation campaigns on the Internet and social media, which have the dangerous potential to influence and persaude the mind of the public. I believe my experiences up to this point position me well for continued success in graduate school and ideally as a future faculty member. I am seeking an NSF-GRFP to further support me in those goals.

Security Challenges in an Increasingly Tangled Web: Motivated by anecdotes documenting the struggles with HTTPS adoption and the challenges with advertisers unwittingly serving malware, I set out to study the security challenges faced in an increasingly complex web. To do this, I instrumented Headless Chrome, which is a non-interactive version of the Chrome browser, to capture all dependencies loaded by websites. I then ran this tool over the Alexa top million websites, a set of domains that represent the top million most popular websites in the world. I used this resultant dataset to quantify the increase of complexity on the web over time, and study how this complexity impacts both HTTPS adoption and trust. This was my first Internet-measurement project, and came with its own set of a challenges. The biggest challenge for me was wrangling the size of the data we were collecting; this was the first project where I had to analyze and gain insight from terabytes of data. Nevertheless, the project was exhilarating, as we were able to identify pain points for both HTTPS adoption and malicious advertising, due to an increased opacity on the web. I support reproducible and accessible research, and to this end, I open sourced my code and my data for this project, which are available at `https://github.com.com/zbrowse` and `https://scans.io/tangled`. It was in completing this project that I first understood the value of measurement at scale, and to this day, this experience focuses the lens by which I view new projects. This work led to a first-author paper at WWW '17 (17% acceptance rate), and I presented the paper at the conference in April 2017 in Perth, Australia [1].

Understanding the Mirai Botnet: The Mirai botnet captured the attention of the security community when it was deemed responsible for a spree of the largest distributed denial-of-service attacks (DDoS) on the public Internet. In addition, it surprised the security community due to its naive attack vector—it simply attempted a set of hardcoded default passwords to amass a botnet of over 300K machines. Our research group set out to understand the rise, growth, and evolution of Mirai, and leveraged many distinct datasets from several organizations. We worked with many industry and academic partners which resulted in a collaboration of 19 researchers from seven organizations. I am lucky to be part of a research group that has so many industry relationships; only through their data perspectives could we have analyzed the entire picture of the botnet.

I was primarily responsible for tracking the evolution of the botnet. I deployed interactive honeypots to collect Mirai malware samples from the Internet and extracted new features that appeared in the binaries. I was surprised to observe how quickly Mirai bots could find my honeypots on the Internet—the first infection happened just 12 seconds after deployment. Next, I augmented this dataset with malware from VirusTotal, a crowd-sourced malware repository. One particularly interesting feature I extracted from each binary was the set of passwords it leveraged to attack new systems. We were able to study the evolutionary pattern of the malware by observing changes in

the password dictionaries—over time, the number of passwords and complexity of passwords grew in size, indicating an adaptive and uninhibited botmaster. This was my first exposure to measuring an instance of Internet abuse. I am fascinated by the devastating effects Internet abuse can have on real systems, and hope to study other flavors of Internet abuse in future work. This project resulted in a second-author paper at USENIX Security Symposium '17 (16% acceptance rate) [2].

Tracking Certificate Misissuance in the Wild: In the last year alone, there have been a number of high-profile cases where Certificate Authorities (CAs) were caught misbehaving, often in spectacular ways. As a result, the public key infrastructure (PKI) community published a number of standards to improve the transparency, reliability, and security of the ecosystem. Unfortunately, the community failed to build tools to systematically analyze the effects of these standards. With colleagues from the University of Michigan and the University of Illinois, I worked to bridge this gap by building ZLint, an open-source system that checks whether an *X.509* certificate is conformant to community standards. After building these tools, I set out to investigate the state of certificate misissuance in the wild. I found that certificate misissuance has drastically decreased in the last four years, indicating a positive trend for the ecosystem. Unfortunately, I uncovered a long tail of actively misissued certificates, indicating much more work to be done. When analyzing the CAs who issue these misissued certificates, I found evidence of two distinct classes of problematic CAs. On the one hand, there are small players that consistently misissue every certificate. These players are dangerous due to their disproportionate power—the way the PKI works, any small player can sign a trusted certificate for any entity on the Internet. The other class of CAs were large players that instead misissue the largest number of raw certificates. Luckily, many of these players are currently being scrutinized by major browser vendors. ZLint was recently integrated into Censys and `crt.sh`, two popular tools used by the PKI community to track and analyze certificates found in the wild. I am the lead developer on the ZLint project, which is available at `https://github.com/zmap/zlint`. ZLint is actively used by the PKI community to track and remedy misissuance in the wild—and further, motivated members of the community are engaging with the project on Github to add new, useful features. I am the first author on a paper that documents our full analysis of misissuance in the certificate ecosystem, which is currently in submission at a top-tier security venue.

Teaching Experiences There is no feeling better than leading a student to an "aha" moment, when they finally realize the intuition required to understand a new concept. As a result, my formal teaching experiences have been some of the most rewarding. During my last year as an undergraduate at the University of Michigan, I served as a teaching assistance for EECS 388, the introductory security course. As a TA, I was responsible for curating new projects and assignments, leading a weekly discussion section (a group of 30-40 students), and holding office hours to help students understand course material. I was most fulfilled when interacting with students, so much so that I would often hold extra office hours or extend my discussion sections just to help students out. I continued my teaching experiences as a graduate student at the University of Illinois, where I served as a head graduate teaching assistant for CS 461, the introductory security course at UIUC. I was surprised by the new challenges I faced at UIUC—despite the students being approximately at the same level of education at both universities, my teaching style needed to drastically change due to nuanced differences in teaching ideologies at the two universities. Certainly, my teaching methodology will continue to evolve as I encounter a wider breadth of students. Ultimately, I hope to lead a course as an instructor at some point during my graduate studies.

Broader Impact: Since my second year of undergrad, I have led a number of projects that seek to improve the awareness and excitement around innovation in technology. This started In 2014, when I co-founded Project Cinta, a platform to connect motivated high-school students with college entrepreneurs. Our intuition was that high school students often have great insight, but lack guidance from elder students and mentors to help those projects come to fruition. We launched a pilot program at Saline High School in Saline, Michigan in the Fall of 2014, which ultimately led to three successful collaborations with students at the University of Michigan.

In the summer of 2014, I was honored to be named a hackNY fellow. hackNY is an organization that "aims to federate the next generation of hackers for the New York innovation community". During my time at hackNY, I volunteered at workshops sponsored by Girls Who Code, an organization dedicated to closing the gender gap in technology. It was fulfilling to contribute to this organization, even in a small way. In addition, I am involved in the larger CS graduate student community here at UIUC, where we are working with undergraduate women in CS to help connect them to women graduate-students to increase interest in graduate school.

During the summer of 2016, I served as a technical counselor for the ICOS Big Data Bootcamp, a program taught at the University of Michigan. The goal of the program is to introduce non-STEM graduate students to introductory concepts in data processing that will translate to transferable skills in their research. I helped graduate students in sociology, psychology, and linguistics work with data models specific to their research, and taught them basic programming techniques in Python. I still keep in touch with several of these students and the program organizers, and have maintained those connections as future potential research collaborators.

Future Directions: I am fascinated by how Internet-scale abuse can cause societal harm. A recent example of this is disinformation campaigns on social media, which have garnered recent media attention due to their purported role in the 2016 U.S. election. Researchers have preliminarily started researching this phenomena, and noticed that many of these campaigns appear to be *automated* and exhibit signs of botnet-like behavior. Given my prior experiences in leveraging Internet-wide measurements to study botnets and other forms of abuse, I believe we can gain a new perspective into this ecosystem by analyzing disinformation campaigns through the lens of Internet abuse. I feel my prior experiences with Internet-measurement and working with industry partners position me well to tackle this problem, as it will require a large collaboration with many involved parties. Outside of research, one problem I have observed in my first year of graduate school is a lack of community within the CS department. I have joined as a board member of the CS Graduate Student Organization (CSGSO) to start building a community, and am leading the charge on a variety of projects. I hope to instill a sense of community in the department and inspire future graduate students to do the same. Ultimately, I hope to be a professor, giving back to students and training the next generation of technologists, researchers, and professors. Support from the NSF-GRFP would be invaluable to this goal.

References [1] D. Kumar, Z. Ma, Z. Durumeric, A. Mirian, J. Mason, J. A. Halderman, and M. Bailey. Security challenges in an increasingly tangled web. [2] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou. Understanding the mirai botnet.