Research Statement

My research interests interests lie in exploring the security, performance, and availability properties of computing systems. My work seeks to both fundamentally enhance our understanding of the sciences underlying the computing subfields of security and privacy as well as to inform the development of these systems to address immediate, important societal problems.

Enhancing the Sciences of Security and Privacy

I have sought to enhance the reputation of the networking and security fields as true sciences by directly and specifically demonstrating that these fields satisfy a variety of concrete scientific criteria including:

Reproducible results. Research into Internet-wide or infrastructure-level attacks are hampered by a lack of Internet-wide datasets. As part of the DHS PREDICT and IMPACT projects, I helped pioneer new frameworks that provide researchers with representative datasets for the reproducible evaluation of their work and shared over 100 TB of networking security data [32, 34, 54]. Most recently, my colleagues and I introduced Censys, a public search engine and data processing facility backed by data collected from ongoing Internet-wide scanning. Censys has 30,000 users who have performed 20 million auditable actions. 100+ academic papers have directly used this data [17].

Well-defined experimental methods. Empirical measurement at-scale often requires the creation of new methodologies and instruments. As an example, my early research focused on a scalable hybrid network monitoring architecture for measuring, characterizing, and tracking a broad spectrum of Internet threats [20, 28, 35, 40, 41, 42, 43, 56]. At its peak, our instrument consisted of more than 60 distinct monitored blocks at over 25 organizations across the Internet and was actively used by Internet Service Providers as an early warning system for worms, denial of service, scanning, and misconfiguration behaviors at-scale. More recently, my work has explored active measurement and documented how Internet-wide, fast scanning can be used to measure phenomena at-scale. Each well-defined experimental methodology has been used in hundreds of additional experiments to date.

Scientific integrity. Early networking and security research evolved without significant concern for human subjects, leading to instances where ethical considerations were either absent because researchers failed to understand their relevance, or lacked any standards for assessment, accountability, or oversight. I have been a leader in helping define these standards [47, 50, 65, 64]. For example, I was a member of the Menlo Report working group that created a community document, inspired by the Belmont Report, that provides a framework for ethical guidelines for computer and information security research. Along with my colleagues, I built tools, such as the ethical impact assessment (EIA), that can be used to help computer science researchers evaluate the ethical impact of their work.

Use Inspired Security and Privacy Research

In addition to fundamentally enhancing our understanding of the sciences underlying the computing subfields of security and privacy, my work informs the development of these systems to address immediate, important societal problems including:

Network and Systems Security. Annualized traffic volumes on the public internet are expected to reach 4.8 ZB by 2022, but this is challenged by literally millions of network attacks with volumes that may exceed TB per second. My primary contribution to network security practice has been in understanding and defending the network against such attacks. Early work exploring the mechanisms of attackers included work in virus and malware detection [27, 38, 39, 51], analysis of self-propagating code and Internet worms [43, 65], and modern botnet infrastructures [7, 53]. I explored attacker behavior seen by the attacker's infrastructure including scanning [23, 36], spam [37], and denial of service attacks [11, 19, 22, 57]. Beyond simply measuring and characterizing these threats, I have been involved in the creation of deployed infrastructures for their direct mitigation. For example, as director of engineering at Arbor networks (now NETSCOUT, NASDAQ: NTCT) I led a team of engineers who built Peakflow SP; a distributed, non-intrusive, scalable DDoS protection product. As of Q3 2013, these tools monitored roughly 70 terabits per second (Tbps) of global Internet traffic, an estimated one third of the total, average, global Internet bandwidth at the time.

Web Security. The world Wide Wide Web (WWW) has become a critical component of our everyday lives, impacting everything from how we read the news to how we buy groceries. Recent estimates place the number of user at 3.5b with at least 5b indexed pages. As its importance and size continues to grow, so does the need to assure the trustworthiness of this critical resource. A major thrust of my work over the last five years has been in assuring the security and privacy of the WWW [1, 6, 8, 9, 10, 21, 29].

Three recent pieces of work stand out in this research thrust as significant examples of my research style and impact. In the paper entitled "The Matter of Heartbleed" which appeared in the Proceedings of the 14th ACM SIGCOMM Conference on Internet Measurement (IMC '14) [21] we explored the impact and community response to one of the most troubling vulnerabilities in the Internet security ecosystem in the last several years. The Heartbleed vulnerability, which represented a flaw in the cryptographic protocol that guarantees confidentiality and integrity on the WWW, took the Internet by surprise in April 2014. Using extensive active scanning, my team assessed who was vulnerable, characterizing Heartbleed's scope across popular HTTPS websites and the entire IPv4 address space. We conducted a global notification effort, in which they emailed the abuse contacts for all networks containing vulnerable hosts, helping to patch over 250,000 Internet hosts. In addition to its direct impact, perhaps the most interesting lesson from the study of Heartbleed is the surprising impacts that these direct notifications of network operators can have on patching. These results informed our fundamental understanding of patching behavior and led to a follow on paper at USENIX Security '16, as well as an NSF large grant with research collaborators. For our efforts, this paper was given a best paper award at the IMC conference.

The world wide web relies on a public key infrastructure (PKI) to anchor trust in the websites visited (i.e., attesting, for example, that amazon.com is really Amazon, Inc.). Certificate Authorities (CAs) which maintain this PKI, regularly make mechanical errors when issuing certificates. In the paper entitled "Tracking Certificate Misissuance in the Wild" which appeared in IEEE Security & Privacy 2018 (S&P '18) [6], we set out to measure, model, and correct these errors. To quantify these errors, we built ZLint, a certificate linter that codifies the policies set forth by the CA/Browser Forum Baseline Requirements and RFC 5280 that can be tested in isolation. We found that although certificate misissuance has drastically reduced since 2012, there is a long tail of small CAs that regularly misissue the majority of their certificates. ZLint was released as an open source project and is being rolled out in the most important CAs, such as Digicert, GlobalSign, and Let's Encrypt. Over 90% of certificates on the public Internet are now validated via this tool.

The final example I want to highlight involves the threat of in-browser cryptojacking, which entails hijacking of a victim's processing power to mine cryptocurrency without consent. The financial success of cryptocurrencies, which are valued at over 280 billion USD in market capitalization as of 2018, has drawn the attention of malicious actors, who attempt to accumulate wealth through illegitimate means. In the paper "OUTGUARD: Detecting In-Browser Covert Cryptocurrency Mining in the Wild" which appeared at the Web Conference 2019 (formerly WWW) [1], we discuss explore this cryptojacking threat. In our work we build an open-source cryptojacking detection system that we deployed in the wild enabling us to discover 6,302 cryptojacking sites; predominantly in the long tail of lower popularity sites that promote illicit content such as torrents or copy-righted video streaming. While other highlighted work focuses on the security of web protocols or the trustworthiness of the PKI, this paper highlights the challenges faced by abuse of web browsers beyond traditional drive by downloads. For our efforts this work was awarded best paper.

Internet of Things. Internet of Things (IoT) represents an unprecedented integration of sensing, computation, storage, communication, and control into physical systems. By 2020 experts estimate that there will be 20.4b connected things, $3t in hardware spending, and that 65% of enterprises will adopt IoT. However, a lack of technical maturity, and more subtly, designs based on incorrect models and assumptions, has given rise to a variety of threats to the confidentiality, integrity, and availability of IoT devices. Assuring the security of these new devices has been a new area of work for my team [4, 7, 60].

A first foray into this new domain is represented by the work entitled "Understanding the Mirai Botnet" which appeared in the 26th USENIX Security Symposium (USENIX Security '17) [7]. This work is among the first papers to document the dangers of IoT devices at scale by tracking the Mirai botnet. The Mirai botnet, composed primarily of embedded and IoT devices, shocked the internet when it was response for several high profile denial of service attack in late 2016 and 2017 (e.g., twitter, spotify). We provided a seven-month retrospective analysis of Mirai's growth to a peak of 600k infections and a history of its DDoS

victims. As the IoT domain continues to expand and evolve, Mirai has served as a call to arms for industrial, academic, and government stakeholders concerned about the security, privacy, and safety of an IoT-enabled world. The paper is one of the top 100, normalized by year, in computer security with over 150 citations in its first year (1000%+ over average).

The second of these new pieces of work is the paper entitled "Skill Squatting Attacks on Amazon Alexa" which appeared in 27th USENIX Security Symposium (USENIX Security '18) [4]. The proliferation of the Internet of Things has increased reliance on voice-controlled devices to perform everyday tasks. Although these devices rely on accurate speech recognition for correct functionality, many users experience frequent misinterpretations in normal use. In this work, we conducted an empirical analysis of interpretation errors made by Amazon Alexa and found that attackers can use these errors in unexpected ways. This work and its discussion of the security implications of speech interpretation errors have resulted in changes to popular systems, such as Amazon Echo.

Cyber Incident Prediction. Data breaches, such as those at Equifax, Target, JP Morgan, and Home Depot highlight the increasing social and economic impact of cyber incidents. Breaches cost on the average $2.6m per organization and over $100 per record. Cyber incident prediction enables the development of effective risk management schemes such as cyber insurance, which introduces monetary incentives for the adoption of better cyber security policies and technologies. In the wake of recent breaches, the market for such policies has soared, with current written annual premiums estimated at $2b and growing quickly (est. $14b in 2022). I have worked extensively in this area [18, 25, 45].

Perhaps the most indicative work in this research area is our paper entitled "Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents" which appeared in the proceedings of the 24th USENIX Security Symposium (USENIX Security '15) [18]. In this work we characterized the extent to which cyber security incidents can be predicted based on externally observable properties of an organization's network. Our method, based on 258 externally measurable features collected from a network's mismanagement symptoms and malicious activities and can achieve a 90% true positive rate with a 10% false positive rate. This work, as well as papers preceding it, have made up a patent portfolio (e.g., US Patent 9,729,558, US Patent 10,038,703) licensed by FICO as a basis for their FICO Cyber Risk Score in an effort to provide "empirical cyber risk scoring, prudent disclosure of contributing risk factors, sound model governance practices, and the enablement of direct client involvement in the resolution of data and definitional issues."

Other Areas For breveity, I have highlighted only a few of the active research areas over the last five years. Other topics not covered include Cloud Computing [26, 36, 48, 62], IPv6 [15, 24, 28], USB [5, 14, 61], cryptocurrencies [1, 2, 3], and others.

Peer Assesment

I am the author of 80+ published works that have been cited over 5500 times (h-index of 36). The last five years (2014-2018 inclusive) have been a particularly productive part of that total with 29 published works and 3,400 citations (h-index since 2014 of 28). My efforts have won serveral research specific accolades (e.g., University of Michigan Research Faculty Recognition Award, Kenneth M. Reese Outstanding Research Scientist Award). During the during the last five years, specifcially, I was awarded the Google Security and Privacy Research Award in 2016, the IETF/IRTF Applied Networking Research Prize in 2015, and best paper awards at the Internet Measurement Conference in 2014 and the Web conference (formerly WWW) in 2019.

Conclusion

In conclusion, my research focuses on the security and availability of complex distributed systems and has informed not only the development of such systems, but also the sciences of computer security, networking, and distributed systems. The impacts of this work are not only visible in my scholarly publications, but also through the transitioning of new technologies into practice.