# Research Statement

Alexander Gamero-Garrido

My research focuses on the intersection of computer networking systems and public policy, with emphasis on online privacy, and yields empirical evidence on contentious questions in technology policy. I use conventional methods from both computer networking (large-scale Internet measurements) and the social sciences (surveys) to conduct empirical research on Internet technologies and related policy, cybersecurity and privacy. My research focuses on three areas: ***(i)*** capabilities for government surveillance created by the macroscopic structure of the Internet; ***(ii)*** Internet policy issues in the United States; ***(iii)*** commercial surveillance enabled by consumer devices. My research in these areas is published in Association for Computing Machinery (ACM) conferences on both cybersecurity [CCS '17] and networking [SIGCOMM '18, IMC '21]. My studies have won multiple awards at networking conferences.

## Surveillance Enabled by Transit Networks

My dissertation research identified nations exposed to mass surveillance and selective tampering of consumer Internet traffic by transit networks, which provide connectivity to consumer-serving Internet service providers (access networks), rather than to consumers directly. As such, they are often unaccountable to consumers and, thus, create opportunities for government surveillance and disinformation campaigns without facing political backlash. A key challenge is the dearth of data on inter-operator connectivity: proprietary *Border Gateway Protocol* (BGP) announcements, which operators use to exchange routing information, are publicly disclosed by select networks. Incomplete BGP data hinders the inference of networks that are present on Internet routes toward a country. To tackle this challenge, I built a methodological framework with several new metrics. A central component of this framework is the *Country-Level Transit Influence* (CTI) metric that filters noise caused by BGP data biases. CTI is built on graph theory and measures the capabilities of a network to observe or tamper with a country's traffic. Applying this research, I identified 34 nations—primarily in marginalized regions of Africa, Latin America, and Asia—that are particularly vulnerable. These nations have a single transit network that serves over 40% of the nation's IP addresses, creating a centralized surveillance point with a large footprint on traffic (potentially exposing unencrypted flows and metadata). Funded by a Microsoft Research Dissertation Grant, my study was published in a networking venue [PAM '22], where it won *Best Dataset Award*.

My ongoing work in this area is focused in two directions. First, to improve the reliability of Internet connections in low-income countries, the **Internet Society** (a global nonprofit) is incorporating a metric from my CTI study into the *Internet Resilience Index*, a measure of the resiliency of country-level networks (broadening my study's impact). For this purpose, I produced a longitudinal analysis of the *peering readiness* of U.N. Member States: whether the infrastructure in each country enables interconnection between domestic and foreign networks. Further, the **International Telecommunication Union** (a specialized **U.N. agency**) has shown interest in my research and we are discussing future collaborations. Second, I work with political scientists to quantify structural differences in country-level networks between authoritarian regimes and democracies. My paper [IMC '21] identifies state-owned network operators globally and is a key input in our current research, which quantifies the footprint of each state-owned conglomerate on both access and transit networks. Findings suggest that authoritarian states use transit networks rather than access networks to build surveillance capabilities (in democracies, state-owned operators are not dominant). We are preparing the manuscript describing this work for **submission to *Science*** [PrePrint-Sci].

## U.S. Internet Policy I: Telecommunications

Episodes of persistent congestion that **degrade performance of end-user applications** may result from under-provisioned links between telecommunications operators (links at which network traffic demands routinely exceed available capacity). The goal of my study on interdomain congestion [SIGCOMM '18] (awarded *Best Paper*) is to provide **empirical grounding on network neutrality disputes in the U.S.** Access networks (*e.g.*, Comcast) often demand payment for delivering content providers' (*e.g.*, Netflix) large traffic volumes and infrastructure upgrades, which content providers argue consumers already pay for, rendering any additional payment unnecessary. This study tackled three challenges in inferring instances of persistent congestion: ***(i)*** effectively identifying latency differentials in router-level links between operators (interdomain links) at the scale of U.S. broadband providers; ***(ii)*** gathering evidence of performance degradation experienced by users as a consequence of these recurring congestion events (for example, additional data collection showing packet loss during the event); ***(iii)*** longitudinally collecting both of these large-scale datasets (from several months to two years) to evaluate trends of localized congestion between specific access networks and content providers.

Leveraging large-scale, longitudinal measurements from U.S. access networks, this study [SIGCOMM '18] did not find evidence of widespread congestion on links across networks, suggesting that access networks and content providers have resolved most sustained disputes (though there were instances of severe congestion between specific networks). In a previous, longitudinal study on U.S. residential broadband [TPRC '15], I argue that issues with infrastructure operated by both access and content providers can cause performance degradation events perceived by users. These causes include instances of localized congestion between specific access networks and content providers.

## U.S. Internet Policy II: Cybersecurity

For my study published in a cybersecurity venue [CCS '17], I gathered empirical evidence of the "chilling effect" imposed by U.S. **federal regulations** on **cybersecurity research**. These policies may **disincentivize necessary security audits** of consumer products by external researchers and render these **products less secure**. Since most vulnerabilities are identified in this manner, audit deterrents can cause broad harms to consumers (for example, enabling large-scale exfiltration of sensitive data by malicious actors). The regulations, including the Digital Millennium Copyright Act (DMCA), enable persecution of researchers who identify vulnerabilities in commercial software. I investigated whether companies expressly permitted security audits of their consumer products. To accomplish this, I led an experiment where real researchers requested permission to audit the security features of popular U.S. consumer products sold by 75 large companies, and analyzed the coded responses across multiple communication modes. My study's findings show that companies generally do not permit audits and are averse to waiving their rights for legal recourse if a vulnerability is disclosed.

With a survey of security researchers, my study also shows that legal issues associated with vulnerability discovery impact researchers' decisions about whether to investigate specific products. The survey's principal instrument is a Likert scale (or rating scale) on respondent agreement with reasons *not* to audit a target: "legal challenges" and four distractors. The study shows that *legal challenges* are a more pressing issue for security researchers than four other factors (distractors) that might reasonably impact product selection. A substantial minority reported personally facing legal threats from companies in response to security audits or vulnerability disclosures. In aggregate, my study's findings suggest that U.S. federal regulations likely discourage vulnerability discovery in consumer products.

## Online Privacy and Commercial Surveillance

My current work on online privacy explores how effective the General Data Protection Regulation (**GDPR**), an EU regulation, is in limiting data transfers to third countries (**data localization**). Such transfers **expose EU users** to (potentially unlawful) **commercial surveillance of consumer traffic by advertisers**. Despite these potential harms, little is known about whether or how companies operate their infrastructure to comply with the GDPR. I tackle this question by empirically measuring the extent to which servers that process EU requests are located outside of the EU. The key challenge is that both browser measurements (to infer relevant *endpoints*) and data-plane measurements (to infer relevant *IP addresses*) are needed, but no large-scale public infrastructure allows both. I built a novel methodology that combines browser and data-plane measurements from separate platforms (at a scale that is representative of the continent-sized EU jurisdiction). The study's findings show that an important fraction of popular-website servers that collect data from EU users are located in the U.S. and Russia. These results raise concerns regarding EU data sovereignty and question the effectiveness of data localization policies for safeguarding consumer privacy. This study, funded by a Northeastern Future Faculty Fellowship, is currently under submission [PrePrint-Privacy].

## Future I: Algorithmic Discrimination by Voice Assistants

I am expanding my research vision in the realm of commercial surveillance by focusing on voice assistants, which pose a particular privacy risk given the **breadth of information** they capture and upload to servers. Once recorded, this information is potentially **vulnerable to observation** by the product vendor, attackers who may steal the information, and surveillance by law enforcement. Since voice assistants are operated by proprietary algorithms, these platforms may also **reinforce discrimination against marginalized communities**, for instance, by exposing voice recordings of U.S. Spanish-speakers at higher rates than the baseline of users who speak only English. To evaluate these potential biases, I plan to investigate whether consumer devices expose private information of marginalized groups at higher rates than the general population by measuring differences in voice assistant behavior across national origin, race, and gender. These differences would reveal biases in algorithms with harmful consequences for historically marginalized groups in the U.S.

This future direction fits within my existing theme of commercial surveillance by voice assistants for targeted advertising. My co-authored study in this area is under submission [PrePrint-Security] and was featured in the popular press (*Chicago Tribune*, *The Verge*) and presented at the FTC's *PrivacyCon*. I plan to write an **NSF *Secure and Trustworthy Cyberspace* (SaTC) proposal** to investigate these questions of consumer privacy and algorithmic discrimination (2023).

## Future II: Internet Connectivity at U.S. Anchor Institutions

I will broaden my research agenda on U.S. Internet policy to illuminate Internet connections at *anchor institutions*: community-serving organizations that provide **critical services**, including those in **education and healthcare**. The COVID-19 pandemic accelerated the digitization of an ever-increasing number of critical sectors, which brought attention to inadequate institutional connections in the U.S. In such institutional settings, **non-performant Internet access** may have **adverse consequences** on key outcomes like educational attainment. These issues are broadly consequential in ***digital equity*** **as institutional networks often supplement or replace residential connectivity** for marginalized communities. These communities include rural and tribal areas (where high-speed broadband is often unavailable) along with low-income urban areas (where broadband affordability is frequently an issue). Despite these challenges, little is known about the Internet connectivity of U.S. anchor institutions, as most computing studies focus on residential connections.

My research will fill this gap by designing Internet measurement techniques that enable mapping and characterizing the Internet connections at anchor institutions along three dimensions: *availability*, *reliability* and *performance*. Beyond providing a single point of connectivity to the network, well-connected anchor institutions with one or more redundant access links offer particularly robust connectivity. For instance, a school with a reliable Internet connection may be able to continuously provide educational services during a temporary service outage using a backup link, whereas another school with a less reliable connection would suffer from educational downtime. Further, modern applications often hinge on high-quality connectivity *from each device*, with strict requirements on latency and throughput; anchor-institutional networks with under-provisioned connections are, therefore, unable to adequately serve their users. Over time, these connection inadequacies can create serious disadvantages for citizens that rely on anchor institutions for essential activities, such as completing homework and submitting job applications.

To quantify connection quality at anchor institutions, I will use large-scale Internet measurements to identify relevant network endpoints and links. My initial research in this area proposes a *reverse IP geolocation* technique to identify institutional networks in tribal areas from physical addresses and network measurements (Ford Foundation Fellowship); tribal areas are often both rural and low-income and, thus, likely to suffer from insufficient connectivity (low *availability*). These methods need to be scaled up to the entire country, which will require advances in automatically labeling networking resources such as domains.

In further stages of this research, I will estimate the *reliability* of institutional networks by developing mapping techniques at fine granularities: most anchor institutions lease connectivity from external providers, rather than operating an autonomous network. Consequently, my research will map router-level links within and across operators, infer their failure rate, and estimate their capacity (correlates with *performance*). For anchor institutions with poor reliability or performance, my research will identify operators present in the area that could build additional infrastructure (for example, following government investment). To broaden this work's impact and relevance for policy, I collaborate with the Schools, Health & Libraries Broadband Coalition (SHLB), a national nonprofit that advocates for improved broadband access at anchor institutions. This organization will support my **NSF *Networking Technology and Systems* proposal** (*NeTS)*, with planned submission in Dec. 2022.

## Concluding Remarks

Emerging Internet technologies will have consequences on broader society at a scale that is difficult to overstate. Public policy on consumer privacy and digital equity will continue to be a central societal concern. While computer scientists are not directly tasked with drafting legislation, our research on these technologies can and should provide valuable input in the policymaking process (in the U.S. and beyond). My research agenda shows that our discipline's technical expertise (on computer networks, cybersecurity, and online privacy) is ideally positioned to illuminate questions at the broad intersection of Internet technologies and public policy.

## References

[PAM '22] **Alexander Gamero-Garrido**, Esteban Carisimo, Shuai Hao, Bradley Huffaker, Alex C. Snoeren, and Alberto Dainotti. 2022. Quantifying Nations' Exposure to Traffic Observation and Selective Tampering. In *Passive and Active Measurement*. Springer.

[CCS '17] **Alexander Gamero-Garrido**, Stefan Savage, Kirill Levchenko, and Alex C. Snoeren. 2017. Quantifying the Pressure of Legal Risks on Third-party Vulnerability Research. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*.

[TPRC '15] **Alexander Gamero-Garrido**. 2015. Characterizing Performance of Residential Internet Connections Using an Analysis of Measuring Broadband America's Web Browsing Test Data. In *Telecommunications Policy Research Conference*.

[IMC '21] Esteban Carisimo, **Alexander Gamero-Garrido**, Alex C. Snoeren, and Alberto Dainotti. 2021. Identifying ASes of State-Owned Internet Operators. In *ACM SIGCOMM Internet Measurement Conference (IMC)*.

[SIGCOMM '18] Amogh Dhamdhere, David D. Clark, **Alexander Gamero-Garrido**, Matthew Luckie, Ricky K. P. Mok, Gautam Akiwate, Kabir Gogia, Vaibhav Bajpai, Alex C. Snoeren, and kc claffy. 2018. Inferring Persistent Interdomain Congestion. In *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM) Conference*.

[PrePrint-Privacy] **Alexander Gamero-Garrido**, Kicho Yu, Sumukh Vasisht Shankar, Sindhya Balasubramanian, Alexander Wilcox, and David Choffnes. Empirically Measuring Data Localization in the EU. Under submission, *privacy venue*.

[PrePrint-Security] Umar Iqbal, Pouneh Nikkhah Bahrami, Hao Cui, Rahmadi Trimananda, **Alexander Gamero-Garrido**, Daniel Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, Zubair Shafiq. Your Echos are Heard: Tracking, Profiling, and Ad Targeting in the Amazon Smart Speaker Ecosystem. Under submission, *cybersecurity venue*.

[PrePrint-Sci] Eda Keremoğlu, Nils B. Weidmann, **Alexander Gamero-Garrido**, Esteban Carisimo, Alberto Dainotti, Alex C. Snoeren. Network Topology Facilitates Internet Traffic Control in Autocracies. *Science*, planned submission in Dec. 2022.