Esteban Carisimo | **Research** Statement

Introduction

Originally conceived as an academic tool, the Internet has evolved into the largest engineered system ever built, spanning all continents and becoming a fundamental pillar of modern life. Addressing the challenges of a network serving billions of users requires approaching the Internet from two critical perspectives: as **a technical system** and as **an integral part of society**.

Recognizing these dual perspectives, my research focuses on the intersection of the Internet's technical infrastructure and its societal implications. **I employ Internet measurements to map critical infrastructures and assess their implications for resilience, user-perceived performance, and exposure to security threats**. Leveraging public Internet measurement platforms for experimentation, public measurement datasets, and proposing new methodologies and tools [1, 2, 3], I create detailed multi-layer views of the Internet as a basis for my analysis.

The multidisciplinary nature of my research integrates empirical network measurement data analysis with socio-political considerations. I have collaborated with 38 co-authors from 11 institutions worldwide, affiliated with departments including computer science and social sciences. The interdisciplinary part of my work focuses on the role of governments on the Internet, ranging from how they serve digital content (§2.3) to policies for developing critical infrastructures (§3.2 and §2.1) and the use of shutdowns to censor the Internet during coups and protests (§2.2).

Throughout my career, **my work has appeared multiple times in top-tier venues such as ACM SIGCOMM, ACM IMC, and ACM SIGMETRICS**, where I have published **first-author papers in all these prestigious conferences**. Of my 14 publications, I led the research on half of them, resulting in first-author papers. These works have received multiple distinctions, including **the PAM Best Dataset Award** [4], and have been invited for presentation at various technical forums such as the APNIC Blog, the LACNIC/LACNOG Conference, and the Lawfare Podcast.

In the following paragraphs, I outline my research contributions and future plans across three interconnected themes: (1) **Government Influence on Internet Architecture and Access** (§2), (2) **Critical Internet Infrastructures and Global Connectivity** (§3), and (3) **Evolution of Content Delivery Networks and Network Performance** (§4). I also discuss the advanced tools and methodologies I have developed and conclude with my future research agenda.

My ultimate goal is to design strategies for creating a robust, high-performing, and secure Internet architecture that lives up to society's expectations. By continuing to explore the complex interplay between technical systems and societal factors, I aim to contribute to an Internet that is more resilient, secure, and capable of meeting the evolving needs of users worldwide.

Government Influence on Internet Architecture and Access

The Internet has become integral to nearly every aspect of our lives, including government operations and services. Governments play a key role in shaping Internet architecture and access through policies, infrastructure investment, and regulatory frameworks. My research examines how government actions influence Internet development, access, and control.

State-Owned Internet Providers and Market Dynamics

An important aspect of **my research examines how governments influence Internet architecture through state ownership of Internet providers**. Governments often utilize state-owned Internet providers as a means to directly invest in infrastructure and improve the quality of Internet access within their countries. However, before our work, there was a significant gap in understanding the global role and prevalence of these state-owned providers.

To address this gap, our initial study compiled a list of state-owned Internet providers where federal-level governments (or their equivalents) control these companies. Our methodology employed a multi-stage approach that integrated network data — determining where each Internet provider operates — with economic reports to ascertain state ownership. By carefully examining state bulletins, legislation, sovereign fund compositions, and companies' annual reports (for publicly traded entities), **our paper published at ACM IMC [5]** was the first study to demonstrate the widespread state ownership of Internet providers, identifying approximately 1,000 state-owned providers operating in 123 countries across all regions.

Interestingly, this trend of employing **state-owned Internet providers in the telecommunications market is independent of the political regime's nature; it is prevalent in both democracies and autocracies**. While most state-owned Internet providers aim to

reduce the digital divide by connecting citizens to the Internet and providing high-quality service, in authoritarian regimes, when these providers dominate the market, governments gain significant control over network traffic, including the ability to manipulate or censor content.

Building on this work, I participated in follow-up studies — including interdisciplinary research centered on political science [6] — to compare the structure of state-owned transit providers in democracies and autocracies. By developing a metric called *Country Transit Influence* (CTI) to quantify the dependency of transit routes on each provider, and combining this with our compilation of state-owned Internet providers, **our findings revealed that network designs in autocratic states tend to concentrate around state-owned transit providers**. This concentration enables easier access to traffic manipulation and disruptions [4, 6], highlighting the potential for governments in autocratic regimes to exert control over the Internet infrastructure to suppress dissent and control information flow.

Censorship and Network Shutdowns

Given the privileged position that state-owned Internet operators afford to authoritarian governments for censoring the Internet, my research aims to determine whether governments exert censorship through their networks.

Building on our previous findings, we conducted a follow-up study in collaboration with political scientists and legal experts to investigate government strategies for network censorship, focusing on extreme measures such as government-ordered network shutdowns. **By combining network data with ground-level information** collected by local activists associated with the organization Access Now, **our paper published at ACM SIGCOMM [7]** identified at least 55 national-scale shutdowns between 2018 and 2022, including events like Sudan's 2022 shutdown during protests. Our data demonstrates a correlation between the presence of state-owned Internet providers and the implementation of censorship through shutdowns, highlighting the significant impact of government control on Internet accessibility.

Digital Government Services and Infrastructure

Government engagement in the Internet includes numerous initiatives to develop e-government services, bringing functionalities online to meet citizens' growing demand for simplicity and efficiency. As the footprint of government digital services expands, it raises critical debates, one of which closely aligns with my research agenda: the dilemma of where to host governmental content. While third-party solutions for content delivery, such as Content Delivery Networks (CDNs) and cloud services, have been rapidly adopted over the past decade, governments may hesitate to delegate this role to external providers due to cybersecurity concerns, even when serving institutional data that contains no private sensitive information.

Our research examines how governments manage content delivery on the public Internet, specifically whether they utilize third-party providers such as CDNs and cloud services or rely on their own infrastructure. In a study of 1 million websites across 62 federal governments, **published at ACM IMC [8]**, we found a significant reliance on third-party hosting, with 62% of URLs and 53% of data bytes delivered through external providers. However, some governments predominantly deliver their data bytes through their own infrastructure, contrasting with popular commercial sites. We are extending this research to sub-national levels to assess whether lower-level governments face greater technical challenges in content delivery. This ongoing work aims to provide a comprehensive understanding of the factors influencing governments' decisions on content hosting, ultimately informing strategies to enhance the efficiency and security of e-government services.

Critical Internet Infrastructures and Global Connectivity

Despite its vastness and decentralized nature, **the Internet heavily relies on a few critical infrastructures that hold the system together**. The importance of these infrastructures arises from various factors. For instance, scarcity plays a role in Submarine Cable Networks (SCNs), which are limited due to their high cost. Natural aggregation points such as Internet Exchange Points (IXPs) enable local traffic exchanges, while DNS nameservers manage mappings between users and the servers hosting content.

My research focuses on these critical components — specifically SCNs and IXPs — to determine their role in content distribution and assess their resilience and impact on performance and society. This endeavor combines collecting and analyzing extensive measurements with the development of new approaches and tools to thoroughly analyze and enhance these infrastructures.

Submarine Cable Networks

Submarine Cable Networks (SCNs) are the backbone of global data exchange, supporting nearly 100% of intercontinental traffic. Despite the growing popularity of Low Earth Orbit (LEO) satellite constellations for providing Internet access, such as Starlink, SCNs still remain the primary means of intercontinental communication. SCNs enable critical activities such as distributed system

synchronization across continents, real-time global interactions (e.g., online classes, video calls, gaming), and the operation of global financial services. Despite their crucial role, the reliance of different countries and services on SCNs is often underestimated until a high-profile incident occurs. For example, when submarine cables are damaged due to natural disasters (e.g., mudslide affecting West Africa [9]) or human activities (e.g., the Houthi militia sinking ships in the Red Sea [10]), connectivity in affected regions can be severely disrupted.

To understand the role of submarine cables, our work identifies their spans and the locations they connect. By analyzing seven years of Internet topology data containing more than 250 million traceroute instances, our research challenges the traditional understanding of network topology related to submarine termination points. Our findings, published at ACM SIGMETRICS [2], reveal that **routers acting as ingress and egress points to SCNs are completely decoupled from the physical landing points on the seashore where SCNs emerge from the sea**. This decoupling results from recent shifts in network architecture that have integrated segments of various submarine cables into cohesive Long-Haul Links (LHLs), enabling direct connections across vast distances.

For example, we demonstrated that routers in Seattle, USA, and Singapore—over 8,000 miles apart and with no direct submarine cable connecting each other — can be directly linked in a single hop due to advances in optical layer routing. Terrestrial infrastructures can also leverage these optical capabilities, allowing **SCN termination points to be located far inland, such as in Chicago**, thousands of kilometers away from the coast. This development has led to "routerless" connections and increased network opacity, complicating assessments of upper-layer services' reliance on submarine infrastructure. Such opacity challenges failure detection, resilience assessments, and national cybersecurity efforts to understand data flow paths and whether they cross or are exposed to adversaries.

Internet Exchange Points and Regional Infrastructures: A Focus on Latin America

IXPs are another critical infrastructure piece, as they gather hundreds to thousands of networks in a single location to generate efficient local traffic exchange. Latin America, spanning 20 million km^2 with a population of 650 million, presents a unique case for studying the relationship between critical Internet infrastructure and socio-economic factors. The region exhibits remarkable attributes, with São Paulo, Brazil, hosting the world's largest IXP by traffic volume. Despite this, the region's Internet infrastructure has been historically overlooked.

Our research was the first to focus on Latin American IXPs [11], contrasting with the extensive studies conducted in Europe. We revealed that most **IXPs in the region were created as a result of state-sponsored initiatives, highlighting the prominent role of governments in promoting critical infrastructures**. Through Looking Glasses of Packet Clearing House and NIC.br distributed across Latin American IXPs, our results showed that countries like Brazil, Argentina, and Chile have successfully developed extensive IXP networks, while others with monopolistic markets dominated by state-owned incumbents have failed to establish effective IXPs. This illustrates how government involvement and market structure determine the success of IXPs.

As shown, governments in Latin America have a significant impact on the development of critical Internet infrastructures, inviting further investigation into their role in deploying and managing other essential infrastructures.

Venezuela serves as a compelling example of how severe socioeconomic conditions and flawed government policies can devastate Internet infrastructure. Over the past decade, the country has experienced a 70% contraction in GDP per capita and inflation rates soaring at 32,000%, leading to widespread public health crises and mass emigration. In stark contrast to the broader region, **our work published at ACM SIGCOMM [12]** revealed that Venezuela has seen no investments in critical infrastructures.

Our work characterized the extent of Venezuela's Internet stagnation by combining a wide array of Internet measurement datasets with social and macroeconomic statistics over a longitudinal period of at least ten years. Among multiple indicators of eroding critical infrastructure, the most prominent findings include the construction of a single submarine cable —connecting only to Cuba, the scarcity of peering facilities, and the **median bandwidth of less than 3 Mbps, which is far below the regional median of 20 Mbps**. These issues are derived from misguided policies, such as the government's seizure of the largest telecom provider, CANTV, and the decision to invest in a cable to Cuba, which collectively have deterred investment and undermined Internet development.

Securing Critical Infrastructures

An essential aspect of critical Internet infrastructures is ensuring their readiness providing robustness against attack vectors. Operating a system like the Internet — designed and developed before widespread commercial use —involves managing legacy components not originally designed with modern security considerations. The Internet's continuously evolving nature introduces new security challenges that must be proactively addressed.

In my recent research, I have focused on identifying potential vulnerabilities in public systems. Our work in collaboration with UC San Diego **published at ACM IMC [1]** introduces a method called *Darksim*, designed to detect patterns of network traffic associated with distributed scanning activities, which could indicate the prelude of cyber attack. At its core, *Darksim* utilizes Dynamic Time Warping (DTW) to identify scanning profiles occurring simultaneously across the network or at different times. By detecting these profiles, we can uncover coordinated scanning sequences aimed at identifying hosts vulnerable to exploitation.

Darksim was evaluated using the UC San Diego Network Telescope, **a dataset of unsolicited network traffic collected over more than 25 years, with $\approx$ 4TB of compressed data generated daily** in recent months. By employing **high-performance computing resources** to analyze this massive dataset, ***Darksim* successfully identified scanning activities that coincided with the public disclosure of Common Vulnerabilities and Exposures (CVEs)**, underscoring attackers' attempts to exploit these known vulnerabilities in systems that were not yet patched.

Evolution of Content Delivery Networks and Network Performance

Content delivery is crucial in today's digital landscape, and Content Delivery Networks (CDNs) play a significant role in optimizing global content delivery. CDNs, pioneered by companies like Akamai and Limelight Networks, are strategically deployed to reduce latency and improve data transfer by placing content closer to users. Major tech companies such as Google, Meta, Netflix, Microsoft, and Twitch have developed their own extensive global CDNs. With the top seven content providers estimated to manage 70% of global Internet traffic, understanding the evolving architecture of CDNs is essential. My research focuses on evaluating the efficiency and adoption of CDNs and their broader impact on the network ecosystem.

Predicting Emerging CDNs and Traffic Patterns

A key strategy for **maintaining pace with a rapidly evolving network is identifying emerging sources of content delivery** that promise to become major traffic generators in the near future. By pinpointing fast-growing CDNs, Internet Service Providers (ISPs) and Internet Exchange Points (IXPs) can establish strategic peering agreements with these networks, resulting in lower operational costs and enhanced user experiences.

To address the challenge operators face in identifying these emerging CDNs, **I employ graph-theoretical methods applied to Autonomous System (AS)-level connectivity graphs to forecast the growth of CDNs based on network connectivity density [13, 14]**. Based on the experience of the deployment of today's large CDNs, this methodology analyzes the density and structure of AS connectivity to predict which CDNs are likely to experience rapid expansion. Our approach has successfully predicted the rise and fall of private CDNs, exemplified by Spotify's transition from its own CDN infrastructure to reliance on third-party providers. Our research helps ISPs and IXPs make informed decisions about network planning and peering strategies, optimizing network performance and contributing to a more efficient Internet ecosystem.

Challenges in Content Delivery: Performance

Efficient content delivery is a shared goal among Internet stakeholders, yet disputes often arise over capacity upgrades between Internet Service Providers (ISPs) and Content Providers (CPs). When these disputes remain unresolved, users suffer from insufficient capacity to meet their needs, resulting in recurrent congestion and a degraded perceived user experience.

Although the Internet community has been dealing with congestion since Van Jacobson's seminal work in 1988 [15], **managing network congestion and identifying bandwidth bottlenecks remain central challenges**. To address these issues, we developed *Jitterbug* [16], a novel, non-invasive congestion detection method designed to minimize the impact on hosts. Unlike traditional tools that focus solely on latency, ***Jitterbug* focuses on stochastic changes in jitter signals**. This enables the identification of both recurring and sporadic congestion events, making it the first of its kind to identify the latter case. This approach provides a more detailed assessment of content delivery quality across networks, improving our ability to effectively manage and alleviate congestion. By offering a deeper understanding of traffic patterns and bottlenecks, Jitterbug facilitates better decision-making for capacity planning, infrastructure upgrades, and informing users and regulators.

Properly Capture the Size of CDNs: AS-to-Organization Mapping Techniques

Although a single Autonomous System Number (ASN) is sufficient for an independent organization to manage its own routing, **organizations often use multiple ASNs due to mergers, acquisitions, and network segmentation**. Managing multiple ASNs is a challenge that extends beyond the Content Delivery Network (CDN) domain but is particularly common among CDNs. This is evident in multiple units within large organizations and legacy infrastructure following mergers and acquisitions, such as Google and Google Cloud, YouTube and Google, GitHub and LinkedIn under Microsoft, and Twitch and Amazon, among others.

To address the discrepancies between individual ASNs and their associated organizations, we developed the *as2org+* methodology [3]. This approach integrates organizational data from PeeringDB and employs Information Extraction techniques to organize unstructured data effectively. ***as2org+* has successfully grouped previously disjoint entities, resulting in a more accurate representation of network structures** and enhancing our understanding of organizations' footprints and routing policies. Currently, we are advancing this methodology by incorporating machine learning techniques to further refine and automate the mapping process, thereby improving the accuracy and scalability of our analysis.

Future Research Agenda

Looking ahead, there are several exciting challenges at the intersection of technology and society that I would like to tackle. Building on my previous efforts, I plan to complement my previous studies by focusing on three interconnected themes: **(1) Exploring the Socio-Political Dynamics of Internet Control**, **(2) Critical Infrastructures: Resilience, Failure Detection, and Recovery**, and **(3) Securing Critical Infrastructures**.

1. Exploring the Socio-Political Dynamics of Internet Control

The socio-political landscape plays a crucial role in shaping Internet architecture and accessibility. My research will continue to investigate the role of governments in influencing the Internet and their participation in the network as both access and content providers. A better understanding of governments' roles and footprints is essential to addressing critical issues across the spectrum, from how to empower users to defend themselves against government censorship to how to help governments efficiently and securely deliver content to their citizens.

This line of work will explore how government infrastructures are exposed to cyberattacks and the strategies implemented to mitigate such threats. By working with scientists from various fields, I hope to create a thorough framework for analyzing how government policies impact their outcomes.

2. Critical Infrastructures: Resilience, Failure Detection, and Recovery

A key to maintaining a robust and adaptive Internet is the ability to identify and mitigate vulnerabilities within its critical infrastructures. My future work will further map and analyze Submarine Cable Networks (SCNs), Internet Exchange Points (IXPs), and other essential infrastructures to uncover vulnerabilities, optimize performance, and ensure resilience against technical failures and geopolitical disruptions. Building on our findings of increased network opacity due to the rise of optical routing, I will focus on characterizing failures that often result in "soft failures"—instances of congestion rather than total disruption—due to the availability of backup transparent resources.

To detect and localize these soft failures effectively, my research will involve compiling large-scale datasets and developing advanced congestion detection tools, such as *Jitterbug*. These non-invasive monitoring tools will provide real-time reading of network performance and enable prompt responses to congestion events. By enhancing our ability to manage and mitigate congestion, this work aims to improve the overall reliability and user experience of content delivery networks.

3. Securing Critical Infrastructures

Securing the Internet's critical infrastructures is key to guaranteeing connectivity and essential services. My research will advance infrastructure geolocation techniques and enhance security threat detection to better assess and protect these vital components. Understanding the precise locations of critical infrastructure, such as core routers and submarine cables, is crucial for identifying vulnerabilities. For instance, a bomb explosion in front of an AT&T facility in Nashville caused long-lasting outages in Alabama and Kentucky [17], demonstrating the significant impact of infrastructure localization on connectivity and public safety.

To address these challenges, I plan to integrate multiple physical and network datasets—such as rights-of-way, submarine cable deployments, and peering facilities—to accurately infer the physical infrastructures used to reach specific destinations. I also contribute to creating standard datasets and metrics for evaluating frameworks for detecting attacker signatures in large volumes of unsolicited traffic, simplifying the comparison of different methodologies and helping develop more effective defense mechanisms. This initiative will enhance our ability to detect and respond to security threats proactively, ensuring the resilience and integrity of critical Internet infrastructures.

My research agenda reflects a holistic approach to understanding and enhancing the Internet's infrastructure and its interaction with societal factors. By addressing critical infrastructures' resilience and security, and exploring the socio-political dynamics of Internet control, my work aims to contribute to a more robust, secure, and equitable Internet. These efforts will ensure that the Internet remains a reliable and adaptable cornerstone of modern life, capable of meeting the evolving needs of users worldwide.

References

[1] Max Gao†, Eric Li, Shubham Kulkarniand, Ricky P. K. Mok, Esteban Carisimo, and k Claffy. Darksim: A similarity-based time-series analytic framework for darknet traffic. In *ACM IMC*, nov 2024.

[2] Esteban Carisimo, Caleb J. Wang, Mia Weaverand, Fabián E. Bustamante, and Paul Barford. A hop away from everywhere: A view of the intercontinental long-haul infrastructure. In *Proc. ACM Meas. Anal. Comput. Syst.*, 12 2024.

[3] Augusto Arturi†, Esteban Carisimo, and Fabián E. Bustamante. as2org+: Enriching as-to-organization mappings with peeringdb. In *Passive and Active Measurement*, pages 400--428, Cham, 03 2023. Springer Nature Switzerland.

[4] Alexander Gamero-Garrido, Esteban Carisimo, Shuai Hao, Bradley Huffaker, Alex C. Snoeren, and Alberto Dainotti. Quantifying nations' exposure to traffic observation and selective tampering. In *Passive and Active Measurement*, pages 645--674, Cham, 03 2022. Springer International Publishing.

[5] Esteban Carisimo, Alexander Gamero-Garrido, Alex C. Snoeren, and Alberto Dainotti. Identifying ases of state-owned internet operators. In *ACM Internet Measurement Conference (IMC)*, page 687–702, New York, NY, USA, 11 2021. Association for Computing Machinery.

[6] Eda Keremoğlu, Nils B Weidmann, Alexander Gamero-Garrido, Esteban Carisimo, Alberto Dainotti, and Alex C Snoeren. Network topology facilitates internet traffic control in autocracies. *PNAS Nexus.*, page pgae069, 02 2024.

[7] Zachary Bischof, Kennedy Pitcher, Esteban Carisimo, Amanda Meng, Rafaek Nunes, Ramakrishna Padmanabhan, Margaret E. Roberts, Alex C. Snoeren, and Alberto Dainotti. Destination unreachable: Characterizing internet outages and shutdowns. In *ACM SIGCOMM*, 09 2023.

[8] Rashna Kumar†, Esteban Carisimo, Lukas De Angelis Rivas, Mauricio Buzzone, Fabián E. Bustamante, Ihsan Ayyub Qazi, and Mariano G. Beiró. Of choices and control - a comparative analysis of government hosting. In *ACM IMC*, nov 2024.

[9] Kentik. Dual subsea cable cuts disrupt african internet, 2023. Accessed on September 17, 2024.

[10] Wired. A ghost ship's doomed journey through the gate of tears. *Wired*, 2024. Accessed on September 17, 2024.

[11] Esteban Carisimo, Julián M. Del Fiore, Diego Dujovne, Cristel Pelsser, and J. Ignacio Alvarez-Hamelin. A first look at the latin american ixps. *SIGCOMM Comput. Commun. Rev.*, 50(1):18–24, mar 2020.

[12] Esteban Carisimo, Rashna Kumar, Caleb J. Wang, Santiago Klein, and Fabián E. Bustamante. Ten years of the venezuelan crisis - an internet perspective. In *ACM SIGCOMM*, 08 2024.

[13] Esteban Carisimo, Carlos Selmo, J. Ignacio Alvarez-Hamelin, and Amogh Dhamdhere. Studying the evolution of content providers in the internet core. In *Network Traffic Measurement and Analysis Conference (TMA)*, pages 1--8, 2018.

[14] Esteban Carisimo, Carlos Selmo, J. Ignacio Alvarez-Hamelin, and Amogh Dhamdhere. Studying the evolution of content providers in ipv4 and ipv6 internet cores. *Computer Communications.*, 145:54--65, 2019.

[15] Van Jacobson. Congestion avoidance and control. *SIGCOMM Comput. Commun. Rev.*, 1988.

[16] Esteban Carisimo, Ricky K. P. Mok, David D. Clark, and Kc Claffy. Jitterbug: A new framework for jitter-based congestion inference. In *Passive and Active Measurement*, pages 155--179, Cham, 2022. Springer International Publishing.

[17] Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, Kimberly C Claffy, and Aaron Schulman. Inferring regional access network topologies: Methods and applications. In *Proc. of ACM IMC*, 2021.

† *indicates mentee*