Duke Law Journal Online

VOLUME 73 OCTOBER 2023

INSIDE THE INTERNET

NICK MERRILL† AND TEJAS N. NARECHANIA ††

ABSTRACT

Conventional wisdom—particularly in the legal literatures—suggests that competition reigns the inside of the internet. This common understanding has shaped regulatory approaches to questions of network security and competition policy among service providers. But the original research presented here undermines that long-held assumption. Where the markets for internet traffic exchange (and related services) have long been thought to be characterized by robust competition among various network services providers, our findings suggest that these markets have consolidated. These trends raise a host of concerns for network reliability, online speech, and consumer choice, among other matters. Indeed, some recent high-profile internet outages reflect some of these concerns. And so we consider how the internet's regulatory infrastructure might respond to these new revelations about the internet's interior network infrastructure. Specifically, we call for regulation to enhance visibility of the internet's interior and to assure a regime of fair carriage for all the internet's users.

† Research Fellow, Center for Long-Term Cybersecurity, University of California, Berkeley.

†† Professor of Law, University of California, Berkeley, School of Law. For helpful comments and suggestions, we thank Mat Ford, James Grimmelmann, Chris Hoofnagle, Tian Kisch, Khushali Narechania, Delia Scoville, Scott Shenker, Erik Stallman, Rebecca Wexler, as well as audiences at the Digital Life Initiative at Cornell Tech and the University of California, Berkeley, School of Law. For outstanding research assistance, we thank Jennifer Sun. We also thank Tom Fogarty and the editors of the *Duke Law Journal Online* for their careful edits and thoughtful suggestions.

INTRODUCTION

On June 8, 2021, the internet seemed to come to a standstill. Suddenly, amazon.com wouldn't respond. CNN, Pinterest, Reddit, Spotify, and Twitch were all down. HBO was inaccessible. Even the official website of the United Kingdom's government—gov.uk—was offline. Sources online speculated that a coordinated cyberattack had caused this sudden series of outages.[1]

In truth, these websites failed simultaneously because a simple error at Fastly, a content delivery network (or CDN for short), unsettled the internet's software supply chain.[2] But the internet is meant to be resilient—to avoid these sorts of cascading, catastrophic failures. In its original architecture, the internet was designed to route requests around outages in any one network services provider.[3]

How, then, could a relatively simple error at one CDN metastasize into such a significant issue? Addressing this question requires a look into the internet's evolving topology, alongside the governance and market structures that attend to the internet's interior. Many consumers understand the basics of the internet's edges: we know, for example, that we need a computer (an Apple MacBook, perhaps) with an internet connection (say, Comcast's Xfinity) to access a website (such as Google). But most know far less about how a user's request for Google's services traverses the *middle* of the internet, from Comcast's network to Google's servers and back.

In this Article, we provide an updated picture of the internet's interior workings, drawing in part on the original internet measurement research developed by one of us (Merrill). Conventional wisdom—particularly in law and policy contexts—suggests that competition reigns the markets at the middle of the internet.[4] But the

---

1. Ryan Browne, *What is Fastly and Why Did It Just Take a Bunch of Major Websites Offline?*, CNBC (June 8, 2021, 10:44 AM), https://www.cnbc.com/2021/06/08/fastly-outage-internet-what-happened.html [https://perma.ccR5YE-SGDL].

2. Clare Duffy, *Two Obscure Service Providers Briefly Broke the Internet. It Could Happen Again*, CNN (June 17, 2021, 2:26 PM), https://www.cnn.com/2021/06/09/tech/fastly-cdn-internet-risk/index.html [https://perma.cc/26FX-QF42]; *see also* Nick Rockwell, *Summary of June 8 Outage*, FASTLY (June 8, 2021), https://www.fastly.com/blog/summary-of-june-8-outage [https://perma.cc/MQ8W-AHLY].

3. *See infra* Part I.

4. *See, e.g.*, JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, DIGITAL CROSSROADS 183–84 (2d ed. 2013) ("By most accounts, transit services are highly competitive today. One reason is that . . . conventional backbone providers now compete not only with one another, but

findings outlined here suggest that such competition is now far less robust than typically assumed.[5] Moreover, other trends in the internet's interior point not only towards *consolidation*, but also towards the increasing *opacity* of the internet's constituent networks. And as these privately owned networks increasingly mediate the internet's core, a lack of public visibility into the internet's structure inhibits risk assessment and disaster planning.[6] Viewed together, these new facts pose significant, but overlooked, internet access, security, and reliability challenges, evinced, for example, by the June 8 outage.

In view of these findings, we contend that regulators must revisit the governance regimes for what has sometimes been known, perhaps too simplistically, as "the market for internet traffic exchange."[7] Specifically, we advocate for new transparency and regulatory regimes to help address the concerns arising out of the consolidation and opacity in these markets. Centralized infrastructures often require centralized risk management, particularly in network contexts.[8] But the opacity and secrecy that shrouds the internet's increasingly opaque interior undermines attempts to plan for a cyberattack, a natural disaster, or even a simple human error. And so we advocate for expanded disclosure mandates as one part of a more comprehensive federal risk management regime.[9] Moreover, this consolidation at the internet's interior renews debates (familiar to the network neutrality context) regarding consumer choice and speech. And so we make further, if tentative, recommendations for regulating these intermediary markets.[10]

This short Article proceeds in four Parts. In the first, we describe a conventional, if dated, understanding of the internet's core. This conventional wisdom regards the markets for internet transit (and related services) as characterized by robust competition among

---

also with alternative mechanisms . . . . Those alternative mechanisms include . . . CDNs."); *see also infra* notes 33–43 and accompanying text.

5. *See infra* Part II.

6. *See infra* Part III.

7. *See, e.g.*, *In re* Restoring Internet Freedom, 33 FCC Rcd. 311, 409 ¶ 164 (2018) (Declaratory Ruling, Report and Order) [hereinafter RIFO].

8. *See* KEVIN STINE, STEPHEN QUINN, GREGORY WITTE & R.K. GARDNER, INTEGRATING CYBERSECURITY AND ENTERPRISE RISK MANAGEMENT (ERM) 2–11 (2020), https://doi.org/10.6028/NIST.IR.8286 [https://perma.cc/U25D-K4HF] (emphasizing the value of centralizing risk management in the cybersecurity context).

9. *See infra* Section IV.A.

10. *See infra* Section IV.B.

network services companies offering, essentially, public carriage of internet content.[11] Moreover, we describe how this view has shaped the regulatory environment thus far. In the second Part, we challenge this conventional wisdom, drawing on the original internet measurement research developed by one of us (Merrill). In particular, this research suggests that the market for network services inside the internet has shifted to a more consolidated set of providers. These providers, moreover, have turned decisively towards relying on privately owned infrastructure that is not available to other downstream users. In other words, these providers have vertically integrated these network services. In the third Part, we describe the security and competition concerns (among others) that attend to this new network and market structure. And so, finally, in the last part, we consider how our regulatory infrastructure ought to respond to these changes in the internet's infrastructure.

I. COMPETITION INSIDE THE INTERNET?

We begin with some brief historical context regarding the modern internet's design.[12] In early conceptions, the internet was envisioned as a point-to-point network. Content, hosted by users in their homes and offices, was globally accessible via decentralized networks—essentially, local internet service providers (or ISPs) such as America Online (AOL) or Comcast. Those ISPs were themselves interconnected via intermediary networks, such as WorldCom.[13] Say, for example, that one user (subscribing to one ISP) requests a website hosted somewhere else. Such a content request would exit that user's ISP, traverse one or several intermediary networks, before reaching the website host's ISP and, ultimately, the site itself. As described in

11. *See, e.g.*, *infra* notes 14–19, 33–43 and accompanying text.

12. *See generally*, Paul Dourish, *Protocols, Packets, and Proximity: The Materiality of Internet Outing*, *in* SIGNAL TRAFFIC: CRITICAL STUDIES OF MEDIA INFRASTRUCTURES (Lisa Parks & Nicole Starosielski eds., 2015) (exploring the physical infrastructure of internet routing); Paul Dourish, *Not the Internet, but This Internet: How Othernets Illuminate Our Feudal Internet*, *in* AARHUS SERIES ON HUMAN CENTERED COMPUTING (2015), https://tidsskrift.dk/ashcc/article/view/21200/18686 [https://perma.cc/MAA2-4YPW] (reviewing possible network alternatives in order "to place 'the Internet' in some context"); David Clark, DESIGNING AN INTERNET (2018) (overviewing the history of the internet from its beginning in the 1970s to modern day).

13. *See* Tung-Hui Hu, *Truckstops on the Information Superhighway: Ant Farm, SRI, and the Cloud*, J. NEW MEDIA CAUCUS (Apr. 2014), http://median.newmediacaucus.org/art-infrastructures-hardware/truckstops-on-the-information-superhighway-ant-farm-sri-and-the-cloud [https://perma.cc/Q96J-T7HM] (comparing the early internet to a decentralized interstate highway).

Figure 1, these requests ascend a stack of tiered providers, from Tier-3, to Tier-2, to Tier-1 providers, and then descend back down again, until finally reaching the destination.[14] Hence, internet access and a spare computer were all that it took to visit—and, critically, create—a website.[15]

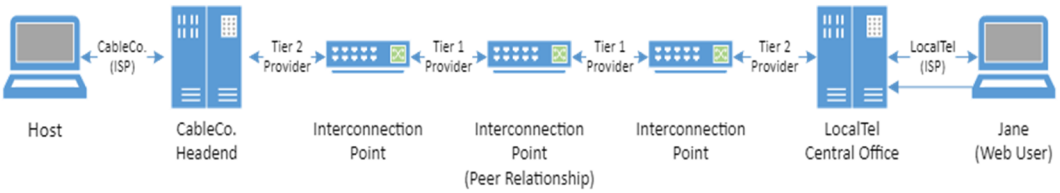

Figure 1. A graphical representation of the point-to-point vision of the internet. Adapted from Tejas N. Narechania & Erik Stallman, *Internet Federalism*, 34 HARV. J.L. & TECH. 547 (2021).

Tier-1 providers sit atop this hierarchy of providers. Collectively, Tier-1 providers can reach any location on the internet without having to purchase carriage from another lower tier provider. Hence, Tier-1 providers form the core of this model of the internet, giving rise to a competitive market for transit.[16] Indeed, some Tier-1 networks sell bandwidth—i.e., transit capacity—on their networks to smaller Tier-2 and Tier-3 providers, including some ISPs and cloud providers (some of whom have also built their own networks).[17] One might worry about the power of Tier-1 providers to collusively fix prices in downstream

14. NUECHTERLEIN & WEISER, *supra* note 4, at 19; *see also In re* Protecting and Promoting the Open Internet, Report and Order on Remand, 30 FCC Rcd. 5601, 5687–92 ¶¶ 196–201 (2015) [hereinafter 2015 OIO] (Declaratory Ruling and Order) ("Backbone service providers interconnected upstream until traffic reached Tier-1 backbone service providers, which peered with each other and thereby provided their customer networks with access to the full Internet.").

15. Internet service from your ISP entitles you to an IP address, a publicly accessible numerical address that uniquely identifies your computer on the global Internet. By associating that IP address with a human-readable name (like a ".com") through the Domain Name System (DNS)–think a phonebook– others can look up your website without having to memorize your IP address (or update their records when your IP address changes).

16. *See* Dennis Weller & Bill Woodcock, *Internet Traffic Exchange: Market Developments and Policy Challenges* 3 (OECD Digit. Econ. Working Paper, No. 207, 2013) ("Operating in a highly competitive environment, largely without regulation or central organisation, the Internet model of traffic exchange has produced low prices, promoted efficiency and innovation, and attracted the investment necessary to keep pace with demand.").

17. *See* 2015 OIO, *supra* note 14, at 5687–92 ¶¶ 196–201 (2015).

bandwidth markets, including those encompassing sales to Tier-2 and Tier-3 providers. But several surveys of the relationships among Tier-1 providers suggest that they operate in an open, competitive market for bandwidth, one which keeps providers honest and prices low for commodity bandwidth.[18]

The competitive market for bandwidth at the internet's core has led policymakers to believe that the internet's core is "efficient[]."[19] These networks are agnostic as to the content carried, and they sell capacity at a market-clearing price. In short, this competition, it is said, resolves concerns about price and potential discrimination.[20]

This belief in an efficient core, however, hinges on the assumption that Tier-1 providers and the market in which they operate matter as much today as they did in the 1990s, when the internet was newer. But the point-to-point internet eventually proved insufficient for today's more bandwidth-intensive consumer internet in at least two ways.

First, under the point-to-point model, requests for geographically distant content suffered from high latency. Such latency originally meant that websites would load comparatively slowly by modern standards, causing browsers to "timeout" before the content could load,[21] or causing users to lose interest and abandon the request.[22] But as so-called "Web 2.0" applications sought to transform the internet from a set of static documents into a more dynamic experience—email

---

18. Weller & Woodcock, supra note 16, at 3, 61.

19. *See* RIFO, *supra* note 7, at 409, 413–414 ¶ ¶ 164, 170 (noting that the "the market for Internet traffic exchange between ISPs and edge providers or their intermediaries historically has functioned without significant Commission oversight" (citation and internal quotation marks omitted) and emphasizing the role of "present competitive pressures in the market for Internet traffic exchange" in disciplining provider conduct); AT&T Servs. Inc., Comments in the Matter of Restoring Internet Freedom WC Docket No. 17-708 47 (July 17, 2017) ("All of these commercial relationships have always been unregulated, and the interconnection marketplace has always functioned efficiently . . . .").

20. *See supra* notes 16–18; *infra* notes 37–44; *see also* BILL WOODCOCK, WHITE PAPER ON TRANSACTIONS AND VALUATION ASSOCIATED WITH INTER-CARRIER ROUTING OF INTERNET PROTOCOL TRAFFIC (2000), pch.net/resources/Papers/routing-economics/index.html [https://perma.cc/ZX3X-HQFU].

21. Roy T. Fielding, Mark Nottingham & Julian Reschke, *Internet Engineering Task Force (IETF) Request for Comments: 9110*, IETF DATA TRACKER (June 2022), https://datatracker.ietf.org/doc/html/rfc9110 [https://perma.cc/VKN8-WFHN] (describing the timeout error code in the HTTP standard in §15.5.9).

22. *See, e.g.*, Steve Lohr, *For Impatient Web Users, an Eye Blink Is Too Long To Wait*, N.Y. TIMES (Feb. 29, 2012), https://www.nytimes.com/2012/03/01/technology/impatient-web-users-flee-slow-loading-sites.html [https://perma.cc/AWC9-W6G2].

inboxes that refresh automatically, or news feeds that are continuously updated—latency hindered the development and utility of these new applications.[23] For example, geographic diversity among ISPs' users, along with the inherent, physical limits of internet facilities, made it difficult or impractical to deliver streaming video or audio reliably for all users.

Second, though this point-to-point model prized dispersed internet content held by independent hosts, an increase in cyberattacks demonstrated how security can be a collective good,[24] highlighting the value in shared defense.[25] For example, the 1990s and early 2000s saw a rise in distributed denial of service attacks—or DDoS attacks—which bombard a service with a large amount of traffic, thereby making it unavailable to legitimate users.[26]

Such attacks, which do not require sophisticated tools to deploy, became widespread and disruptive. In 2000, for example, sixteen-year-old Michael Calce—known online as *Mafiaboy*—brought down CNN,

23. *See*, *e.g.*, Andrea Cardaci, Luca Caviglione, Alberto Gotta & Nicola Tonellotto, *Performance Evaluation of SPDY Over High Latency Satellite Channels* 123–24, *in* PERSONAL SATELLITE SERVS., (2013) (describing Web 2.0 systems that "constantly transmit[] data between the server and the client" to create "real-time collaboration frameworks"). Moreover, some studies have found that latency is a determinant of consumer trust in, say, a retail website or a banking application, suggesting that the public's willingness to adopt these advances has depended on the nature and quality of network access throughout the internet's structure. *See* Gerard Ryan & Mireia Valverde, *Waiting for Service on the Internet: Defining the Phenomenon and Identifying*, 15 INTERNET RES. 220, 221 (2005) (citing Sung-Joon Yoon, *The Antecedents and Consequences of Trust in Online-Purchase Decisions*, 16 J. INTERACTIVE MKTG. 47, 47–63 (2002)); *cf.* Infrastructure Investment and Jobs Act, Pub. L. No. 117–58, 135 Stat. 1199 (defining minimum basic requirement for broadband internet access service, or broadband carriage, as including "a latency that is sufficiently low to allow reasonably foreseeable, real-time, interactive applications").

24. Deirdre K. Mulligan & Fred B. Schneider, *Doctrine for Cybersecurity*, 140 J. AM. ACAD. ARTS & SCI. 70, 80 (2011) (stating "[o]ur doctrine of *public cybersecurity* . . . is rooted in the thesis that cybersecurity is a public good" (emphasis in original)).

25. *Cf.* Mazaher Kianpour, Stewart James Kowalski & Harald Øverby, *Advancing the Concept of Cybersecurity as a Public Good*, SIMULATION MODELLING PRACTICE AND THEORY, Apr. 2022, at 1, https://doi.org/10.1016/j.simpat.2022.102493 [https://perma.cc/HQ84-S88B] ("Evolving malicious cyber activities and increasing cyber risks to individuals, organizations and governments has made cybersecurity a significant challenge and core part of the societal, political and economic decisions.").

26. *See, e.g.*, Ketki Arora, Krishan Kumar & Monika Sachdeva, *Impact Analysis of Recent DDoS Attacks*, 3 INT'L J. COMP. SCI. & ENG'G 877, 882 (2011) (noting a one-thousand-fold increase in DDoS attacks from 2003 to 2011).

Yahoo, Amazon, Dell, eBay, and FIFA with a DDoS attack.[27] Such attacks can be difficult for individual websites to defend against. DDoS attacks use numerous endpoints to "hijack" individual computers that typically source legitimate traffic. The attacker might use malware, already loaded onto these computers, to bombard a target with repeated internet requests.[28] That sudden influx of traffic can overwhelm that target, depleting its bandwidth and computational capacity, making the site inaccessible to legitimate viewers. Moreover, because these malicious requests are camouflaged as legitimate ones, content hosts have difficulty separating bad requests from good ones, making it difficult to end the attack without taking the site offline altogether.[29]

Victims of DDoS attacks, however, can better address these incidents by sharing internet-traffic-related intelligence. Indeed, the most effective defenses against DDoS attacks rest upon large-scale observations of network traffic: a centralized observer, or federated network of observers, can share intelligence about troublesome sources of internet traffic to help separate legitimate requests from malicious ones.[30]

Content delivery networks (or CDNs, for short) are an important innovation due in significant part to how they respond to these two concerns. CDNs duplicate—that is, cache[31]—these internet companies' content in localized servers across the internet to minimize latency, while also offering a collective defense against cyberattacks and the expert management of this distributed internet infrastructure. Where the internet once functioned primarily as a widely distributed network

27. Rick Davis, *The History and Future of DDoS Attacks*, CYBERSECURITY MAG. (Jan. 15, 2021), https://cybersecurity-magazine.com/the-history-and-future-of-ddos-attacks [https://perma.cc/8YZE-B5HM].

28. Commonly, attackers construct "botnets" from compromised machines. However, DDoS attacks have also been launched through collective, volunteer action, as was the case in Anonymous's attacks against the Church of Scientology. *See generally* GABRIELLA COLEMAN, HACKER, HOAXER, WHISTLEBLOWER, SPY: THE MANY FACES OF ANONYMOUS (2014) (describing the anatomy of a DDoS attack).

29. *Understanding Denial-of-Service Attacks*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY, https://www.cisa.gov/uscert/ncas/tips/ST04-015 [https://perma.cc/9TMS-69U5] (last updated Oct. 28, 2022).

30. *See What Is DDoS Mitigation*, CLOUDFLARE, https://www.cloudflare.com/en-ca/learning/ddos/ddos-mitigation [https://perma.cc/HF8W-TFXK] ("Cloudflare's network runs Internet requests for millions of websites, creating an advantage in analyzing data from attack traffic around the globe.")

31. 2015 OIO, *supra* note 14, at 5687–88 ¶¶ 197–98.

of computers, with individual users connected to each other as described above, CDNs now offered an alternate distribution model.

To illustrate the differences between these models, consider a user checking the latest basketball statistics from the NBA. If the NBA does not use a CDN to host its statistics, that user's request might traverse a series of networks to arrive at the NBA's server. This process of "hopping" from one network to another increases both latency—as each hop adds time to the round trip between the NBA's servers and the user's computer—and the risks of packet loss—as each hop is a new opportunity for failure. But if the NBA instead employs a CDN, then the NBA need only deliver the statistics to that CDN. The CDN would then distribute the NBA's content across its geographically dispersed network of servers—moving, say, Golden State Warriors-related content closer to its facilities in California and Boston Celtics-related content to Massachusetts—thereby reducing latency. This practice, known as *caching*, helps to enable more dynamic content, such as video highlights. In either case, users access nba.com. But how nba.com is hosted varies: in the earlier example, users access nba.com at one location—say, NBA HQ in NYC; but in our new paradigm, users access the site from a local cache maintained by a CDN, which is responsible for distributing and replicating NBA content across its own network of servers.[32]

Moreover, a CDN that serves the NBA—and the NFL and the NHL and MLB—has a wider view of internet traffic than any one of those organizations alone. So, based on an understanding of the patterns across all of them, it is better able to detect and mitigate cyberattacks directed at any one of these leagues.

On one view, then, by entering the market for internet traffic exchange, CDNs helped to make this competitive market even more so. The Federal Communications Commission, for example, described CDNs as one class of participant in the general market for internet traffic exchange. In 2018, the agency described the market as "emerging and competitive," and CDNs as one "innovative" "alternative" to other modes of traffic exchange,[33] including the

32. As you might imagine, configuring this geographic dispersal in real-time is automated by algorithms. The challenges for policy of an internet whose core infrastructure is increasingly characterized by such algorithms is discussed at greater length. *See infra* Part IV.

33. *See* RIFO, *supra* note 7, at 412–13 ¶ 169.

traditional modes of transit, via Tier-1 providers, among others, such as direct interconnection.[34]

The Commission's 2018 statement, its most recent, builds on a long line of precedent that understands the market for internet traffic exchange as robustly competitive. Consider, for example, the Commission's 2010 network neutrality rules ("2010 OIO"). It issued those rules in view of consolidation-related concerns in the access network market—i.e., the market for retail broadband subscriptions, such as those users might purchase from Comcast or Verizon.[35] When it came to the market for internet traffic exchange, however, the Commission bluntly noted that it was treating interconnection arrangements as beyond the scope of those rules, implying that the apparently distinct competitive conditions in these respective markets justified the differential treatment.[36]

Likewise, when the Commission issued new network neutrality rules in 2015, it took only cautious steps towards superintending the market for internet traffic exchange. It decided against greater regulation because of the apparent competition among a wide range of services and service providers in the market, including several Tier-1 providers, several CDNs, and a range of other transit service providers.[37] And finally, as noted, the Commission's justified its most recent decision to re-deregulate this market on the view that market discipline, through the continuing "competitive pressures in the market for Internet traffic exchange," are more efficient than regulatory intervention.[38]

---

34. *Id.*

35. Preserving the Open Internet, 25 FCC Rcd. 17905, 17943–44 ¶ 67 n.209 (2010) (Report and Order).

36. *Id.*; *see also* NUECHTERLEIN & WEISER, *supra* note 4, at 290–92 (considering the 2010 Order and emphasizing the availability of "multiple providers" as one critical factor weighing against regulatory intervention). It is true that the 2010 OIO has been criticized on the grounds that network neutrality should rationally extend through, at least, interconnection at the edge. *See id.* at 214–216, 287–290. Our critique is slightly different. While we do not disagree with such critiques insofar as they pertain to interconnection at the edge, interconnection in the interior of the internet, however, presents different concerns. In that latter context, our disagreement with the Commission's approach in the 2010 OIO (and other regulatory proceedings) is founded on our study of the technical and economic structures of the interior of the internet, as our results differ from the FCC's assertions about "the market for Internet traffic exchange." *See supra* note 33 and accompanying text.

37. 2015 OIO, *supra* note 14, at 5687–88 ¶¶ 197–98.

38. *See* RIFO, *supra* note 7, 413–414 at ¶ 170.

The Commission's largely consistent statements over the past decade regarding the state of this market seem to reflect a view shared by a wide range of scholars,[39] policymakers,[40] and market participants.[41] One leading telecommunications policy text, for example, suggests that "transit services are highly competitive," attributing that competition "not only to conventional backbone providers [that] compete . . . with one another" but also "alternative mechanisms" such as CDNs.[42] And AT&T and Comcast, among others, have all reported that the interconnection marketplace is "efficient[]" and "well-functioning."[43] In all, the market for traffic exchange has long been thought to be characterized by a number of different classes of services—transit, CDNs, etc.—as well as a number of providers within each class.

II. CONSOLIDATION INSIDE THE INTERNET

As noted, one view—a dominant view, it seems—is that competition reigns the market for internet traffic exchange. Some recent incidents, however, might give us reason to question that longstanding assumption. In 2021, for example, internet users saw at least two high-profile internet outages—each of which might be traced

39. *See* Kevin Werbach, *Only Connect*, 22 BERKELEY TECH. L.J. 1233, 1253–54 (2007) ("For most of the internet's history, there have been sufficient backbone competitors to limit the market power any one might enjoy." (footnote omitted)); *see also infra* note 42.

40. *See, e.g.*, *Oversight of the Federal Commc'ns Comm'n: Hearing Before the Comm. on Com., Sci., & Transp.*, 114th Cong. 153 (2016) (Response to Written Questions Submitted by Hon. John Thune to Hon. Ajit Pai) ("Indeed, the best evidence in the record suggests the free market for interconnection has been an unmitigated success, with transit rates falling 99 percent over the last decade."); Dissenting Statement of Comm'r Michael O'Rielly, 2015 OIO, *supra* note 14, at 5994 (Dissenting Statement of Comm'r Michael O'Reilly) (remarking that the market for Internet traffic exchange is a "thriving, competitive market"); FED. TRADE COMM'N, BROADBAND CONNECTIVITY COMPETITION POL'Y 26 (June 2007), at 26 ("To date, market forces have encouraged interconnection among backbones and between backbones and last-mile ISPs."); RIFO, *supra* note 7, at 412–13 ¶ 169 ("We believe that market dynamics, not Title II regulation, allowed these diverse [alternative internet traffic exchange] arrangements to thrive.").

41. *See infra* note 43.

42. NUECHTERLEIN & WEISER, *supra* note 4, at 183–84 ("By most accounts, transit services are highly competitive today. One reason is that . . . . conventional backbone providers now compete not only with one another, but also with alternative mechanisms . . . . includ[ing] the CDNs . . . .").

43. *See, e.g.*, AT&T Servs. Inc., *supra* note 19; Comcast Corp., Reply Comments In the Matter of Restoring Internet Freedom 37 (Aug. 2017) (interconnection is a "well-functioning marketplace").

back to an error or glitch at one CDN.[44] Yet, if the internet's interior was robustly competitive, observers might be surprised that an error at any one provider could cause such widespread headaches. Such competition should facilitate the redundancy—the availability of different traffic routes—that is inherent to the internet's design.[45]

In short, these incidents may suggest that today's internet more closely resembles a centralized network with few central, critical points-of-failure, rather than the decentralized map of alternative traffic paths that many imagine. Fortunately, this speculation raises a testable question: how consolidated, really, is the market for internet traffic exchange? Our novel study, described below, helps to answer this question.

A. Methodology

One can begin to address this question with methods and tools used by the community of internet measurement scholars.[46] To compute these results, one of us (Merrill) used data from the world's top websites while working in conjunction with W3Techs, an organization that collects data, via technical means, about the use of various internet technologies.[47]

There are fundamental challenges in surveying the user-facing web.[48] For instance, it is difficult to establish a complete picture of

---

44. *See* Jim Salter, *Today's Massive Internet Outage Comes Courtesy of Akamai Edge DNS*, ARS TECHNICAL (July 22, 2021, 1:36 PM), https://arstechnica.com/gadgets/2021/07/todays-massive-internet-outage-comes-courtesy-of-akamai-edge-dns [https://perma.cc/P33R-Z44L]; Annie Palmer, *Dead Roombas, Stranded Packages an Delayed Exams: How the AWS Outage Wreaked Havoc Across the U.S.*, CNBC (Dec. 9, 2021, 10:51 AM), https://www.cnbc.com/2021/12/09/how-the-aws-outage-wreaked-havoc-across-the-us.html [https://perma.cc/RVW5-5HTE]; *see also supra* notes 1–5 and accompanying text (discussing the Fastly outage).

45. NUECHTERLEIN & WEISER, *supra* note 4, at 290–92.

46. For a discussion of the uses of W3Techs data, see, e.g., Aakanksha Mirdha, Apurva Jain & Kunal Shah, *Comparative Analysis of Open Source Content Management Systems*, IEEE INT'L CONF. ON COMPUTATIONAL INTEL. & COMPUTATIONAL RSCH. 1–4 (2014), https://ieeexplore.ieee.org/abstract/document/7238337 [https://perma.cc/PUQ4-ELGD] (using W3Tech technology to survey content management system usage).

47. This data was compiled in conjunction with the Internet Society, whose financial support funded aspects of this data collection. The code used for all data collection and analysis is available at https://github.com/elsehow/taaraxtak [https://perma.cc/B2EC-V3XJ].

48. *See* Kimberly Ruth, Aurore Fass, Jonathan Azose, Mark Pearson, Emma Thomas, Caitlin Sadowski & Zakir Durumeric, *A World Wide View of Browsing the World Wide Web*, 22 PROC. OF THE 22ND ACM INTERNET MEASUREMENT CONF. 317, 319 (2022), https://zakird.com/papers/browsing.pdf [https://perma.cc/V9AU-U3FT] ("Prior work has shown

websites people visit and use, as the web is vast.[49] Further, it is difficult to rank websites by popularity. Because a large proportion of all web traffic is automated rather than the result of human use, observing internet traffic is unreliable. Moreover, because patterns of use on the web are highly diverse, samples collected from web users may be inaccurate unless the surveyed population is sufficiently large.[50] Together, these factors make it difficult for researchers to identify which parts of the web matter most to end-users' lived experience of the internet.[51] The data collected from Google Chrome's web browser, the Chrome UX Report (or CrUX) dataset[52] is generally treated by the internet measurement community as the most accurate depiction of website popularity worldwide because the dataset is large and the behavior observed originates from human users.[53]

Using CrUX's list of the world's most popular websites, W3Techs used data traces to determine which CDNs, if any, those websites rely

---

that Google Chrome's public Chrome User Experience Report (CrUX) dataset . . . provides the most accurate perspective on site popularity compared to other public datasets.")

49. *Id.*; *see also* Kimberly Ruth, Deepak Kumar, Brandon Wang, Luke Valenta & Zakir Durumeric, *Toppling Top Lists: Evaluating the Accuracy of Popular Website Lists, in* ACM INTERNET MEASUREMENT CONF. PROC. 374 (2022), https://dl.acm.org/doi/pdf/10.1145/3517745.3561444 [https://perma.cc/XJ7M-DDUQ] (describing problems identifying the actual websites users visit based on comparisons to different sources).

50. Ruth et al., *A World Wide View*, *supra* note 48, at 319.

51. *See id.* (describing how factors like identification problems and representativeness limited the conclusions that could be drawn from a survey of internet use).

52. For an overview of CrUX and how it works, see generally *About Chrome UX Report*, GOOGLE.COM (June 23, 2022), https://developer.chrome.com/docs/crux/about [https://perma.cc/BJD4-NK4A] (describing how Google collects Chrome UX data).

53. To understand why CrUX is the most appealing option in surveying the user-facing web, it is helpful to consider available alternatives. Beyond instrumenting end-user devices (as Chrome does), researchers are left to instrument shared internet infrastructure, such as DNS servers or ISPs. Since much traffic on the web is automated, instrumenting that infrastructure results in noisy data. *See supra* notes 49–50 and accompanying text. The only common internet infrastructure that reliably distinguishes between human and automated traffic are CDNs, which must perform this function to deliver DdoS protection. *See* Ruth et al., *Toppling Top Lists*, *supra* note 49, at 375 (explaining that CDNs like Cloudflare provide "authoritative data" on internet traffic because of the way that it acts as an "authoritative DNS provider and reverse proxy for customer websites"). However, instrumenting CDNs only captures data about that CDN's customers, which is insufficient for our purposes; we require a view of websites that use many or no CDNs. CrUX provides such a dataset. However, it is not without limitations. It may over-sample desktop users, for whom the Chrome browser is relatively common, as compared to mobile users, particularly iPhone users for whom Safari is the default browser. Ruth et al. conclude that the CrUX dataset is likely the most representative available dataset of the set of websites Internet users visit. *See* Ruth et al., *A World Wide View*, *supra* note 48, at 319.

upon.[54] Specifically, W3Techs requests each website and inspects the responses to identify which technologies were used to build and deliver that website.[55] When a user requests a website, the website is not delivered as single files, but rather as many discrete components known as *packet*s.[56] These packets are received and "assembled" by the requesting computer's web browser to construct an interactive webpage for the user.[57] Each of these packets contains a "header," which includes metadata about the packet and its content.[58] Aspects of these headers are analogous to the "from" and "to" fields on mailed parcels. The internet's hypertext transfer protocol—or HTTP, as in *http*://www.google.com—relies on this information to transmit and assemble packets correctly. Among that metadata is the IP address that originated the packet. By cross-referencing these IP addresses to lists of known service providers, W3Techs can determine which packets were delivered by particular CDNs. W3Techs then computes the proportion of websites in the sample that use a particular CDN to deliver content. If Cloudflare, for example, is found to deliver 76 percent of the packets in the survey, we would estimate its market share at 76 percent.

From this data, we can compute an overall picture of the market for CDN services over time. We have compiled historical data on the market for CDNs dating from January 2017 to December 2022. The data presented here has been used by, for example, the Internet Society—a nonprofit organization founded by two of the internet's so-called "founding fathers," Vint Cerf and Bob Kahn[59]—to describe the state of the internet.[60]

---

54. For an overview of W3Techs, see generally *W3Techs–World Wide Web Technology Surveys*, W3TECHS, https://w3techs.com [https://perma.cc/QD6K-U276] (describing W3Techs and listing previous surveys).

55. *Frequently Asked Questions*, W3TECHS, https://w3techs.com/faq [https://perma.cc/BRB7-YPFE].

56. *How the Web Works*, MOZILLA CORP., https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/How_the_Web_works [https://perma.cc/B2TA-29PX] (last updated July 24, 2023).

57. *Id.*

58. *HTTP Headers*, MOZILLA CORP., https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers [https://perma.cc/765Z-6RWD] (last updated July 19, 2023).

59. *See, e.g.*, Adam Mann, *Father of the Internet, Vint Cerf, on Creating the Interplanetary Internet*, WIRED (July 5, 2013, 11:19 AM), https://www.wired.co.uk/article/vint-cerf-interplanetary-internet [https://perma.cc/9JND-JTZJ].

60. *Market Concentration*, INTERNET SOCIETY: PULSE, https://pulse.internetsociety.org/ con

Our "whole-web" figures use, as a baseline, the 15,000,000 most popular websites as measured by the CrUX dataset.[61] Of those top 15,000,000 websites, analysis of packets delivered on behalf of those websites reveals that 23.6 percent use a CDN to deliver service. We treat these websites as the effective market for CDN services. Under this model, 91 percent of the packets delivered in response to requests are delivered by one of three CDNs: Cloudflare, Fastly, and Amazon.[62]

We acknowledge some drawbacks of our methodological approach. For example, there is some uncertainty about how, exactly, to calculate the relevant market. As suggested, there is a long tail of less-popular websites—about three-quarters of all websites—that use no CDN at all. We can only speculate as to why these websites do not use a CDN. Perhaps it is because they do not need one, as they generate too little traffic to benefit from a CDN's services. Indeed, this hypothesis is consistent with our data: a website's likelihood of using a CDN plummets as its popularity (ranked by CrUX dataset) declines. Every website in the top one thousand websites uses a CDN, and 99.9% of all websites in the top ten thousand do. Hence, while non-CDN using websites account for a good proportion of all websites, they likely account for a vastly smaller portion of web traffic. Although the exact distribution of traffic across websites is difficult to calculate,[63] web traffic is commonly observed to be power-law distributed, an observation that, applied to top websites, would suggest, roughly, that approximately 80 percent of all web traffic is generated by 20 percent

centration [https://perma.cc/C68V-U6QK]; *Internet Atlas*, UC BERKELEY CNTR. FOR LONG-TERM CYBERSECURITY, https://cltc.berkeley.edu/program/internet-atlas [https://perma.cc/3SQA-3J7S].

61. As noted, this paper's study relies upon CrUX, which internet measurement research has generally found to be the most comprehensive and accurate list of top websites available. *See* Ruth et al., *supra* note 48. Historical data generated by May 1, 2022, relies upon Alexa Internet rankings, a public resource, commonly used in other internet measurement research. *See* Allison McDonald, Matthew Bernhard, Luke Valenta, Benjamin VanderSloot, Will Scott, Nick Sullivan, J. Alex Halderman & Roya Ensafi, *403 Forbidden: A Global View of CDN Geoblocking*, *in* PROC. OF THE INTERNET MEASUREMENT CONF. 218 (2018), 8532.3278552https://dl.acm.org/doi/10.1145/3278532.3278552 [https://perma.cc/M794-FNXG] (using of Alexa Internet rankings to measure CDN usage). The service, founded in 1996 by Internet Archive steward Brewster Kahle, was acquired by Amazon in 1999. While Amazon discontinued this Alexa service on May 1, 2022, the data we report in this paper was generated prior to its shutdown. *See* themadprogramer, *Pulling Rank: The Legacy of Alexa Internet*, THE DATA HORDE, https://datahorde.org/pulling-rank-the-legacy-of-alexa-internet [https://perma.cc/LVU7-PKAL].

62. *See infra* app., tbl.1.

63. *See supra* note 53.

of all websites.[64] Indeed, industry reports suggest that about 50 percent of all web traffic is drive by six websites alone.[65] Hence, it seems quite likely that most, if not all, of these sites do not use a CDN simply because they do not need one. Though these sites might eventually join the market for CDNs, they have not yet, and so we exclude them from the rest of our analysis.[66]

Moreover, some service providers are so large that they build proprietary CDNs, which are also excluded from our analysis.[67] For example, Meta maintains a CDN for its own properties.[68] Netflix also runs a CDN to deliver its streaming services, partnering with ISPs to cache content in a way that minimizes the distance between Netflix's viewers and Netflix's content.[69] These large providers have exited the market for CDNs for a different reason: they have achieved such tremendous scale that are best served by vertically integrating these caching and security services. And since these large content providers do not sell their CDN service to other entities (nor buy it from anyone

64. *See* Aniket Mahanti, Niklas Carlsson, Anirban Mahanti, Martin Arlitt & Carey Williamson, *A Tale of the Tails: Power-Laws in Internet Measurements*, 27(1) IEEE NETWORK 59, 61 (2013) https://nymity.ch/tor-dns/bibliography/pdf/Mahanti2013a.pdf [https://perma.cc/W8PE-TPEB] (listing examples of internet measurements that abide by this 80/20 principle)**.** The assumption of a roughly power-law distribution of website popularity is consistent with the estimate that 50 percent of all internet traffic is driven by the top 6 websites. *See* Sandvine, *When Netflix and MAMAA Rule the Internet*, PHENOMENA: THE GLOBAL INTERNET PHENOMENA REP., Jan. 2023, at 10.

65. Sandvine, *supra* note 64, at 10. We further clarify that our analysis does not capture information about the relative volume of traffic generated by the websites any one CDN serves. It may be the case, for example, that Akamai serves as much or more *traffic* than Cloudflare, even if it serves fewer websites. Cloudflare may, for example, have captured a long tail of smaller content providers, while Akamai serves the largest. Under our analysis, Akamai would have a smaller market share that Cloudflare (given our measure of the denominator).

66. If they ever faced problems, they could sign up for Cloudflare's free CDN service. *Security, Performance, and Reliability–All in One Package*, CLOUDFLARE, https://www.cloudflare.com/plans [https://perma.cc/29RY-QWDX] (listing Cloudflare plans).

67. Other studies of CDNs have studied these providers. *See* Petros Gigis, Matt Calder, Lefteris Manassakis, George Nomikos, Vasileios Kotronis, Xenofontas Dimitropoulos, Ethan Katz-Bassett & Georgios Smaragdakis, *Seven Years in the Life of Hypergiants' Off-nets*, *in* ACM SIGCOMM CONF. PROC. 516–33 (2021).

68. *See* Huapeng Zhou, Linpeng Tang, Qi Huang & Wyatt Lloyd, *The Evolution of Advanced Caching in the Facebook CDN*, META (Apr. 7, 2016), https://research.facebook.com/blog/2016/4/the-evolution-of-advanced-caching-in-the-facebook-cdn [https://perma.cc/U7FJ-9784] (describing the Facebook Content Distribution Network, "FBCDN").

69. *See Open Connect*, NETFLIX, https://openconnect.netflix.com/en [https://perma.cc/GR98-5AXA] (noting that Netflix Open Connect "partner[s] with over a thousand ISPs to localize . . . traffic").

else), they too have exited the market for CDN services. While our analysis does not capture the use of these vertically integrated CDNs, this omission is acceptable for our present purposes, as the companies who use proprietary CDNs are not customers in the market for CDN services.[70]

In all, any uncertainty about the precise baseline against which to measure a CDNs' share does not undermine our basic point. Our data captures information about those popular websites that require a CDN yet lack the capacity to build and deploy one for themselves. Even Spotify, a streaming content company that earned over 11 billion dollars in revenue in 2022, uses commercial CDNs to deliver its content.[71] Hence, our analysis covers 23.6 percent of all websites in the CrUX dataset—and, as noted, these websites comprise a significant majority of all web traffic. Any remaining uncertainty serves to reinforce our point, also elaborated below, that there should be greater transparency in how traffic flows across the internet for a variety of purposes.

B. Results and Analysis

1. *The Rise of—and the Consolidation in—the CDN Market*. In the last thirty years, CDNs have grown rapidly along dimensions of both size, that is, volume of data served and scale, meaning number and variety of users. Before Akamai, a leading CDN, was founded in 1998,[72] no website used a CDN. Today, *every* website in the most

70. Amazon is a bit of special case: Amazon's own web properties seem to employ, predominantly, Amazon Web Services, or AWS, CloudFront for CDN services—and AWS is also sold to as a commercial CDN service to other entities. Stated otherwise, amazon.com is vertically integrated with AWS CloudFront, and AWS CloudFront is also available to other websites. We treat AWS CloudFront as within the scope of our analysis, for Amazon-owned and non-Amazon properties alike.

71. *See* Spotify Engineering, *How Spotify Aligned CDN Services for a Lightning Fast Streaming Experience*, SPOTIFY (Feb. 24, 2020), https://engineering.atspotify.com/2020/02/how-spotify-aligned-cdn-services-for-a-lightning-fast-streaming-experience [https://perma.cc/5YRV-KE2A] (explaining how Spotify got its "new CDN service up and running quickly on Fastly").

72. *See Company History*, AKAMAI, https://www.akamai.com/company/company-history [https://perma.cc/78HM-XW2P] (noting that Akamai was incorporated in August 1998); *see also* Gigis et al., *supra* note 67, at 516 ("Since 2000, Akamai has deployed their servers in hundreds of networks around the globe.").

popular[73] one thousand websites uses a CDN, and over 99.9 percent percent of the top ten thousand websites do.

These results reflect the major change in the structure of the internet described above. To appreciate the magnitude of this change, compare Cogent Communications, considered by many to be a Tier-1 provider,[74] with Akamai, a leading CDN. In March 2022, coverage of Cogent's decision to terminate service to Russia in the wake of the conflict in Ukraine noted that Cogent carried roughly 25 percent of the world's internet traffic.[75] Akamai is responsible for roughly the same volume of traffic.[76] Yet the way these two entities handle traffic is substantially different. Cogent is agnostic about the bits that travel over its networks. To invoke a perhaps tired and tortured metaphor: Cogent

---

73. *See supra* note 61 and accompanying text (describing CrUX, which lists the most popular websites available).

74. Although no firm list of Tier-1 providers exists, a rough proxy for Tier 1 status is the size of its "customer cone"; that is, the number of customers the network can reach either directly or indirectly. *See* Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giostas & KC Claffy, *AS Relationships, Customer Cones, and Validation*, *in* ACM INTERNET MEASUREMENT CONF. PROC. 243–56 (2013), https://dl.acm.org/doi/pdf/10.1145/2504730.2504735 [https://perma.cc/R5Z9-H5GN]. The Center for Applied Internet Data Analysis (CAIDA) at U.C. San Diego assigns Cogent an AS Rank of 3, indicating it has the third highest number of direct and indirect customers by peering relationship of any network of the 75,337 that CAIDA tracks. *See Autonomous Systems Rank*, THE CTR. FOR APPLIED INTERNET DATA ANALYSIS, https://asrank.caida.org [https://perma.cc/GA7Q-PN83].

75. Igor Bonifacic, *Internet Backbone Provider Cogent Cuts off Service to Russia*, ENGADGET (Mar. 5, 2022, 5:31 PM), https://www.engadget.com/cogent-communications-223135454.html [https://perma.cc/T4LE-L6VD].

76. 2015 OIO, *supra* note 14, at 5687–88 ¶ 197 n.491 (quoting Akamai Techs., Inc., Comment on Proposed Rule for Protecting and Promoting the Open Internet Framework for Broadband Internet Service (July 15, 2014) at 4 ("At any given time Akamai delivers between 15–30% of all web traffic, resulting in over two trillion interactions delivered daily.")).

Keen readers may discern a discrepancy between the statistic reported here and our results presented later: Here, we note that Akamai is responsible for about 25 percent of the internet's traffic; whereas later we ascribe to Cloudflare (a competitor) over 80 percent the market (for a total exceeding 100 percent). What gives? We can resolve these readers' apparent dilemma by noting the distinction between the *volume* of traffic served and the *share* of websites served. Akamai delivers lots of content—i.e., it seems responsible for a high volume of traffic—across several large websites (such as eBay). By contrast, Cloudflare appears to deliver content for many websites, even if many of them are quite small (in terms of traffic volume). Stated simply, Akamai rules the top (fewer websites, higher volume) while Cloudflare rules the long tail (more websites, less volume).

We note one further caveat: The explanation we offer here is our best understanding of the data we have (including Akamai's self-reported traffic statistics). But, because of the limited visibility into global internet traffic (a problem of transparency we address more fully *infra*), we can only observe certain slices of observable data (namely, what the data that is ultimately delivered to us and other end-users) and must make inferences about the rest.

offers the rough equivalent of basic postal delivery. Cogent's customers provide Cogent with packages, namely, packets, that have "to" and "from" addresses. Cogent delivers those packages for a price. Akamai, by contrast, is in the business of logistics, including, but not limited to, postal delivery. Akamai offers to warehouse its customers' data, and, in response to user orders—that is, internet requests—Akamai generates packages, completes the "to" and "from" fields, and assumes responsibility for their delivery. For that last step—delivery—Akamai may, but (as we elaborate below) need not, purchase bandwidth from Cogent or other Tier-1 providers.

Akamai and other CDNs thus intermediate the relationship between the internet's core and its users. And given their broad popularity, they play this intermediary role for a vast proportion of internet activity. Even if, then, competition among Tier-1 providers remains strong, the practical benefits conferred by CDNs compels many web properties to rely on the full array of services that CDNs provide, and not the mere delivery services that transit providers sell. Indeed, the widespread prevalence of CDNs, especially among the most popular web properties, may help to confirm their status as a practical necessity.

Hence, CDNs' customers are largely companies that use the internet to conduct business. Such companies encompass a wide range of products and services: some may be e-commerce websites; others might be professional services firms; still others might offer ad- or subscription-funded news, reporting, or commentary. To be sure, some large and technology-first businesses, such as Google and Netflix, operate their own proprietary CDNs.[77] But many enterprises employ an outside CDN to facilitate their business.[78] And many smaller entities may not even understand that, when they build a website, they also

77. *See Cloud CDN and Media CDN*, GOOGLECLOUD, https://cloud.google.com/cdn [https://perma.cc/DL8B-KC3D] (describing "Google's content delivery networks"); *Open Connect*, NETFLIX, https://openconnect.netflix.com/en [https://perma.cc/J398-CQ4S] (explaining that Open Connect partners with ISPs "to deliver [its] content").

78. A comprehensive list of companies that run proprietary CDN infrastructure is difficult to come by. However, hypergiants like Google, Netflix, Meta, and Apple are known to deploy their own infrastructure. *See* Gigis, *supra* note 67, at 516. These providers also serve their own traffic. *See id.* ("Some [hypergiants] also install servers . . . to serve . . . their customers."). For an internet-based company looking to serve content they originate, they must either host it with a large company or pay one of a small handful of CDNs. The approximately 3.7M websites that use some CDN service, identified in our methods, fall into the latter category: we can identify them because they use a well-known commercial CDN to provide service.

implicitly purchase CDN services. These services are, for example, packaged into the widely-available and widely-used online web design services. If, for example, a small business employs Shopify to manage its webstore, that small business becomes reliant on whatever CDN(s) that Shopify has employed.[79]

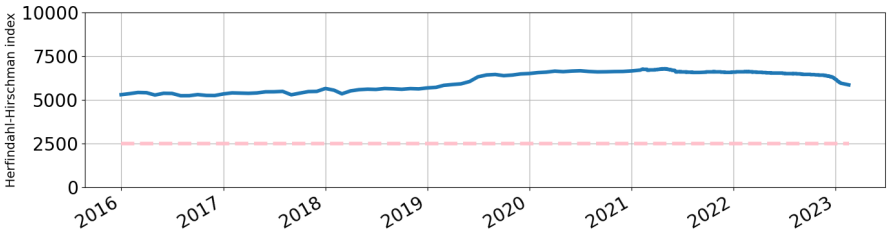

Figure 2. The Herfindahl–Hirschman index (HHI) of the market for CDNs across the whole web, from 2016 to 2023. The dotted line shows the 2,500 threshold designating a highly concentrated market.

Of these CDNs, Cloudflare is far and away the most dominant. Of websites using a CDN—a category that includes, recall, all the top one thousand websites, over 99 percent of the top ten thousand websites, and nearly one-quarter of the top 15,000,000 websites—over 70 percent rely on Cloudflare for such services.[80] After Cloudflare, Fastly serves about six percent of the market, and Amazon's Cloudfront serves just over five percent. In all, as Figure 3 and Appendix Table 1 suggest, only eleven providers control 99 percent of the CDN market.

Antitrust authorities and economists sometimes measure market concentration using the Herfindahl–Hirschman Index, or HHI, which scales from 0 to 10,000.[81] Values over 2,500 generally denote a "highly

79. Some applications may run their own special-purpose CDNs available only to their customers. For example, Shopify runs a CDN specifically for user images. *See Shopify's Content Delivery Network*, SHOPIFY.COM, https://cdn.shopify.com [https://perma.cc/7DZH-5XSJ] (describing how Shopify's CDN transforms users' files). CDNs like this seem unlikely to meaningfully improve consumer choice; even if Shopify users can opt out of using Shopify's CDN for their images, they are still obligated to rely on Shopify's infrastructure for their application, including any CDNs on which Shopify relies.

80. See discussion *supra* note 76 for an explanation of an apparent—but resolvable—conflict in the data reported here.

81. U.S. DEP'T OF JUST. & FED. TRADE COMM'N, HORIZONTAL MERGER GUIDELINES § 5.3 (2010). The Department of Justice's newly proposed merger guidelines recommend lowering the threshold for "highly concentrated" markets to HHI values of 1,800. *See* U.S. DEP'T OF JUST. & FED. TRADE COMM'N, DRAFT MERGER GUIDELINES 6 (2023). Under either approach, this market easily qualifies.

concentrated" market.[82] The market for CDN services currently weighs in at an HHI of 5,846 (Figure 2), and it has remained well above this benchmark for much of the regulatory history treating this market as presumptively competitive.[83] But these market share statistics strongly suggest a concentrated market.

We acknowledge that any complete analysis of market concentration would be more complicated and would include more difficult questions of market definition and the substitutability of other options. But the practical benefits conferred by CDNs, together with the indicia of concentration and pervasiveness, suggest that this market is far less competitive than the one usually ascribed to the Tier-1 providers at the core of the core of the internet.

Stated differently, the single "market for internet traffic exchange" seems, instead, to be two markets: one characterized by the commodity bandwidth, and a second, "interpositioned" between this traditional core and its end users, offering a wide range of "in network processing" services, including caching and security.[84] And control of this second market—one whose caching and security services are a seeming necessity for any major web property—is concentrated among a small handful of CDNs.

82. HORIZONTAL MERGER GUIDELINES § 5.3.

83. For reference, the markets for other internet-hosting-related services—web hosting, for example—are far more competitive, with an HHI of only 153. HHI is computed as the sum of the squares of each firm's market share. *Id.* The resulting value ranges from 0 to 10,000. A high HHI indicates a few firms that dominate; a low HHI indicates many firms with small market shares. As a rule of thumb, the antitrust authorities have considered values above 2,500 to indicate a low degree of competition. *See id.* (designating a highly concentrated market as one with an HHI above 2,500).

84. *See* Scott Shenker, Lloyd Brown, Emily Marx, Christopher Branner-Augmon, William Lin, Catherine Lu, Mark Theis, Zach Van Hyfte, Mark Zhang, Emmanuel Amaro, Ezra Kissel, Inder Monga, Ben Pfaff, Debnil Sur, Arvind Krishnamurthy, James McCauley & Aurojit Panda, Creating an Extensible Internet Through Interposition 1 (2023) (unpublished manuscript) (on file with authors) (noting "[t]he Internet's architectural stasis," but explaining the rise of user-facing networks that "intercept and process all traffic that is intended for" cloud and content providers that user-facing private networks).

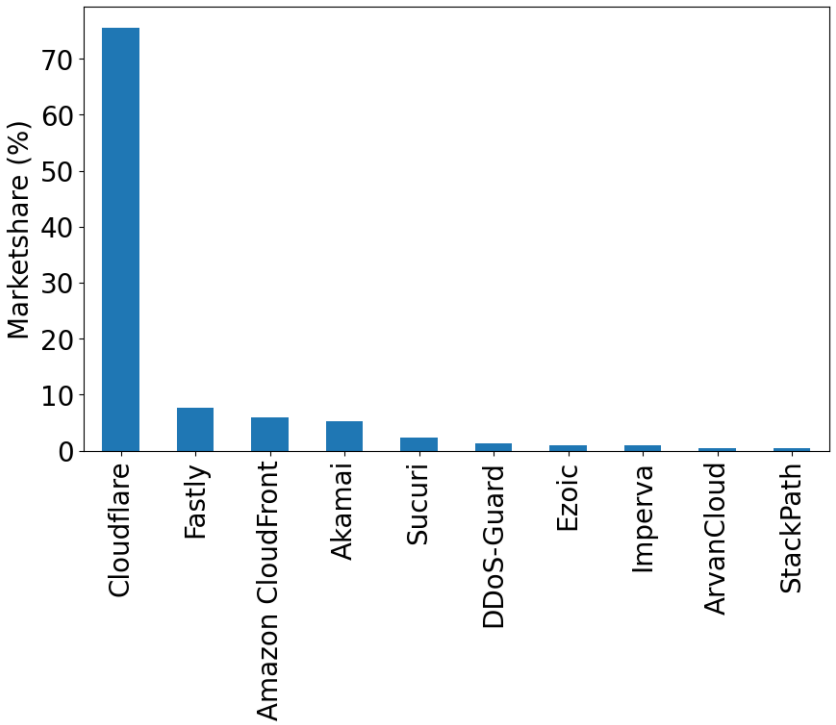

Figure 3. The marketshare of the top ten CDNs worldwide, across all CDN-using websites in the sample.

Moreover, like some other internet infrastructure companies and data intensive services, there is a feedback effect to market consolidation, as CDNs benefit from both network effects and economies of scale.[85] Consider Netflix, which runs its own proprietary CDN.[86] By observing geographic patterns in content consumption, Netflix can cache content files geographically close to the people most likely to watch it. Netflix might, for example, cache episodes of *Sacred Games*, featuring some famous Bollywood actors, in India and in cities with relatively high populations of Indian-Americans—thereby

85. *See, e.g.*, NUECHTERLEIN & WEISER, *supra* note 4, at 3–9 (describing the relationship, in telecommunications contexts, between consolidation and network effects as well as economies of scale and density); *see also* Geoff Huston, *On Internet Centrality and Fragmentation*, RIPE LABS (Aug. 7, 2023), https://labs.ripe.net/author/gih/on-internet-centrality-and-fragmentation [https://perma.cc/8DJC-ARGZ] ("Markets with a high reliance on data have positive feedback loops.").

86. *See Open Connect*, NETFLIX, https://openconnect.netflix.com/en [https://perma.cc/AM24-8HTP] ("Netflix . . . operates its own CDN, called 'Open Connect.'").

yielding faster load times, and better experiences, for users. Some CDNs extend this process across a wide range of content. And as a CDN grows, it gains an increasingly comprehensive view of global internet traffic—which, in turn, helps it to more efficiently cache content, and to more rapidly and accurately respond to emerging attacks.[87]

While we have used relatively simple examples in our exposition here—NBA teams as suggestive of geography, or correlations between demographics and particular content—the truth is that much of the logic behind caching is automated, some driven by machine learning algorithms that become more powerful as a CDN's scale and scope expand.[88] Hence, some CDNs even provide some services for free, offering the advantages of a more centralized architecture to smaller or newer companies, thereby growing their view of the internet traffic—all while fueling superior service to their paying and nonpaying customers alike.

2. *Towards Opaque Networks*. The results described above suggest one further trend: not only do these results highlight the rise of an increasingly concentrated CDN market, one that intermediates the relationship between the internet's traditional core and its users; they also highlight how the internet's tiers are "flattening."[89] This flattening reflects a shift in the internet's topology. Recall that, under the first model of the internet we described, so-called Tier-1 and Tier-2

87. *Cf.* Omer Yoachimik, Julien Desgats & Alex Forster, *Cloudflare Mitigates Record-Breaking 71 Million Request-Per-Second DDoS Attack*, THE CLOUDFLARE BLOG (Feb. 13, 2023) https://blog.cloudflare.com/cloudflare-mitigates-record-breaking-71-million-request-per-second-ddos-attack [https://perma.cc/8FGM-5VN7] (explaining Cloudflare's response to dozens of DDoS attacks).

88. *See* Tejas N. Narechania, *Machine Learning as Natural Monopoly*, 107 IOWA L. REV. 1543, 1584 (2021) (explaining the "virtuous cycle" of machine learning-based applications, which become more accurate and effective the more data they collect and analyze).

89. The observation that the traditional hierarchy of ISPs was "flattening" due to the increased importance of private networks relative to the Tier-1 dates to at least 2008. *See* Phillipa Gill, Martin Arlitt, Zongpeng Li & Anirban Mahanti, *The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles, or Contrived Collapse?*, *in* INT'L CONF. ON PASSIVE AND ACTIVE NETWORK MEASUREMENT 1–10 (2008). Measurement in the intervening years has captured the steady development of this trend. *See* Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide & Farnam Jahanian, *Internet Inter-Domain Traffic*, *in* PROC. OF THE ACM SIGCOMM CONF. 75–86 (2010); Todd Arnold, *Unpacking a Flattened Internet*, PAC. NET. CENT. APNIC (Dec. 4, 2020), https://blog.apnic.net/2020/12/04/unpacking-a-flattened-internet [https://perma.cc/U678-WU5G] (noting how new modes of connectivity among "many of the Internet's networks" bypass "the traditional Internet hierarchy's apex").

providers existed at the apex and middle—respectively—of an imagined hierarchy. Internet traffic was often conceptualized as starting at the bottom of this hierarchy with a local ISP, ascending a stack of tiered providers up to Tier-1 providers, and then back down again.

The internet measurement literature, combined with our original research above, casts further doubt on this model of the internet's core, and not only because individual business relationships with these commodity bandwidth providers is increasingly intermediated by CDNs. In addition, recent findings show that users' relationships with these businesses are also intermediated by CDNs: CDNs can reach over 76 percent of all internet addresses without traversing a Tier-1 or Tier-2 network *at all*.[90] Where web requests used to hop from an ISP to a series of backbone providers and then to another ISP and back again, most requests for web content are now fulfilled simply by moving from an ISP to a CDN's proprietary network and back.[91] In this new, flatter internet, CDNs and other providers broker their own connections with users. They use their own proprietary networks, thereby bypassing the Tier-1 providers that have traditionally made up the internet's core.[92] Hence, though legal scholars and policymakers have long thought of the internet's interior as competitive—internet traffic exchange provisioned by diverse classes of infrastructure providers who compete with each other on terms of price, speed, and reliability—the story seems instead to focus increasingly on CDNs, including CDNs that use their own proprietary networks.

---

90. Todd Arnold, Jia He, Weifan Jiang, Matt Calder, Italo Cunha, Vasileios Giotsas & Ethan Katz-Bassett, *Cloud Provider Connectivity in the Flat Internet*, *in* ACM INTERNET MEASUREMENT CONF. PROC. 2 (2020), https://dl.acm.org/doi/10.1145/3419394.3423613 [https://perma.cc/GXF9-LUA6]. By issuing traceroutes from within cloud providers' networks, Arnold et al. assembled a topology (essentially, a network diagram) to understand how (i.e., by which routes) these major cloud providers reached the rest of the internet. *Id.* By cross-referencing these results against known routing maps published by CAIDA's AS-relationship dataset, the authors estimate that these cloud providers can deliver content to 76 percent of all internet addresses "hierarchy-free," i.e. without traversing the traditional Tier-1 networks. *Id.* Phrased differently, for 76 percent of hosts on the internet, these cloud providers do not need to participate in the Tier-1 market for transit.

91. *See* Arnold, *supra* note 89 (explaining how networks now "establish direct connectivity between each other, effectively bypassing the few dozen networks that occupy the traditional Internet hierarchy's apex").

92. Some in the internet standards community refer to this phenomenon as "the death of transit." *See* Geoff Huston, *The Death of Transit?*, APNIC (Oct. 28, 2016), https://blog.apnic.net/2016/10/28/the-death-of-transit [https://perma.cc/RN3D-ED7W].

We readily acknowledge that there is a long tail of websites and internet endpoints for which the old model of transit still rings true. About three-quarters of websites, as noted, use no CDN provider at all.[93] We do not mean to suggest that these transit services are obsolete altogether. But, the most popular websites responsible for the vast majority of internet traffic[94] as well as of the internet's economic and social value, rely upon CDN services. And the majority of internet users can access a CDNs' network directly.[95] Viewed together, it seems that for a substantial proportion of these popular applications and websites, internet traffic moves directly from a consumer's ISP to the CDN's private network and back again, circumventing the traditional transit system entirely.

One can hardly overstate the significance of this change for the internet's structure: a system once characterized by a robust array of competing network services providers operating on public networks has been replaced by a concentrated set of ISPs interconnected with a concentrated set of CDNs, relying on its own private network.

III. THE CONSEQUENCES OF THIS NEW CORE

The effective core of the internet has thus shifted away from public carriage over the networks of Tier-1 providers and towards the private networks of CDNs. This new core is critically different from the old core composed of a competitive market of Tier-1 providers. Tier-1 providers sell a fungible service—the ability to deliver packets from one address to any other. The services CDNs provide are less fungible—they are the aggregate of a CDN's capacity to deliver traffic *and* its ability to securely, often algorithmically, manage that traffic.[96] Indeed, as noted, CDNs rely on proprietary models to both cache and filter traffic. The result is an internet system whose behavior is less

93. *See supra* note 62 and accompanying text.

94. *See supra* note 73 and accompanying text.

95. We say practically because we do not know—and cannot measure—exactly how data arrives to the locations from which it is served. That is, to return to our NBA example, before a fan in California can watch a locally-served video highlight of a game played in, say, New York, the video recording must travel, at least once, from New York (where it was originally recorded) to California (where it is stored, i.e., cached).

96. Indeed, just as we have described the CDNs as intermediating the traditional relationship with the market for internet traffic exchange, scholars from the communications and computer networking community have described CDNs as offering the "interposition of in-network processing," i.e., the algorithmic management of traffic for caching and cybersecurity (among other) purposes. *See* Shenker et al., *supra* note 84, at 1 ("There is no denying that the interposition of in-network processing in these private networks offers immediate and tangible benefits.").

predictable and scrutable to outside observers, including both users and regulators.

When a user requests a website, what data should be delivered? From where? Are customers in low-income areas served differently from those in high-income areas? Are customers in low-income areas more likely to have their traffic blocked or throttled as suspicious? In the old model of the internet, the internet's core would have little say in such decisions, as competition among Tier-1 providers seemed to force them to set aside such concerns and focus on efficient packet delivery. Instead, individual applications and websites—each with its own filtering, prioritization, and security logic—would decide how to handle such concerns. But today's core—intermediated by CDNs—now takes greater control over such matters.[97]

We reiterate that this new model has helped to deliver better and more secure services to a wide range of the world. It has enabled new applications such as streaming audio and video, even under some capacity-constrained conditions.

But this new model is not costless. The consolidation of these core internet services among private companies using private networks has produced two main externalities.[98]

---

97. The way these models work (and sometimes fail) to filter traffic has raised questions among internet researchers about civil rights and equal access. *See generally*, Anne Jonas & Jenna Burrell, *Friction, Snake Oil, and Weird Countries: Cybersecurity Systems Could Deepen Global Inequality Through Regional Blocking*, 6 BIG DATA & SOC'Y 1, 3–4 (2019). As Anne Jonas and Jenna Burrell explain in their 2019 paper:

> Professionals tasked with preserving online security hope automated identification of patterns of "good" and "bad" usage will produce more accurate and less discriminatory methods for determining who should be able to access their services and who should be flagged as a concern. However, without understanding the systematic social and political conditions that produce differential behaviors online, these systems may continue to embed unequal treatments, and further disguise discrimination behind more complex and less transparent automated assessment . . . The literature on fairness in machine learning has especially considered application domains such as criminal justice, lending, and social services where mechanisms of allocation impinge upon civil rights.

Id.

98. Consolidation among CDNs seems to also be entangled with consolidation in other aspects of the internet's core, such as the Domain Name System (or DNS). The DNS hosts a distributed database that, among other things, maps human-readable "domain names" (e.g., nytimes.com) to machine-readable IP addresses. Although the DNS is itself a decentralized protocol, recent research shows that these DNS services are also increasingly centralized. Among the most centralized are Akamai, Amazon Web Services, and Cloudflare—three of the same organizations that dominate the market for CDN services. *See* Samantha Bates, John Bowers, Shane Greenstein, Jordi Weinstock, Yunhan Xu & Jonathan Zittrain, *Evidence of Decreasing Internet Entropy: The Lack of Redundancy in DNS Resolution by Major Websites and Services*, 1

First, CDNs' consolidated and private infrastructure give rise to *central points of failure* that resist scrutiny and oversight. The CDNs' private networks are black boxes to outsiders, confounding efforts at oversight and risk management.[99] When the internet was characterized by providers offering public carriage, we could more easily map, visualize, and assess the internet's infrastructure. But we now know far less about how the internet's various interconnected networks fit together, largely because we have a very limited understanding of how CDNs route traffic internally, that is, on their private networks for private carriage, and over to one another. Without some window into the CDNs' private networks, our public security and reliability efforts are frustrated.

Second, the increasing consolidation in this market gives rise to *centralized points of control.* These CDNs can—and sometimes do—filter the content available to internet users, sometimes in ways that are invisible to these users and often in ways that consumers may not avoid.[100] Existing regulatory frameworks, moreover, do little to address

---

J. QUANTITATIVE DESCRIPTION: DIGITAL MEDIA 1, 28 (2021) (discussing diversification rates among DNS external providers); *see also* Rashna Kumar, Sana Asif, Elise Lee & Fabian Bustamante, *Each at Its Own Pace: Third-Party Dependency and Centralization Around the World*, 7 PROC. OF THE ACM ON MEASUREMENT & ANALYSIS OF COMPUTING SYS. 1, 2 (2023). How is it that this ostensibly different, and ostensibly decentralized component of the internet's core is increasingly centralized among the same providers? It is for good reason, as CDN providers bundle DDoS protection in their DNS service, offering cost savings and convenience for cybersecurity services throughout various aspects of the internet's technical systems. But this increasing consolidation reproduces the concerns for fragility (among others) described throughout this Part.

99. For example, Gigis et al. investigated the structure of major CDNs. To do so, they had to reverse engineer aspects of the CDNs' mechanisms, using a side-channel (characteristic DNS records that act as "fingerprints") to identify CDNs' endpoints within ISPs. The research took a total of seven years to complete. Researchers' difficulty in seeing even the edges of CDNs' networks speaks to their opacity: understanding how those endpoints are networked to one *another* likely requires insider access to CDNs' networks and operation. *See* Gigis et al., *supra* note 67, at 517–18.

100. A structural biologist got stuck in a "Kafka-esque" loop when trying to access scientific software, in which Cloudflare's anti-bot software required he prove himself to be human and, in so doing, triggered its anti-bot software. *See* James Hawley, *Blocked by Cloudflare*, JAMES HAWLEY BLOG (Aug. 7, 2023) https://jrhawley.ca/2023/08/07/blocked-by-cloudflare [https://perma.cc/34SU-V6SX]; *see also* Jonas & Burrell, *supra* note 97, at 4 ("In each of these cases, people can find their intentions online thwarted by their locale . . . the decisions of corporate actors to restrict access based on geo location.").

concerns about these decisions, giving rise to possibilities for abuse by CDNs and creating targets of opportunity for bad actors.[101]

A. Central Points of Failure

We begin with the risk, noted above, that consolidation yields central points of failure. Our opening example, in which an error at Fastly led to a cascade of problems for properties across the internet, helps to highlight this risk. What caused these widespread failures—even among websites that had no apparent commercial relationship with Fastly?

A closer look at the software supply chains that help to form the internet's content is instructive. Individual webpages, numbering in the *billions*, are built in real-time atop *thousands* of different software supply chains—assembling HTML and JavaScript code alongside specialized fonts and images, among other resources, many of which rely upon some combination of *hundreds* of common web development tools. And many of these tools and resources are housed and delivered by one of the approximately *eleven* CDN providers described above. Hence, disruptions at the base of this inverted pyramid—that is, at one of the CDNs—can send ripple effects through the entire supply chain, yielding large-scale and difficult-to-predict patterns of failure.

101. *Cf.* Karen Kopel, *Operation Seizing Our Sites: How the Federal Government Is Taking Domain Names Without Prior Notice*, 28 BERKELEY TECH. L.J. 859, 860 (2013) (describing how Immigrations and Customs Enforcement (ICE) can use court orders to make online content inaccessible, if only in a relatively blunt way).

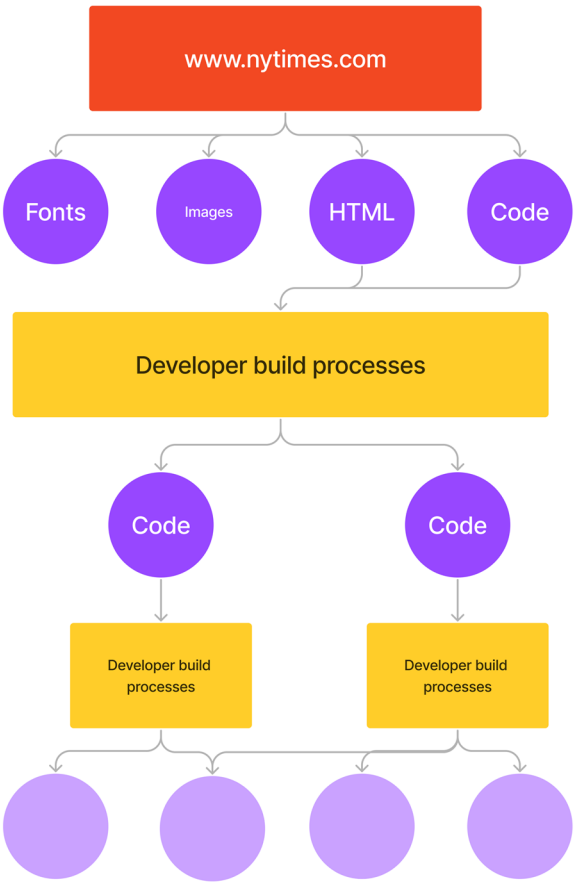

Figure 4. A schematic diagram of a typical software supply chain. Websites that appear in your browser rely on a variety of static assets (like images and fonts) as well as software assets (like build tools). Those software assets themselves rely on development processes, which themselves rely on software assets, creating a recursive supply chain of "nested" dependencies that can be dozens if not hundreds of layers deep. Assets in purple represent reliances on content that is likely stored with one or more CDN provider(s). A failure to deliver any of these assets could cause all downstream products to become unavailable, behave unpredictably, or become impossible to update. As the number of providers decreases, the impact of an outage in any one provider increases. Where a server outage in the Tier-1 model of the internet might affect thousands of websites, one CDN failing might affect millions.

Moreover, these failures can extend far beyond the scope of any clear, existing commercial relationships. Recall that Fastly has captured only about 5 percent of the CDN market; and yet an error there gave rise to effects felt far beyond such a footprint. How? Again, one can look to the internet's software supply chains for an answer. Some websites will not load—they will break altogether—if even one component in the supply chain does not fall into place. For example, even if the Financial Times does not directly employ Fastly's services for its consumer-facing news sites, something in its software supply chain might. It might, for instance, require a font hosted on a CDN. If that CDN suffers from some error, then users may be unable to load *any Financial Times* content. For want of a font, the whole site is lost.[102]

Consolidation can thus cause problems at any one CDN to echo across the web's software supply chain in unpredictable ways. These problems are compounded by the private nature of the CDNs' networks. The public knows very little about how individual CDNs are physically connected to each other. Stated otherwise, we not only lack information about exactly what particular CDNs host, and for whom, we also lack information regarding the internal network connection within and across distinct CDNs. Recall that even Amazon was affected by Fastly's outage—even though Amazon has its own CDN. Hence, even large providers that maintain their own infrastructure have upstream dependencies that may rely on other providers. This interdependence illustrates the complexity of this network of reliance. In short, providers of all scales may be systemically vulnerable to one another.

Since the internals of CDN networks are private—as are the mechanisms by which traffic is routed within them—it is difficult to map the logical routes in the software supply chain. That makes it difficult for risk managers within organizations, such as cybersecurity professionals, to manage risk within their firms. It also makes it difficult for public sector risk managers, such as Cybersecurity and Infrastructure Security Agency (CISA) officials, to manage sector- or economy-wide risk. In short, this opacity imposes tremendous

102. Such problems, moreover, can happen at any point in the supply chain. Or the error might happen upstream of the user: the developers who build the Financial Times website may rely on some tool, or set of tools, themselves stored on CDNs. If that goes down, no one will be able to produce the Financial Times–or any of the other websites that rely on that tool.

structural challenges on any attempt to map these physical and logical routes—and to plan for troubles on those routes.[103]

Consider, for example, Hurricane Sandy. In 2012, Sandy's landfall caused widespread damage to physical infrastructure, causing outages among Tier-2 and Tier-3 providers in the New York metro region.[104] Those outages made certain internet addresses unroutable—that is, unavailable online—for extended periods of time.[105] Put simply, Sandy knocked New York off the internet. More unexpectedly, Sandy's landfall had cascading effects in far-flung places such as Brazil and Russia, which researchers measured by examining changes in the traffic patterns between other of the internet's constituent networks.[106] If such localized outages give rise to such far-flung and unpredictable effects, imagine what might happen if even larger providers are affected. And that is part of the point. We can only imagine what might happen, since our public visibility into these networks is so limited. Fastly's outage—accidental and short—offers a preview of the possible effects; but a more sophisticated cyberattack or significant disaster could inflict more widespread and longer-lasting troubles.

The opacity of CDNs' private networks prevents the public and its representatives from establishing risk profiles and mitigation strategies for the internet's infrastructure. If Fastly had a partial outage, what would be inaccessible? If Cloudflare had a complete outage, how much damage would it do? Imagine an error at—or, worse, attack on—Cloudflare instead. Given the expansive nature of Cloudflare's scope, if its systems go offline for whatever reason, the scale of the outage would be tremendous. But what *exactly* would be affected? Our limited view into the complex interdependencies among web properties means that we can barely anticipate, let alone prepare for, such a scenario. While these risks begin with technical, engineering, and security failures, the fact that we cannot anticipate and prepare for them is a policy failure. Our limited insight into these providers' inner

103. *See* Gigis, *supra* note 67, at 517 (noting that large CDN providers "typically use IP addresses announced by the hosting network . . . making it impossible to identify the server . . . using traditional techniques such as inspecting BGP feeds").

104. Marguerite Reardon, *Hurricane Sandy Disrupts Wireless and Internet Services*, CNET (Oct. 30, 2012, 10:08 AM), https://www.cnet.com/tech/mobile/hurricane-sandy-disrupts-wireless-and-internet-services [https://perma.cc/8Z8V-GK4Y].

105. *Id.*

106. Geoff Huston, *Superstorm Sandy and the Global Internet*, RIPE LABS (Dec. 3, 2012), https://labs.ripe.net/author/gih/superstorm-sandy-and-the-global-internet [https://perma.cc/8KDQ-BEDA].

workings—their customers, what they do for those customers, and how their physical datacenters connect to one another and to those of other firms'—confounds our ability to establish specific and actionable disaster plans.

B. Central Points of Control

The growing intermediation of the internet's core by CDNs—and the consolidation among CDNs—has not only complicated our public efforts at disaster planning, it has also made the internet more subject to a limited number of *central points of control*.

Recall that one advantage of the shift towards CDNs is a benefit to cybersecurity.[107] CDNs offer collective security to web properties by observing and responding to global traffic trends. But that collective security can come at a cost to individual access. CDNs can—and sometimes do—act as a gatekeeper to internet content. CDNs can filter incoming requests to any of their customers—CNN, NYTimes, gov.uk—in a fine-grained way. Stated otherwise, they may be filtering the content that particular users—identified by, say, their country of origin, or their preferred browser technology—can view.

For example, security measures implemented by CDNs sometimes prevent populations in the Global South from accessing websites. Some automated systems treat traffic from countries like Ghana—or, more precisely, traffic that is *estimated* to originate from countries like Ghana—as presumptively suspicious.[108] CDNs, on behalf of their customers, estimate the risk-reward tradeoff for the traffic they are asked to serve. If it is the case *both* that traffic from Ghana is more likely to be malicious than traffic from the U.S., *and* that traffic from Ghana is less likely to be lucrative than traffic from the U.S., then the CDN will be more—perhaps much more—likely to block Ghanaian than U.S. traffic. While such filtering may satisfy the security objective of the CDN and its customers, it comes at an important cost—namely, restricted internet access—to disfavored populations. An internet that was once imagined offering a life-raft of economic opportunity for developing nations can act, instead, as a moat that excludes them from an ostensibly global market for goods and services.

---

107. *See supra* Part I.

108. Jonas & Burrell, *supra* note 97, at 4.

These security measures also have the effect of limiting the spread of new, potentially useful technologies. Tor, for example, is a privacy-protecting browser.[109] But because it is often used for—and has thus become associated with—illicit purposes, some CDNs have gotten in the habit of regularly challenging Tor-based requests for content, rendering the browser practically unusable.[110] To be sure, there are equities on both sides: On one hand, users may use Tor to access medical websites in order to privately obtain information about a sensitive diagnosis, or to privately seek out abortion-related care in locales where doing so may lead to substantial liability.[111] On the other, Tor's frequent association with bad actors might justify a security concern.

Indeed, this conflict recalls one of the earliest network neutrality controversies, in which Comcast blocked access to BitTorrent, a service which can be used for legitimate purposes, such as a Bible study.[112] But BitTorrent was also frequently associated with copyright infringement.[113] And so the question here—as it was there—regards the allocation of decisionmaking authority. Who should decide which applications and content are allowed: the internet's users, its infrastructural providers, or some other entity altogether? And how does one exercise discipline over the decisions made by infrastructure providers that face little meaningful competition? Even Cloudflare's chief executive, Matthew Prince, raised such questions after banning

109. *History*, TOR, https://www.torproject.org/about/history [https://perma.cc/J4LX-ZJVG].

110. In particular, Cloudflare detects traffic coming from a known Tor endpoint. Associating Tor traffic with fraud and denial of service attacks, it guards the page with a CAPTCHA, attempting to slow down potential attacks. Matthew Prince, *The Trouble with Tor*, CLOUDFLARE BLOG (Mar. 30, 2016), https://blog.cloudflare.com/the-trouble-with-tor [https://perma.cc/Z3HC-44MF]. When used correctly, Tor provides meaningful privacy guarantees. It helps people worldwide evade state censorship. But users find it frustrating to use in practice, in part because Cloudflare makes it difficult to browse the internet with Tor. *See* Kevin Gallagher, Sameer Patil, Brendan Dolan-Gavitt, Damon McCoy & Nasir Memon, *Peeling the Onion's User Experience Layer: Examining Naturalistic Use of the Tor Browser*, *in* ACM SIGSAC CONF. ON COMPUTER AND COMMUNICATIONS SECURITY 1290–1305 (2018).

111. *Cf.* Texas Heartbeat Act, Tex. S.B. 8, 87th Leg., Tex. Health & Safety Code, ch. 171, § 208(2) (2021) (authorizing civil suits against a party that "aids or abets the performance or inducement of an abortion").

112. Formal Complaint of Free Press and Public Knowledge Against Comcast Corporation for Secretly Degrading Peer-to-Peer Applications, Memorandum Opinion and Order, 23 FCC Rcd. 13028, 13042, 13052 n.191 (2008), *rev'd sub. Nom.* Comcast Corp. v. FCC, 600 F.3d 642, 644 (D.C. Cir. 2010).

113. *Id.* at 13030 n.7.

8Chan—a online community strongly associated with neo-Nazi groups and others advocating violent white supremacy.[114] He noted that while he thought it appropriate for Cloudflare to terminate service for platforms that "directly inspire tragic events and are lawless by design," doing so thrusts the company into the "incredibly uncomfortable . . . role of content arbiter," and suggested that public governance structures are better suited to resolving these disputes.[115]

But so far, no such governance structures have emerged—even as such conflicts persist. In late 2022, for example, Cloudflare was once again under pressure to stop providing CDN services to Kiwifarms, a close cousin of 8Chan that had "become notorious for waging online harassment campaigns against [LGBTQIA+] people, women, and others."[116] At first, Cloudflare explained that it would not discontinue service for Kiwifarms, elaborating its "view that cyberattacks not only should not be used for silencing vulnerable groups, but are not the appropriate mechanism for addressing problematic content online."[117] It resolved to continue to provide Kiwifarms with defenses from cyberattacks and other CDN services.[118] But only a few days later, Cloudflare reversed course in what it called an "extraordinary decision" due to "an unprecedented emergency" arising out of increasingly threatening content on Kiwifarms' website.[119] No matter whether one thinks Cloudflare got it right at first, or after its reconsideration, the essential point is that only Cloudflare controlled Kiwifarms' online destiny—notwithstanding its own view that Cloudflare lacks "the political legitimacy to determine generally what is and is not online by restricting security or core internet services."[120]

---

114. *See* Diana Rieger, Anna Sophie Kümpel, Maximilian Wich, Toni Kiening & Georg Groh, *Addressing the Extent and Types of Hate Speech in Fringe Communities: A Case Study of Alt-Right Communities on 8chan, 4chan, and Reddit*, 7 SOC. MEDIA + SOC'Y 1, 1 (2021).

115. Matthew Prince, *Terminating Service for 8Chan*, CLOUDFLARE BLOG (Aug. 4, 2019), https://blog.cloudflare.com/terminating-service-for-8chan [https://perma.c/592E-XBV9].

116. Casey Newton, *How Cloudflare Got Kiwi Farms Wrong*, THE VERGE (Sept. 6, 2022), https://www.theverge.com/2022/9/6/23339889/cloudflare-kiwi-farms-content-moderation-ddos [https://perma.cc/YZ29-YMEL].

117. Matthew Prince & Alissa Starzak, *Cloudflare's Abuse Policies & Approach*, CLOUDFLARE BLOG (Aug. 4, 2022), https://blog.cloudflare.com/cloudflares-abuse-policies-and-approach [https://perma.cc/P5HV-Z6M8].

118. *Id.*

119. Matthew Prince, *Blocking Kiwifarms*, CLOUDFLARE BLOG (Sept. 3, 2022), https://blog.cloudflare.com/kiwifarms-blocked [https://perma.cc/XE8V-48XE].

120. Prince, *supra* note 115.

So too for Tor—when Cloudflare blocks (or effectively blocks) Tor traffic, that's the end of Tor.[121] Because so much of the web relies on Cloudflare, Tor users lack meaningful access to the internet's most popular destinations. That gives rise to a self-fulfilling prophecy, as Cloudflare's decision gives Tor little practical use outside of the so-called "dark web"—and no one but Cloudflare had any meaningful input over the development of this competing browser technology.

CDNs are thus a hidden vector of consolidated power. Their services have positioned them as key players inside the internet, where they function as sites of control. Just as network neutrality, among other policy efforts, is a response to a competition problem, the growing consolidation among CDNs suggests that similar responses may be appropriate. Our network neutrality debates have focused on the edges of the internet, where some ISPs enjoy monopoly status and so may block, throttle, or prioritize traffic with impunity; but policymakers should look inside the internet, too, and consider who should decide what entities can access and traverse the inside of the internet, and how such decisions ought to be made.

IV. WHAT CAN WE DO?

Challenges in scaling the internet's original design led to the emergence of a more centralized structure characterized by CDNs, whose power over the internet has since increased immensely. Indeed, the CDNs' control over this infrastructure not only includes the power to distribute content or help prevent cyberattacks, it also encompasses decisionmaking power over various aspects of the internet—who can use it, and on what terms. Specifically, the CDNs' increasingly private infrastructure confounds the public's ability to reason collectively and strategically about the internet's physical structure. We, as a public, know comparatively little about the interrelated dependencies inside the internet, and so face significant troubles in planning for and addressing outages that cyberattacks, natural disasters, or accidental configuration errors might cause. Similarly, the CDNs' consolidated control over so much internet content grants these infrastructural providers gatekeeper power over the path between users and content providers.

We do not, to be sure, mean to suggest that the answer is a return to the old model of the internet. CDNs help to solve important

121. Prince, *supra* note 110.

problems by providing critical caching and collective defense services. But that CDNs offer an improvement over the prior status quo need not imply that we must accept their shortcomings, or that we should ignore any new problems that these solutions to the old problems introduce. And so we consider policy responses to improve meaningful governance—via both public regulation and market discipline—over these mission-critical providers.

A. Transparency and Security

As noted, public regulators, and the public more generally, lack clarity on both hard infrastructure—that is, cable connectivity patterns within and across providers—as well as aggregate traffic flow patterns—who sends traffic to whom, and to what degree that traffic matters for particular applications.[122] This opacity is the main structural barrier to appreciating systemic risks to the internet's stability and resiliency.

Currently, neither national regulators nor the community of internet measurement and cybersecurity scholars can assess strategic contingencies, due, in large part, to this lack of sufficiently detailed connectivity data. Say, for example, that Fastly has a datacenter in Cheyenne and another in Asheville: Which is more critical to protect or restore? Or if Cloudflare uses certain algorithms to route traffic, how might that logic help public officials decide, in the event of a widespread outage, which internet exchange point to restore first? Stated simply, the public doesn't know enough about how private networks operate internally, how they connect to one another, or how they relate to broader internet to adequately assess strategic contingencies.

One simple and straightforward step is thus to mandate greater disclosure of datacenter connectivity and aggregate statistics about traffic flows between them. If government risk management agencies—for example, the United States Cybersecurity and Infrastructure Agency (CISA)—had such data, they could better prepare for cyberattacks; or they could prepare disaster plans that, for example, prioritize certain web hosts, datacenters, and internet exchange points, to restore service most quickly to the widest population. FEMA, for

122. The amount of reverse engineering the Internet measurement community performs simply to observe proxies of this connectivity is, relative to the centrality of this infrastructure to global trade, commerce, communication, and emergency response, incredible. *See, e.g.*, Gigis et al., *supra* note 67.

example, has detailed maps—roads, topography, and so on—that help it plan for natural disasters. CISA needs network maps, too.

Developing these resources will require input from computer scientists, civil society, and, of course, industry. But disaster planning is—and ought to be—the role of a public agency, such as CISA. Unlike industry, which can make decisions in relative secrecy and by reference to private incentives that need not align with public goals, federal agencies must conduct their work in the open, respond to public input, and account for their decisions to elected leaders and the voting public. Policymakers should charge CISA with the resources needed to carry out this responsibility.

B. Fair Carriage and Gatekeeper Power

As noted, the consolidated nature of the new market for internet traffic exchange gives CDNs the power to control users' access to content. Moreover, CDNs' exercise of this power—intermittent as it may be—is not governed by any public standards. As Matthew Prince, Cloudflare's CEO, noted, it may seem easy, in any one given egregious case, to deplatform an entire service from the internet, but it is much more "hard [to] defin[e] the policy . . . [to] enforce transparently and consistently going forward."[123] It is troublesome that such power over internet access—perhaps the most important modern utility—sits entirely with entities that could render their decisions in the dark, guided by private, rather than public, incentives. Even if the decisions rendered by, say, Cloudflare strike us so far as good, correct, or public-minded, they remain so only for now. We may rightfully wonder how long such private power will be vested in trusted actors and used for purposes with which we may broadly agree.[124] Even Cloudflare, which, like other CDNs, routes and filters internet traffic, has admitted some discomfort with this de facto power over internet speech, explaining that questions about content standards "are real societal issues that need politically legitimate solutions."[125] All the while they rely—

123. Prince, *supra* note 115.

124. Indeed, recent transactions and news reports suggest that some powerful individuals and entities have sought control of other private speech channels to amplify certain voices. *See* Zoe Schiffer & Casey Newton, *Elon Musk's Reach on Twitter is Dropping — He Just Fired a Top Engineer Over It*, THE VERGE (Feb. 9, 2023, 3:25 PM), https://www.theverge.com/2023/2/9/23593099/elon-musk-twitter-fires-engineer-declining-reach-ftc-concerns [https://perma.cc/M2GG-AMSJ] (describing Elon Musk's desire for increased engagement on his personal tweets).

125. Prince, *supra* note 115.

perhaps understandably—on comparatively opaque algorithmic systems to secure and protect network systems.

Future work might formulate a fair carriage rule that prohibits CDNs from discriminating among internet users along such dimensions as national origin or other protected characteristics.[126] It could require that CDNs ensure access to lawful content by means of lawful applications. And it could acknowledge that CDNs perform a beneficial security function and must be given some leeway to protect and secure the internet's constitutive networks. In general terms, such a fair carriage rule would prioritize access to internet speech and content over a CDN's efforts to curate that speech.[127] Vesting the

126. *See* James B. Speta, *Can Common Carrier Principles Control Internet Platform Dominance?*, Marquette University Law School 2022 Robert F. Boden Lecture (Sept. 22, 2022), *in* NW. UNIV. PRITZKER SCH. L. PUB. L. & LEGAL THEORY SERIES, No. 22-29, at 4, https://ssrn.com/abstract=4228208 [https://perma.cc/2TWK-ZK9S] (proposing that common carriage rules would increase diversity and availability of platforms). We note that James Speta's proposal for regulating infrastructural providers, such as Cloudflare, is aimed primarily at resolving questions of competition among user-facing platforms, such as Facebook and Twitter, on the theory that "applying [common carriage] rules to [these infrastructural] support layers could increase the diversity of platforms." By contrast, our proposal for regulating these providers is aimed at resolving competition-related problems *among these providers themselves*.

127. We recognize, of course, that any effort to regulate the control CDNs exercise over internet content will raise First Amendment concerns that echo in the debates over both network neutrality and content moderation. Advocates for network neutrality highlight the consolidation in local markets for internet access, contending that ISPs (such as Comcast) should not have the power to decide which streaming services, say, a user can access (all of them, or perhaps only the Comcast-owned Peacock). *See, e.g.*, Tejas N. Narechania, *Symmetry and (Network) Neutrality*, 119 MICH. L. REV. ONLINE 45, 57–59 (2021). Meanwhile, critics of legislated standards for content moderation contend that the First Amendment guarantees platform providers the discretion to block offensive content, and that policymakers do not fully grasp the impossibility of the problem of moderating at scale. *See, e.g.*, Brief for Respondents at 6–8, Moody v. NetChoice, LLC, —S. Ct.—, 2023 WL 6319654 (No. 22-277) (Oct. 24, 2022) (outlining the discretion used by platform providers in regards to content moderation and arguing for continued freedom to moderate). In our view, the problems presented by CDNs are more closely related to those implicated in the network neutrality debate. Cloudflare's massive capacity, to be sure, presents some difficult problems of moderating at scale—and we do not mean to say that the CDNs have no First Amendment interest in the content that flows over its network. (However, we equally do not concede that they do.) But any of the CDNs' First Amendment concerns must be weighed against the speech interests of users—Ghanaian residents, for example; or those who wish to access sensitive content (information, say, on abortion access in certain states) discreetly. *See, e.g.*, Turner Broad. Sys. V. FCC, 520 U.S. 180, 226 (1997) (Breyer, J., concurring) (noting the speech interests of both the intermediaries exercising editorial control and putative speakers and listeners). Indeed, the CDNs' apparently vast market power likely diminishes the strength of any of First Amendment challenge to new fair carriage rules. *See, e.g.*, *id.* at 181 (describing the relationship between competition and permissible regulation under the First Amendment); USTA v. FCC, 825 F.3d 674, 743–44 (D.C. Cir. 2017) (noting that open internet rules are not barred by the First Amendment).

power to enforce such a rule in a public agency—the FCC, for example—would help to ensure "a certain degree of democratic or quasi-democratic control over infrastructure that undergirds the modern world."[128]

That is not to say that policymakers should not do more to improve competition in the CDN market. Only a few CDNs control the critical paths to the internet's most popular content—and do so in a way that consumers cannot readily avoid. So greater competition or stronger carriage rules can both help to guarantee content access for the internet's users and ensure that CDNs do not leverage their gatekeeper power into adjacent markets.[129] But the barriers to entry are high—CDNs require massive investments in data centers worldwide as well as in sophisticated network engineering and cybersecurity tools. Moreover, as noted above,[130] existing providers benefit from scale economies and network effects, leaving new entrants far behind.

Even though entry into the CDN market is difficult, regulation may help to improve competition among the market's existing players.[131] Regulators might, for example, address switching costs among CDNs, such as (but certainly not limited to) egress fees. This

128. Daniel T. Deacon, *Institutional Considerations for the Regulation of Internet Service Providers*, 74 FED. COMM. L.J. 111, 337 (2022).

129. *See* Tejas N. Narechania, *Network Nepotism and the Market for Content Delivery*, 67 STAN. L. REV. ONLINE 27, 34–35 (2014).

130. *See supra* note 85 and accompanying text.

131. In regulatory comment filed with the Federal Trade Commissions, Cloudflare contended that other cloud providers of other services engaged in anticompetitive conduct that merited close scrutiny. *See, e.g.*, Cloudflare, Comment Letter on Business Practices of Cloud Computing Providers (June 21, 2023), https://www.regulations.gov/comment/FTC-2023-0028-0085 [https://perma.cc/TQ6P-8QF6]. There, Cloudflare explained that these providers charge egress fees, that is, fees when customers attempt to switch providers and take their data from one cloud services provider and to another, or take it in house. *Id.* at 4. Cloudflare's comment explains that such egress fees produce an artificial lock-in effect. *Id.* We agree, as noted *infra* note 132 and accompanying text. But egress is not the only source of such a lock-in effect. Web properties that rely on cloud-based infrastructural services (like CDNs) face special difficulty in switching providers, as even temporary misconfigurations during the transition can result in extended, worldwide downtime for such internet-based businesses. Second, as described above, many web properties are implicitly reliant on CDNs, given the structure of the internet's software supply chains. Even if a business or government manages to extricate themselves from a direct relationship with any one CDN, they may find that some of their embedded services or other providers rely on that very CDN for other purposes. *See supra* Figure 4. Moreover, the data feedback effects that accrue across customers and even within a single customer over time drive lock-in effects for the largest CDNs.

would allow, for example, a provider of reproductive health content to more easily switch from a CDN that blocks private Tor connections to one that allows them.[132] Policymakers might even consider developing a public CDN option—a publicly-run service, definitionally subject to the First Amendment's prohibitions against speech discrimination, that can both discipline other CDNs' terms and rates through competition and give content providers another option.[133] Or policymakers might support the development of technical standards that enable a more competitive CDN market. Specifically, regulators might encourage extensible internet architectures that provision caching and security at the network layer, with the effect of structuring a more competitive, and responsive, market.[134] But, in the meantime, regulators should not merely wait for competition to come to this market. They should instead take action to ensure that internet carriage practices reflect public values and account for broad public concerns.

---

132. You may also be wondering: Can't providers just change CDN providers? In practice, answers to this question depend on whom the CDNs' decisions affect. Providers like Cloudflare regularly block traffic originating in countries like Ghana, treating it as intrinsically suspicious. Jonas & Burrell, *supra* note 97, at 1, 4. How many companies have stopped using Cloudflare in response? Very few, likely because few large tech companies have customers or engineers who notice. Likewise, how many companies have stopped using Cloudflare because they block Tor—itself a serious issue for people trying to circumvent Internet censorship globally? Again, given Cloudflare's persistent dominance, the answer seems to be "very few." Besides, switching reverse proxy providers isn't as easy as you might think. If you've ever used a 'standard' web library like Bootstrap or JQuery on your webpage, you probably used a version hosted on a reverse proxy—probably Cloudflare. Even if you stop using Cloudflare, the libraries you depend on might still. Switching all that stuff over can be a pain at best, and, at worst, could temporarily break your website. The incentives to stick with one's existing provider are high.

133. *Cf.* Yotam Harchol, Dirk Bergemann, Nick Feamster, Eric Friedman, Arvind Krishnamurthy, Aurojit Panda, Sylvia Ratnasamy, Michael Schapira & Scott Shenker, *A Public Option for the Core*, *in* ANN. CONF. OF THE ACM SPECIAL INT. GRP. ON DATA COMMC'N ON THE APPLICATIONS, TECHS., ARCHITECTURES, & PROTOCOLS FOR COMPUT. COMMC'N 377 (2020).

134. Historically, internet researchers have assumed that only intermediaries, be they private or public-sector, can provision the interposing functions that CDNs provide (such as content delivery and DDoS protection). This assumption has been challenged in recent literature. Network design proposals around the so-called "extensible Internet" demonstrate that these interposition functions can be deployed in software at the edge; that is, without intermediating the internet's core. *See generally*, James McCauley, Yotam Harchol, Barath Raghavan, Scott Shenker & Aurojit Panda, *Enabling a Permanent Revolution in Internet Architecture*, *in* ANN. CONF. OF THE ACM SPECIAL INT. GRP. ON DATA COMMC'N ON THE APPLICATIONS, TECHS., ARCHITECTURES, & PROTOCOLS FOR COMPUT. COMMC'N (2019) (proposing a "backwards-compatible architectural framework" designed to create a more "extensible Internet").

CONCLUSION

Although many regard the inside of the internet as robustly competitive—a view that has shaped the regulatory approach to the market for internet traffic exchange—that view is flawed. Perhaps unsurprisingly, the internet's infrastructure has adapted to more modern uses of the internet. A concentrated set of CDNs now intermediate the relationship between the internet's users and its traditional "core." In many respects, this is good. CDNs offer advances in speed, reliability, and security.

But there are tradeoffs. While CDNs offer these advances, they come at the expense of transparency and gatekeeper control. We may want our internet infrastructure to deliver on several promises: access to lawful internet content that is ungated by intermediaries; privacy; and protection from cyberattacks. CDNs can implement these in different ways. CDNs might, for example, use automated processes that inspect internet content before deciding whether to carry it, thereby guaranteeing security and some modicum of privacy. But they do so at the expense of the network neutrality norms that have long governed the internet's core. Or CDNs might require that internet users authenticate themselves before agreeing to carry traffic on equal terms, thereby ensuring network neutrality, at least for authenticated users, and security. But they do so at the cost of privacy.

Resolving such trade-offs is likely to be a core internet governance question in the coming years. Our specific response to these concerns is not to take us back to the old model of the internet—one less secure and less adapted to sorts of applications we are now accustomed to using. Instead, we imagine some policy reforms—new disclosure requirements and fair carriage rules—that directly address these emerging concerns.

Our primary focus for now, however, is on *who* decides on the rules that govern our internet infrastructure rather than on *what* those rules are in their details. At present, our core internet infrastructure is governed by private industry and guided by private incentives. We would much prefer that such rules and decisions come, in the spirit of the internet, by way of our systems of democratic governance and public participation.

APPENDIX

Provider	Market Share
Cloudflare	75.6
Fastly	7.7
Amazon CloudFront	5.9
Akamai	5.3
Sucuri	2.3
DDoS-Guard	1.3
Ezoic	0.9
Imperva	0.9
ArvanCloud	0.4
StackPath	0.4
Variti	0.3
CDNetworks	0.1
Bunny CDN	0.1
Edgio	0.1
GoCache	0.1
QUIC.cloud	0.1
Qrator	0.1
Section	0.1

Appendix Table 1. Whole-web market shares of CDN providers as of February 17, 2023. This table can be regenerated from the code provided in this Article's Supplemental Materials. *See supra* note 47.