NBER WORKING PAPER SERIES

THE IMPACT OF THE GENERAL DATA PROTECTION REGULATION ON INTERNET INTERCONNECTION

Ran Zhuo
Bradley Huffaker
kc claffy
Shane Greenstein

Working Paper 26481
http://www.nber.org/papers/w26481

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2019

The authors are grateful to the Doctoral Office at Harvard Business School for financial support for field work. This research is partially supported by NSF OAC-1724853, NSF C-ACCEL OIA-1937165, U.S. AFRL FA8750-18-2-0049. The views and conclusions do not necessarily represent those of the Air Force Research Laboratory, the U.S. Government, or the National Bureau of Economic Research.

The Impact of the General Data Protection Regulation on Internet Interconnection
Ran Zhuo, Bradley Huffaker, kc claffy, and Shane Greenstein
NBER Working Paper No. 26481
November 2019
JEL No. L00,L51,L86

ABSTRACT

The Internet comprises thousands of independently operated networks, where bilaterally negotiated interconnection agreements determine the flow of data between networks. The European Union's General Data Protection Regulation (GDPR) imposes strict restrictions on processing and sharing of personal data of EU residents. Both contemporary news reports and simple bilateral bargaining theory predict reduction in data usage at the application layer would negatively impact incentives for negotiating interconnection agreements at the internet layer due to reduced bargaining power of European networks and increased bargaining frictions. Considerable empirical evidence at the application layer confirms this prediction. Using a large sample of interconnection agreements between networks around the world in 2015–2019, we empirically investigate the impact of the GDPR on interconnection behavior of network operators in the European Economic Area (EEA) compared to network operators in non-EEA OECD countries. All evidence estimates precisely zero effects across multiple measures: the number of observed agreements per network, the inferred agreement types, and the number of observed IP-address-level interconnection points per agreement. We also find economically small effects of the GDPR on the entry and the observed number of customers of networks. We conclude that the short-run costs for GDPR are concentrated at the application layer.

Ran Zhuo
Department of Economics
Harvard University
Cambridge, MA 02138
USA
rzhuo@g.harvard.edu

Bradley Huffaker
University of California, San Diego
San Diego Supercomputer Center
9500 Gilman Dr. Mail Stop 0505
La Jolla, CA 92093-0505
USA
bradley@caida.org

kc claffy
University of California, San Diego
San Diego Supercomputer Center
9500 Gilman Dr. Mail Stop 0505
La Jolla, CA 92093-0505
United States
kc@caida.org

Shane Greenstein
Technology Operation and Management
Morgan Hall 439
Harvard Business School
Soldiers Field
Boston, MA 02163
and NBER
sgreenstein@hbs.edu

1. Introduction

The Internet comprises thousands of independently owned, managed, and operated networks where networks voluntarily exchange data via bilaterally-negotiated agreements between operators (The Internet Society, 2015). The success of the Internet in creating economic surplus depends on efficient and cost-effective interconnections negotiated by these networks. Hundreds of billions of dollars in transactions depend on the internet's operation in the US alone, and these revenues have been growing rapidly.[1]

This paper investigates whether the General Data Protection Regulation (GDPR) affected investment in the interconnection and growth of the Internet at the internet layer.[2] The GDPR is the European Union (EU)'s stringent regulation to protect consumers' personal data. The GDPR was approved by the EU in April 2016 and came into effect in May 2018. The regulation imposes strict requirements in collecting, processing and transferring personal data of EU residents. The GDPR is a landmark privacy law, inspiring a wave of privacy regulations in countries such as Brazil, India, Japan, and South Korea. The US is also considering federal privacy regulation to harmonize state privacy laws led by California (Goldberg, Johnson & Shriver, 2019). The unprecedented scale and scope of the GDPR to protect consumer privacy make it the most important privacy policy since the commercialization of the internet in the 1990s.

A growing literature measures the impact of regulations' on outcomes at the application

[1]From 2012 to 2017, payments for access to wireline forms of Internet access reached $88.7 billion, growing more than 30% in those five years. Payments for access fees to wireless service reached over $90.0 billion, an increase of 57%. In 2017, online advertising contributed $105.9 billion in revenue to the GDP (Gross Domestic Product) among Internet Publishing and Broadcasting and Web Search Portals. That has grown 250% since 2012. The Census Bureau estimates electronic retailing at over $545 billion for just electronic shopping and mail order houses (NAICS 4541), a growth of 65% over the same period.

[2]We will provide a precise definition of "the internet layer" in Section 2.1.

layer.[3] The pioneering paper in this literature is Goldfarb & Tucker (2011b), which examined how the EU's 2002 e-Privacy Directive reduced online display ad effectiveness in the EU relative to other countries. Our study arrives against a backdrop of growing literature assessing the impacts of privacy regulation such as the GDPR. To date there is little research on the impact of these prior policies' at the transport or internet layer of the Internet. Our paper adds to the body of research on the impact of online privacy regulation along this important margin. Specifically, we ask: How does the GDPR affect investment in growing and interconnecting the internet layer of the Internet?

Based on all studies to date, and on all available news reports, we expect GDPR to reduce investment. Most studies to date find significant costs associated with implementation of the GDPR. Goldberg, Johnson & Shriver (2019) found a large and significant 10% decline of recorded page views, visits, orders and revenue for a set of EU firms using Adobe Marketing Cloud after the implementation of the GDPR. Johnson & Shriver (2019) found the week after the GDPR's enforcement, website use of web technology vendors fell by 15%. They also found websites were more likely to drop smaller vendors, which increased the relative concentration of the vendor market by 17%. Jia, Jin &Wagman (2018, 2019) show the implementation of the GDPR strongly reduced venture capital investment in technology start-ups in Europe compared to their US counterparts and far-away investors were more likely to respond negatively. Moreover, the GDPR has led to use of fewer third party cookies (Libert, Graves, & Nielsen 2018) and changes in online privacy policies (Degeling et al. 2018 and Mohan et al. 2019). Godinho de Matos & Adjerid (2019) studied the effectiveness of a campaign for obtaining GDPR-compliant

[3]We will provide a precise definition of "the application layer" in Section 2.1.

consent for personal marketing, and, again, found a costly effect. These studies follow considerable publicity for GDPR, which also forecast large implementation costs which we summarize inside the body of this paper. An exception is Iordanou et al. (2018), which found few changes in the amount of data flow associated with web tracking and in the percentage of this data flow attributed to tracking servers hosted in EU around the GDPR implementation window.

Theoretical considerations also lead to an expectation that GDPR will reduce investment. After reviewing the setting and background, we first formalize the common intuition that the demand for and usage of data at the application layer alters investment incentives to interconnect at the internet layer. To do this we develop a simple theoretical model of bilateral bargaining between network operators. This model is largely drawn from the parsimonious model presented in Besen et al. (2001).

We then dive into the empirical analysis. Our descriptives show persistent and similar growth in internet interconnection of EEA (European Economic Area) countries versus non-EEA OECD (Organization of Economic Cooperation and Development) countries, though the levels of interconnectedness differ (see Appendix Table A1 for precise definitions). We treat the GDPR's April 2016 approval and May 2018 enforcement as two cutoff dates for periods of post policy treatment. We hypothesize that, if there is any effect of the GDPR at the internet layer, it occurs after the policy approval date and especially after the policy effective date. We offer several reasons for this assumption. We expect the effect of the GDPR at the internet layer to occur as network operators respond with changes in bargaining outcomes due to changes in actual data flows. Moreover, recent empirical work offers evidence that the effects at the application layer occurred immediately after the stark enforcement date in May 2018 (Goldberg, Johnson & Shriver, 2019 and Jia, Jin, & Wagman, 2018, 2019). Though the discussion of a new data protection regulatory framework began in 2012, there is a lot of evidence suggesting organizations operating at the application layer did not respond to the policy change prior to its approval. Rather, within the two years between the GDPR's approval and effective dates, many organizations choose to roll out their compliance strategy only days and weeks before the effective date (Jia, Jin, & Wagman, 2018). This behavior justifies our use of stark cutoff dates for treatment.

For the difference-in-difference approach, we contrast interconnection activities by networks owned by organizations headquartered in the EEA and networks owned by organizations headquartered in other countries before and after the approval and implementation of the policy. This approach is in line with prior literature (Goldfarb &Tucker, 2011b, 2012, Jia, Jin, & Wagman, 2019). Our treatment group consists of networks owned by organizations headquartered in EEA countries, as the GDPR predominantly affects data processing organizations located within the EEA, and the policy explicitly restricts transfer of personal data outside of the EEA. The control group is chosen to be networks owned by organizations headquartered in non-EEA OECD countries. This is a relevant comparison because networks in developed countries have similar growth rates of interconnection prior to the GDPR. As we will show in a series of graphs later in the result section, the parallel pre-trends needed for the difference-in-difference approach are visually apparent for the treatment and control groups across all outcome variables of interest. Networks in non-EEA non-OECD countries and territories are excluded from the control group for worries that networks in developing countries might behave differently from networks in more developed countries prior to the GDPR. We note that the GDPR may affect networks outside of the EEA, therefore it is appropriate to interpret our results as the differences between the impact of the GDPR on EEA networks relative to non-EEA OECD networks.

Our data comes from various data sources collected by the Center of Applied Internet Data Analysis (CAIDA) at the University of California, San Diego, and represents the state-of-the-art in inferring the presence of interconnection agreements and their types between networks on the world-wide scale, based on large collections of raw data on global network and IP address level topology of the Internet. Our data includes ownership information of all operating networks around the world, the number of observed agreements per network and the inferred type of each agreement. Using this network level data we can estimate the number of networks that are customers to a given networks. By combing the topology we can infer the number of IP-address-level connection points between pairs of networks with interconnection agreements. The datasets used in this paper are collected quarterly, monthly or even daily. Most of the datasets go as far back as to early 2000s and are publicly accessible through CAIDA's website at http://www.caida.org/data/overview/.[4]

Using this data, we are able to infer changes in European networks' interconnection behavior before and after April 2016 and May 2018 relative to non-European OECD networks'. We estimate precise zero effects across multiple measures. Networks in the EEA are similar to networks in non-EEA OECD countries in terms of the growth in the number of interconnecting parties and types of agreements reached. The observed numbers of IP-address-level interconnection points between each pair of interconnecting networks are also unaffected by whether an EEA network is involved. We also find economically small effects of the GDPR on the

[4]For more information about the data sources used in this paper, please see the Data Appendix.

entry and the number of networks that are customers of networks in EEA countries relative to non-EEA OECD countries. Overall, we discover no discernible change in European networks' interconnecting behavior, rejecting the theory that the negative impact of the GDPR at the application layer has thus far propagated to the internet layer. The European networks are growing in numbers and interconnectedness as rapidly as their non-European OECD counterparts. An obvious policy implication of our paper is that even stringent internet privacy regulation that has strong negative impact at the application layer does not impact the short-run growth of the Internet infrastructure and the incentive of network operators to interconnect the Internet. In the last part of the paper we discuss a number of possible reasons for this result.

An additional contribution of our paper is the data. While there is a theoretical literature on network operators and interconnection agreements (see for examples, Besen et al. (2001), Choi, Jeon, and Kim (2015) and Laffont et al. (2001)), empirical research in Economics has been scant. To the best of our knowledge, the type of data used in our paper has only been used once in prior Economic literature, where D'Ignazio and Giovannetti (2008) obtained data from the London Internet Exchange (LINX) of its member networks and one type of agreement (peer-to-peer) between the members. Our data is a significant improvement from their data as it covers virtually all operating networks in the world, a large number of agreements of both peer-to-peer and provider-to-customer types, and is publicly accessible. Such data may be valuable for a range of economic- and policy-relevant research. Across the academic and policy arena, the lack of well-measured data describing the interconnectivity and traffic flow in the Internet has brought great attention, especially in issues such as net neutrality, international trade in digitally delivered goods, and market power of big technology firms whose data flows dominate global internet traffic (see discussions in Weller and Woodcock (2013), US International Trade Commission (2014), Meltzer (2014), Nicholson and Giulia (2016) for a few examples). Our data may represent a small step towards filling the data gap in growing needs to analyze internet-related economic and policy issues.

The rest of the paper is organized as follows. Section 2 provides the background in network interconnection and the GDPR. Section 3 provides the theoretical underpinning of our empirical approach by presenting a simple bilateral bargaining model between network operators. The section also surveys additional theoretical literature on internet interconnection in Economics and Computer Science. Section 4 describes the data. Section 5 presents results across a number of measures of the impact of the GDPR on interconnection. Section 6 concludes.

2. Background

2.1 Internet Interconnection

The Internet was designed with four layers of data exchange in mind: application, transport, internet, and link.[5] Each layer uses a specific set of protocols, shared state, and provides a connection for higher layers. Processes in each layer communicate both with the layer directly above and below, but also across the same layer through connections provided by lower layers. Figure 1 provides a visual illustration of the four layers and how data is exchanged between and across each layer. Application-layer software maintains state specific to each application, such as Skype or a web browser, which communicates to other applications hosted on other devices

[5] In an official specification document for the Internet regarding requirements for Internet hosts, the Internet Engineering Task Force (RFC1122, 1989) describes the four layers and specifies protocols associated with each layer.

across the global Internet. This is the layer where personal data is most relevant. Application layer communication relies on lower layers of internet infrastructure and communication protocols. The transport layer is responsible for making sure data from applications arrives correctly and reliably between end point devices. Protocols at this layer break data into packets before handing them off to the internet (network, or IP) layer. The internet layer is responsible for maintaining global routing state, and routing data packets to their destination address by selecting the next closest router. Below this layer, the link layer is responsible for forwarding the packets to immediately adjacent (the "next hop") routers.

At the internet layer, the Internet can be conceptualized as a collection of different networks, each with its own set of routers and routing policies. In order to reach other networks, individual networks make direct connections with each other, as well as indirect connections through other networks that transport data traffic on their behalf. As described by the Internet Society (The Internet Society, 2015), these networks are typically classified into one or more of the following roles:

- Internet Service Providers (ISPs) own or resell services, i.e., access to networked facilities, that bring Internet access to residential and business end users. ISP end users both consume and generate Internet traffic.

- Governments, private companies, and universities often operate their own networks and interconnect with other networks to achieve global reachability on the Internet.

- Content providers are producers and distributors of Internet content. Examples include web-hosting companies and social media networks. In recent years, many content providers have built their own distribution networks or have chosen to buy services from content delivery networks that specialize in distributing content to end users.

- Content delivery networks (CDNs) aim to efficiently and reliably distribute content on behalf of their primary customers: content providers. CDNs aim to place their content as close to the end user as technically and commercially possible. CDNs often have servers in many data centers around the globe to make it easier to interconnect with ISPs close to end users.

- Regional/global transit providers. They provide access to the global Internet to their customer networks, allowing customers to access distant networks.

Network operators typically have interconnection relationships with many different players and use a mix of agreements. As described by the Internet Society (The Internet Society, 2015), these agreements can be broadly understood to be one of two types:

- Provider-to-customer (p2c) or customer-to-provider (c2p) is an agreement by which the provider network agrees to provide its customers with connectivity to the rest of the Internet for a fee.

- Peer-to-peer (p2p) is an agreement by which two networks agree to a mutual exchange of traffic to and from their customer networks. Peering arrangements reduce the amount of traffic a network must send through its upstream transit provider, lowering the average cost of traffic delivery. If the peers have similar negotiating power, they form a settlement-free agreement. If there is an imbalance, the weaker network pays the other under a paid peering agreement.

p2c and c2p agreements are typically priced by the transit provider as a metered service outside of the residential market on a per-megabit-per-second (Mbps) basis. The market for providing transit is very competitive and prices have been on a strong declining trend from 1998 until present. Estimates based on a sample of US transit providers show that per Mbps transit prices averaged $12.00 in 2008 and averaged $0.63 in 2015 and yearly decreases between 2008 and 2015 ranged from 28% to 52% (Norton, 2014, Table 2-2). Transit providers may also provide pricing discounts for pre-committing to certain volumes of traffic. The duration of p2c and c2p agreements can be as short as one month or as long as multiple years. Due to the strong declining trend in prices, even multiple-year agreements are usually renegotiated yearly. Other than the potential legal cost of breaking a multiple-year agreement, the cost for customers to switch to a different provider is low relative to the potential gains of bringing the unit transit prices in line with the current market price. Another common practice in the transit market is to use extremely short-term agreements, typically with no volume commits and with a duration of just one month, to fully capture the ever-decreasing market prices for transit (Norton , 2014).

Though the prices of p2c agreements have declined rapidly, p2p agreements may reduce the cost of traffic exchange even further when the volume of traffic is high. p2p agreements are typically negotiated on a case-by-case basis between potential peers. Traffic volume is often "a key determinant" whether a peering agreement results from the negotiation as "the decision hinges upon whether or not there is sufficient value from peering to justify spending time and money" (Norton, 2014). A portion of the cost of peering involves purchasing circuits of fixed capacities between the peers at the peering point and this cost scales with the capacities of the circuits. When a network does not have a Point of Presence (POP) at the agreed peering location, additional cost is incurred to bring its traffic to the peering point. Networks incur additional cost associated with colocation, equipment, and peering ports. The split of the cost is specific to the agreement and net payment between networks may occur, resulting in a paid peering agreement. The cost of peering can also vary significantly by the peering location and the geographic proximity of the two networks to the peering location. A very rough estimate of the total cost of a p2p interconnection with a 10Gbps capacity at a European peering point using cross-continent transport stands at $11,000 per month in 2014 (Norton, 2014, Table 5-1).

When an interconnection agreement is formed, the two parties decide the technical aspects of interconnection. In many cases, setting up an interconnection does not require the deployment of additional hardware (Norton, 2014). The two parties may simply utilize existing assets, such as configuring an existing port or purchasing circuits between their existing POPs. The process to interconnect can take as little as minutes. When the physical assets for interconnection, such as optical fibers and undersea cables, are not present, it can take substantially longer to install the hardware to interconnect, often in years. As we will discuss in more detail in later sections, our empirical analysis contrasts interconnection activities of networks in the EEA versus networks in non-EEA OECD countries and the vast majority of interconnections that we study involve both parties in developed countries. We expect the availability of physical hardware to have little constraint on the incentives to interconnect, at least in the short run that we study, and therefore interconnections can be established or terminated reasonably quickly in response to policy changes.

In summary, the internet is comprised of many agreements between different organizations – ISPs, transit providers, content providers, and CDNs – all oriented towards exchanging data. As described below, basic economic reasoning predicts that GDPR will lead to fewer data flows at the application layer, and, therefore, fewer negotiated interconnection agreements at the internet layer. In our statistics we will not distinguish between different organizations, nor will we focus on the motives for making the agreements other than to focus on how GDPR changed behavior at the internet layer, as described in the text below. For some analysis purposes we will focus on the differences between p2c, c2p, and p2p, which we will describe in the Section 4.

2.2 The GDPR

The GDPR was approved in April 2016 and became effective in May 2018. The regulation imposes strict restrictions on how personal data of EU residents should be collected, processed and transferred. Individuals receive the right to access, correct, erase, and port their personal data elsewhere as well as the rights to object to data processing and decisions based on automated processing. The GDPR also expands the definition of personal data beyond personally-identifiable data to include individual-level data like cookies and IP addresses (Goldberg, Johnson & Shriver, 2019).

Under the GDPR, organizations must obtain explicit opt-in consent from users to collect personal data and are required to request consent in an intelligible and easily accessible form, using clear and plain language.[6] Organizations must also encrypt and anonymize personal data (data protection by design), minimize data collection (data protection by default), and allow individuals to exercise their rights in an easy and timely manner (Goldberg, Johnson & Shriver, 2019). Moreover, organizations must appoint a Data Protection Officer to oversee compliance

[6]EU GPDR.org. "GDPR Key Changes." https://eugdpr.org/the-regulation/. (Accessed September 23, 2019).

activities and audit internal data processes. In the event of a data breach, organizations must promptly notify the regulator and affected individuals. Organizations in breach of the GDPR can be fined up to 4% of annual global turnover or 20 million Euros, whichever is greater. These obligations impose potentially large compliance and opportunity costs on firms and may impact online activities of both firms and users (Goldberg, Johnson & Shriver, 2019).

It is important to note that the GDPR specifies how data should be processed and shared at the application layer but imposes no explicit restriction of data exchange at the internet layer. For example, if personal data is electronically routed through a network located in a non-EEA country at the internet layer but the transfer of personal information is actually from one organization located in an EEA country to another organization also located in an EEA country at the application layer, it is not a restricted transfer. Though there is no directive that requires changes at the internet layer, indirect effects could matter. The indirect effects may be driven by requirements for compliance of the regulation at the application layer. The GDPR primarily applies to controllers and processors located in the EEA, and individuals risk losing the protection of the GDPR if their personal data is transferred outside the EEA. On that basis, the GDPR restricts transfers of personal data from organizations within the EEA to organizations outside the EEA at the application layer.[7] When data transfers between organizations within and outside the EEA occur at the application layer, data needs to be routed through interconnections between networks within and outside the EEA at the internet layer. Therefore one may expect data flow at the internet layer between EEA countries and other countries to decrease, driven by decline in data transfers initiated by firms that collect and process personal data at the application layer. In

[7]Information Commissioner's Office, United Kingdom. 2019. "Guide to the General Data Protection Regulation (GDPR)." https://ico.org.uk/for-organisations/guide-to-data-protection/.

addition, one may also expect building of more storage facilities within Europe corresponding to the shift from international transfer and storage of data to local storage of data.

A broad range of behaviors also could drive indirect effects. The discussion of the GDPR in the popular media, and in numerous public "White Papers" paints an alarmist and negative picture of the broad impact of the regulation due to reduction in data usage and increased frictions. A sampling from (the most credible) news sources gives a good sense of the range of concerns voiced at the time GDPR became binding. Every editorial regarded the GDPR as important and impactful, with the most sweeping editorials declaring "...the web will never be the same..."[8] and the internet had reached the end of "industry self-regulation."[9] Some editorials bemoan that European regulators had become the globe's watchdog,[10] while others foresaw their action as steps towards splintering the internet into three regimes.[11] Some editorials express concerns about confusion and uncertainty,[12] and costly adjustments for business,[13] while other editorials forecast that large firms would benefit at the expense of the small.[14] Several news articles focus on extreme reactions from businesses, such as "...a number of businesses have decided to deal

[8]Petzinger, Jill & Jason Karaian. 2018. "What It's Like to Use the Web in Europe after the Arrival of GDPR." Quartz. https://qz.com/1289152/gdpr-has-changed-what-its-like-to-use-the-web-in-europe/.

[9]Downes, Larry. 2018. "GDPR and the End of the Internet's Grand Bargain." Harvard Business Review. https://hbr.org/2018/04/gdpr-and-the-end-of-the-internets-grand-bargain.

[10]Satariano, Adam. 2018. "G.D.P.R., a New Privacy Law, Makes Europe World's Leading Tech Watchdog." New York Times. https://www.nytimes.com/2018/05/24/technology/europe-gdpr-privacy.html.

[11]The Editorial Board. 2018. "There May Soon Be Three Internets. America?s Won't Necessarily Be the Best." New York Times. https://www.nytimes.com/2018/10/15/opinion/internet-google-china-balkanization.html.

[12]Cool, Alison. 2018. "Europe's Data Protection Law Is a Big, Confusing Mess." New York Times. https://www.nytimes.com/2018/05/15/opinion/gdpr-europe-data-protection.html.

[13]Trentmann, Nina. 2018. "Companies Worry That Spending on GDPR May Not Be Over." Wall Street Journal. https://www.wsj.com/articles/companies-worry-that-spending-on-gdpr-may-not-be-over-1527236586.

[14]Davies, Jessica. 2018. "The Impact of GDPR, in 5 Charts." Digiday. https://digiday.com/media/impact-gdpr-5-charts/.

Bershidsky, Leonid. 2018. "Europe's Privacy Rules Are Having Unintended Consequences." Bloomberg. https://www.bloomberg.com/opinion/articles/2018-11-14/facebook-and-google-aren-t-hurt-by-gdpr-but-smaller-firms-are.

Kostov, Nick & Sam Schechner. 2019. "GDPR Has Been a Boon for Google and Facebook." Wall Street Journal. https://www.wsj.com/articles/gdpr-has-been-a-boon-for-google-and-facebook-11560789219.

with GDPR by getting rid of European services altogether,"[15] or not make their sites available to European readers,[16] or move physical facilities outside of the jurisdiction of European regulators.[17] As of this writing, these views continue to be the consensus. In extensive online search of news articles and editorials since the implementation of the GDPR, we have found no opinion or report to suggest any other impact on business than a costly impact.

This news coverage would lead one to expect significant indirect effects of the costs at the application layer on derived demand for the internet layer. The effects could take many forms, such as a decline in internet-related investment, fewer firms, less growth in applications, less growth in traffic, and fewer partnerships between European and non-European firms. One may also expect confusion and uncertainty to create delays in investment within Europe. Regardless of the mechanism, all the forecasts point in the same direction.

Notice also the potential flaw of such predictions. The popular media makes broad speculations about the negative impacts of the policy, but only a small sample of experiences provides support for the speculations. Neither systematic data collection, nor a census of experience across a range of circumstances, informs the conclusions, and much of it stresses the costs in unspecific terms. Sound empirical work is needed to support or refute such uninformed speculation.

[15] Burgess, Matt. 2018. "Help, My Lightbulbs are Dead! How GDPR Became Bigger than Beyonce." Wired. https://www.wired.co.uk/article/happy-gdpr-day-gdpr-hall-of-shame.

[16] O'Connor, Joseph. 2018-2019. "Websites not Available in the European Union after GDPR." https://data.verifiedjoseph.com/dataset/websites-not-available-eu-gdpr.

[17] Hern, Alex. 2018. "Facebook Moves 1.5bn Users out of Reach of New European Privacy Law." Guardian. https://www.theguardian.com/technology/2018/apr/19/facebook-moves-15bn-users-out-of-reach-of-new-european-privacy-law.

3. Theory

In this section, we formalize the common intuition that the demand for and usage of data at the application layer alters investment incentives to interconnect at the internet layer with a simple theoretical model of bilateral bargaining between network operators. The model largely draws from Besen et al. (2001). Though this model abstracts away many issues, such as interdependence of interconnection decisions, customers' choices of networks and the rich set of considerations different types of networks have in making interconnection decisions, it is parsimonious and delivers neat analytical solutions of the bargaining outcome and the amount of transfers. At the end of this section, we also briefly survey additional theoretical literature in the Economics and Computer Science literature on modeling interconnection decisions.

First let there be two network operators O_1 and O_2. The two networks decide whether to interconnect. Let mass M_1,M_2 account for the combined value of each network's content and users, and the value of its customers not reachable through the other network. So M_1 is the value reachable through O_1 or O_1's customers and not reachable through O_2 or O_2's customers. Let I_1, I_2 be the combined value of all content and users on the Internet not reachable through the other network or its customers. M_1 is a subset of I_1 and M_2 is a subset of I_2. When network i is a large transit provider, I_i would be equal to all content and users on the Internet minus M_j. Examples of value are a content provider's video content, and the ISP's video subscribers.

For transit providers and content delivery networks, customers would be other networks depending on them to connect to other parts of the Internet. For governments, private companies and universities, their customers are just themselves. Assume that from O_1's perspective, forming a peer-to-peer interconnection with O_2 would allow O_1 to reach mass M_2 more efficiently. O_1 can in term generate revenue from its customers due to improved service. Assume also that O_2 would reach mass M_1 more efficiently under the peer-to-peer agreement. If O_1 is the provider, forming a provider-to-customer link with O_2 would allow O_1 to reach M_2 more efficiently. While from O_2's perspective, forming a customer-to-provider link with O_1 would allow O_2 to access I_1.[18]

Let $f(m)$ be function of revenue collected by the network operator O_i per unit mass of its customers, where m represents the mass of customers in the internet O_i is able to reach in a reliable and efficient manner for its customers. For networks whose customers are themselves, we can think of f as the benefit of connecting their networks to the Internet in monetary values. Assume f is an increasing function and is concave. Let $C_{p2p}(M_1 + M_2)$ be the cost of a p2p interconnection and $C_{p2c}(I_1 + M_2)$ be the cost of a p2c interconnection between O_1 and O_2, which are increasing functions in the total masses that depend on the interconnection. C is concave, evident from decreasing per Mbps interconnection fees in this industry. Moreover $\frac{dC_{p2p}(m)}{dm}|_{m=\tilde{m}} < \frac{dC_{p2c}(m)}{dm}|_{m=\tilde{m}}$ for all m, reflecting the fact that p2p agreements have more rapidly declining per Mbps cost than p2c agreements and significantly reduce cost of interconnection especially when m is large. Let τ be any additional cost associated with negotiating an agreement.

Assume any disruption to data exchange between O_1 and O_2 is only sustained during bargaining[19] and customers do not change their networks during bargaining or in response to the

[18]Note M_1, M_2, I_1 and I_2 are specific to the negotiation between O_1 and O_2. If O_1 and O_2 form a p2c agreement where O_1 is the provider, under a negotiation between O_1 and another network O_3, M_2 becomes part of O_1's combined value of content and users M_1'.

[19] In the event of no agreement between O_1 and O_2, customers in M_1 and M_2 experience less efficient service in reaching I_2 and I_1. In practice, data usually takes a longer and inefficient path through a series other networks between O_1 and O_2.

bargaining outcome. We also hold fixed the interconnection agreements between either of O_1, O_2 and all other networks. Assume these other agreements allow each network to access mass G_1 and G_2. O_1 and O_2 can either form an agreement with one of the three agreement types: (a) a p2c agreement where O_1 is a provider to O_2, (b) a p2p agreement, (c) a c2p agreement where O_1 is a customer to O_2, or take the outside option (d) no agreement. In practice, as the relative masses and bargaining power of the two networks strongly influence the type of agreement formed[20], we first assume networks compare one of (a), (b), (c) with the outside option (d), rather than comparing all of the four options simultaneously, and derive comparative statics. We then discuss potential substitutions between agreement types.

3.1 Peer-to-Peer Agreements (p2p)

When the two networks have relatively similar masses and bargaining power, they consider either a p2p agreement or no agreement. The bargaining outcome according to the noncooperative bargaining theory with short times between offers is approximately the same as that of the Nash bargaining model, provided the payoff each earns during the period of disruption is treated as the Nash threat point (Binmore, Rubinstein, & Wolinsky (1986), Besen et al. (2001)). The total surplus to be divided when O_1 and O_2 reach an agreement is $M_1 f(G_1+M_2)+M_2 f(G_2+M_1)-C_{p2p}(M_1+M_2)-\tau$, while the threat point is $(M_1 f(G_1), M_2 f(G_2))$. We further assume when the mass M_i of a network O_i increases, this change has a higher impact on O_j's threat point value than on the cost of interconnection, that is $\frac{d(M_j f(G_j+M_i))}{dM_i}|_{M_i=\tilde{M}} > \frac{dC_{p2p}(M_i+M_j)}{dM_i}|_{M_i=\tilde{M}}$ for

[20]For reference, if we measure M_1, M_2 and I_1 purely in terms of the number of IP addresses and let O_2 be the smaller network in an agreement, the average ratio of M_2 to M_1 is 0.81 for p2p agreements, while the average ratio of M_2 to the full routed IP address space (which is close to I_1, given the relatively small size of M_2) is 0.00016 for p2c agreements.

all M_i, M_j.

O_1 and O_2 would decide to interconnect if the gains from agreement

$$g = M_1 f(G_1 + M_2) + M_2 f(G_2 + M_1) - C_{p2p}(M_1 + M_2) - \tau - M_1 f(G_1) - M_2 f(G_2) \geq 0. \quad (1)$$

At a noncooperative bargaining outcome, the two networks divide equally any gains relative to the threat point, so the resulting bargaining payoff for network O_1 is

$$\pi_1 = \frac{1}{2}[M_1 f(G_1 + M_2) + M_2 f(G_2 + M_1) - C_{p2p}(M_1 + M_2) - \tau + M_1 f(G_1) - M_2 f(G_2)]. \quad (2)$$

and for network O_2 is

$$\pi_2 = \frac{1}{2}[M_1 f(G_1 + M_2) + M_2 f(G_2 + M_1) - C_{p2p}(M_1 + M_2) - \tau - M_1 f(G_1) + M_2 f(G_2)]. \quad (3)$$

With interconnection, O_1 would be able to earn a revenue of $M_1 f(G_1 + M_2)$ from M_1 and needs to share half the cost of the interconnection $\frac{1}{2}[C_{p2p}(M_1 + M_2) + \tau]$. Let O_1's profit be $\rho_1 = M_1 f(G_1 + M_2) - \frac{1}{2}[C_{p2p}(M_1 + M_2) + \tau]$, then the excess

$$\pi_1 - \rho_1 = \frac{1}{2} M_2 [f(G_2 + M_1) - f(G_2)] - \frac{1}{2} M_1 [f(G_1 + M_2) - f(G_1)]. \quad (4)$$

is the negotiated net payment from O_2 to O_1. Define $h(M) = \frac{[f(G+M) - f(G)]}{M}$, then O_1 receives a positive payment from O_2 if and only if $h(M_1) - h(M_2) > 0$. In such a case, O_1 and O_2 are in a paid peering agreement. When $h(M_1) - h(M_2) = 0$, the two networks are in a settlement-free peering agreement.

Now suppose O_1 is a network serving customers in the EU while O_2 is some other network outside the EU that connected with O_1 before the GDPR was implemented. We can work out the comparative statics for changes in bargaining outcomes following changes in model parameters due to the GDPR. We consider two different changes in model parameters: (a) a decrease in M_1, and (b) an increase in τ. Goldberg, Johnson & Shriver (2019) shows large and significant 10% decline in recorded page views, visits, orders and revenue of EU customers after the implementation of the GDPR. Jia, Jin, & Wagman (2018) show decline in venture capital investment in technology start-ups, particularly in the total amounts raised across funding deals, the number of deals, and the amount raised per individual deal. The effects are especially pronounced for newer and data-related ventures. Both papers provide some evidence of decline in the mass of EU customers, both in terms of the number of users and the amount of content supplied to the rest of the Internet. This change is represented by a decrease in M_1 in our model. As the new legislation rolled out, it creates uncertainty in the business environment and additional burden in making sure both interconnecting parties and their customers are GDPR-compliant, increasing bargaining frictions. We represent this change by an increase in τ in our model.

Taking the derivative of the gains from agreement with respect to M_1, we have

$$\frac{dg}{dM_1} = f(G_1 + M_2) - \frac{dC_{p2p}(M_1 + M_2)}{dM_1} - f(G_1) + M_2 \frac{df(G_2 + M_1)}{dM_1} > 0. \tag{5}$$

It is also easy to show $\frac{d[h(M_1) - h(M_2)]}{dM_1} > 0$. Together, these derivatives imply two changes when

M_1 decreases: (1) Gains from agreement fall. If the gains fall below zero, the agreement between O_1 and O_2 breaks. (2) O_1 receives a reduced amount of transfer from O_2, though we do not observe transfers in our data. Using similar derivations, an increase in τ would also imply higher chance of termination of the interconnection agreement, though it does not have an effect on the transfers.

3.2 Provider-to-Customer Agreements (p2c)

When O_1 has substantially more mass and bargaining power than O_2, the networks consider either a p2c agreement where O_1 is the provider or no agreement. Using the same set of assumptions as above for the p2p agreements, O_1 and O_2 would decide to interconnect if the gains from agreement

$$g = M_1 f(G_1 + M_2) + M_2 f(G_2 + I_1) - C_{p2c}(I_1 + M_2) - \tau - M_1 f(G_1) - M_2 f(G_2) \geq 0. \quad (6)$$

The negotiated net payment from O_2 to O_1 is

$$\pi_1 - \rho_1 = \frac{1}{2} M_2 [f(G_2 + I_1) - f(G_2)] - \frac{1}{2} M_1 [f(G_1 + M_2) - f(G_1)] > 0. \quad (7)$$

Suppose O_1 is a transit provider in EU and a significant portion of I_1 are EU users and content. The GDPR might result in a decrease in I_1. Taking the derivatives of Equations 6 and 7 with respect to I_1, we derive two changes when I_1 decreases: (1) Gains from agreement fall. If the gains fall below zero, the agreement between O_1 and O_2 breaks. (2) O_1 receives a reduced amount of transfer from O_2, though we do not observe transfers in our data.

If we instead suppose O_2 is a EU network seeking access to I_1 and the GDPR decreases M_2, we take the derivatives of Equations 6 and 7 with respect to M_2 and derive two changes: (1) Gains from agreement fall. If the gains fall below zero, the agreement between O_1 and O_2 breaks. (2) O_1 receives an increased amount of transfer from O_2, though we do not observe transfers in our data.

3.3 Substitution between Agreement Types

When O_1 has a larger mass than O_2, it is possible the two networks decide between a p2c agreement where O_1 is the provider and a paid p2p agreement. The two networks would enter a paid p2p agreement if Equation 1 holds and the gains from a p2p agreement are greater than the gains from a p2c agreement

$$g_\Delta = M_2[f(G_2+M_1) - f(G_2+I_1)] + C_{p2c}(I_1+M_2) - C_{p2p}(M_1+M_2) \geq 0 \tag{8}$$

Suppose O_1 is a EU network and the GDPR negatively impacts both M_1 and I_1. $\frac{dg_\Delta}{dM_1} > 0$, implying that a decrease in M_1, holding all else fixed, would make it more likely for the two networks to enter a p2c agreement. However, $\frac{dg_\Delta}{dI_1} < 0$, implying that a decrease in I_1, holding all else fixed, would make it more likely for the two networks to enter a p2p agreement. The overall effect is unclear and depends on the relative changes to the masses M_1 and I_1 offered, their prices and O_2's revenue function.[21] Suppose instead O_2 is a EU network and the GDPR

[21] An intuitive way to understand this situation is to use the second-degree price discrimination framework. One can view the p2c agreement as the product with a larger quantity and a higher price and the p2p agreement as the product with a smaller quantity and a lower price. The choice between the two products depends on the consumer's preferences as well as the structure of non-linear pricing.

negatively impacts M_2. Taking the derivative, $\frac{dg_\Delta}{dM_2}$ can either be positive or negative, depending on the cost functions, O_2's revenue function, and the relative masses M_1 and I_1.

In summary, this simple model formalizes the intuition that negative impacts of the GDPR on the application layer negatively impact European networks' bargaining positions. European networks would have fewer agreements of all three types and receive a reduced amount of transfers. The effect of the GDPR on the potential substitutions between agreement types is unclear. Additional theoretical literature in Economics have modeled other aspects of interconnection decisions. Laffont et al. (2001) analyze competitiveness of pricing of interconnection. Choi, Jeon, & Kim (2015) analyze the effect of second-degree price discrimination when quality of interconnection is heterogenous. Badasyan & Chakrabarti (2008) model ISPs' choices between peering and transit agreements. There is also a considerable theoretical literature in Computer Science and Information Systems that models complex interactions between network operators, often relying on numerical simulations for solutions. These complex models may incorporate multiple operator types,[22] different peering and pricing strategies,[23] the network formation process,[24] prioritized access and net neutrality,[25].

[22]See, for examples, Tan, Chiang, & Mookerjee (2006), Suksomboon, Pongpaibool, & Aswakul (2008), Ma et. al (2010).

[23]See, for examples, Huston (1999), Laffont, et al. (2001), Chang, Jamin, & Willinger (2006), Faratin et al. (2007), Jahn & Prufer (2008), Lodhi, Dhamdhere, & Dovrolis (2012b, 2014b).

[24]See, for examples, Dhamdhere & Dovrolis (2010), Lodhi, Dhamdhere, & Dovrolis (2012a, 2013).

[25]See, for examples, Gyarmati, et al. (2007), Ma & Misra (2013), Tang & Ma (2014), Choi, Jeon, & Kim (2015), Ma (2017), Ma, Wang, & Chiu (2017).

4. Data

In this section, we describe our data. Our data comes from various data sources collected and compiled by the Center of Applied Internet Data Analysis (CAIDA) at the University of California, San Diego. Since 1998, CAIDA has been studying interconnectivity of the Internet by actively probing the Internet using its many monitors placed at various vantage points around the world. Its current flagship active measurement infrastructure, Archipelago, collects interconnectivity data on the IP-address-level from more than 200 monitors located on 6 continents in over 60 countries. CAIDA also collaborates with many organizations and compiles data collected from their monitors. Most notably, it collaborates with the Route Views Project at the University of Oregon and The Réseaux IP Européens Network Coordination Centre (RIPE NCC) in Europe to collect BGP routing tables that contain network-level interconnection paths announced across the Internet. Our main data on the network-level interconnection agreements comes from the routing tables, while our lower IP-address-level interconnection points for each agreement come from the active probes (Figure 2 visualizes the different levels at which data is collected and their relationships). CAIDA also gathers records of network registration information from the world's five regional Internet registries (RIRs), allowing us to identify countries or territories of organizations owning individual networks.[26]

Table 1 provides a summary of the variables used in this paper, describing their units of observations, frequency, sources and definitions. Table 2 presents summary statistics of variables described in Table 1. In the remainder of this section we discuss the data collection process and the caveats of data sources. For additional information, please refer to the Data Appendix

[26] A complete list of countries and territories in our sample is presented in Table A1.

section at the end of this paper.

As shown in Table 1, a number of our key variables come from a dataset called *AS Relationships*. The dataset contains network-to-network level interconnection agreements extracted from routing tables contributed by Route Views and RIPE NCC. To correctly route data across the Internet, networks exchange routing and reachability information through a protocol called the Border Gateway Protocol (BGP). Each network router using the BGP protocol maintains a routing table. The table contains the connectivity information of the network and its immediate neighbors in the Internet and lists paths to particular network destinations. By placing monitors that peer directly with large networks, we can extract the full set of agreements used between the collecting networks and all visible destinations.

We then annotate the extracted agreements with algorithmically-inferred agreement types, as network operators consider the details of their business relationships as proprietary information and do not generally make them public. Our inference algorithm (Luckie et al., 2013) draws from a long literature of this type of inference including Gao (2001), Subramanian et al. (2002), Di Battista et al. (2003), Erlebach et al. (2002), Xia and Gao (2004), Dimitropoulos et al. (2007a) and Dimitropoulos et al. (2007b). It achieved over 98% accuracy of agreement type inference via direct validation with a set of network operators (Luckie, et al., 2013). The algorithm is able to infer 96% of the agreement types in our sample.

The AS Relationships dataset is computed monthly. We use data from January 2015 to June 2019 for our analysis. We first count the number of observed agreements each network has in this data and make the variable $numAgNtwrk_{kt}$. We then aggregate individual agreements to the number of agreements between networks owned by each pair of countries (or territories) i and

j to construct the variable $numAg_{ijt}$. Breaking down the number of agreements between each country (or territory) pair by their agreement types, we make three variables $numProvAg_{ijt}$, $numPeerAg_{ijt}$, $numCustAg_{ijt}$ for when country (or territory) i's networks are providers to, peers to, and customers of country (or territory) j's networks respectively. We measure a network's centrality in the Internet by its customer cone, a commonly used measure of the number of networks that pay it directly or indirectly for transit. A network's customer cone is defined as itself and all the networks it was observed reaching following provider-to-customer agreements. Networks with larger customer cones have an especially important role in interconnecting the global Internet. We make the variable $NtwrkCustCone_{kt}$ for each network.

Our IP-address-level interconnection points within each agreement come from a dataset called *IPv4 Prefix-Probing*. The dataset consists of daily traceroutes from a subset of our Archipelago monitors to every announced BGP routing prefix in the Internet. Each traceroute tries to reach each destination prefix and records the entire IP address-by-IP address path it takes. We then map each IP address to its network with the help of Route Views Prefix-to-AS mappings dataset (CAIDA, 2013) and bdrmapIT tool (Marder et al., 2018), identify IP pairs that form inter-network links and label the observed interconnection links by their IP addresses and network identifiers. The IPv4 Prefix-Probing dataset is available since December 2015 on a daily basis from multiple monitors, so we use data from December 2015 to June 2019. We do two aggregations. First we aggregate daily captures from multiple monitors to weekly captures of unique IP-address-to-IP-address connections. Then we aggregate individual connections to the number of connections between each pair of networks k and l. The resulting variable is $numAgIP_{klt}$.

Although we know of no more rigorous data collection efforts of interconnection on the internet layer, we recognize that our data has limitations. First, networks owned by organizations headquartered in a particular country or territory can have multiple points of presence (PoP) in many countries and locations within a country and a single Internet interconnection can represent multiple geographically distinct physical connections. Geolocating points of presence is a hard and an open question, so it is important to note the country subscripts of our variables indicate network ownership by organizations headquartered in those countries or territories instead of the exact physical locations of the networks. This measure is especially problematic for large global transit providers and content providers which have PoPs both within and outside the EEA. However, we note that though the relatively few large networks account for a substantial portion of global internet traffic, the typical network is small and has limited geographic reach beyond its country of origin.[27] Throughout this paper, we use unweighted measures of the number of networks and the number of interconnections. This to some extent alleviates the concern that the imperfect measurement of locations of a few large networks drives the results.

Second, the number of agreements we capture, though extremely large, is a subset of all agreements. Individual routers do not maintain a full set of Internet paths, but rather a set of "best" paths for each destination based on local preferences. Networks also do not announce their peer-to-peer paths to their providers so many peer-to-peer agreements are not observable in the data we use. A truly complete set of agreements would require collecting BGP tables and traceroute data from vantage points in the majority of Internet networks, while our data

[27] For reference, if we measure the combined value of an organization's users and content purely in terms of the number of IP addresses in its customer cone, an organization at the 95% percentile only accounts for 0.01% of the full routed IP address space, an organization at the 99% percentile accounts for 0.2%, while Amazon.com, Inc. accounts for 1.21%.

collection is limited to vantage points where we have our own or partner monitors. Over time, monitors were added at new vantage points, resulting in more visibility in parts of the Internet and hence a greater number of discoverable agreements. To keep visibility consistent throughout our sample periods, we extracted agreements only from a set of monitors that operated throughout our sample periods, January 2015–June 2019 for AS Relationships and December 2015–June 2019 for IPv4 Prefix-Probing.

Moreover, sometimes technical problems occur with monitors, resulting in changes in visibility of some paths. In October 2018, configuration changes in three RIPENCC partner monitors placed in Amsterdam, Barcelona and Zurich caused permanent disappearance of around 2450 network-to-network interconnections from our sample. We dropped all of the affected interconnections throughout our sample. We also note interconnection agreements are more complex than allowed for in our approach. The types of agreements between the same two networks can differ by peering location or even by prefix. Our inference algorithm oversimplifies these cases by assigning a single agreement type to each pair of networks (CAIDA, 2015-2019a). Finally it is important to note that connectivity is not traffic, though there is evidence that IP address space advertised by BGP tables are strongly positively correlated with networks' self-reported traffic volume for a large set of peer-to-peer interconnections (Lodhi, et al., 2014a). We do not know how much traffic is exchanged across an interconnection or how that traffic has changed over time. If major changes in traffic occurred purely through existing interconnections, causing increased or decreased investment in Internet infrastructure, it would be invisible in our data. More monitor vantage points and additional sources of data would further improve our data quality.

5. Results

In this section, we outline our empirical strategy as well as present regression results for each outcome variable shown in Table 1. Our descriptives, which we present in a series of figures below, show persistent and similar growth in internet interconnection of EEA countries versus non-EEA OECD countries, though the levels of interconnectedness differ. This motivates our difference-in-difference approach. We treat the GDPR's April 2016 approval and May 2018 enforcement as two cutoff dates for periods post policy treatment. We define two variables for the post period: $POST_e$ is an indicator variable equal to 1 if the observation is made after the GDPR became effective, $POST_a$ is an indicator variable equal to 1 if the observation is made after the GDPR was approved. We expect the main effect of the GDPR on internet interconnection, if any, to occur after the GDPR became effective, since our model predicts changes in the bargaining outcomes due to changes in actual use. In addition, this date emerges as central in all available empirical evidence at the application layer.

The assignment of treatment status requires more explanation. As shown in Table 1, there are two types of outcome variables: ones where the units of observations are networks or countries, and ones where the units of observations are network or country pairs. For outcome variables where the units of observations are networks or countries, we simply assign networks or countries in the EEA to the treatment group. We assign networks or countries not in the EEA but in the OECD to the control group, because networks owned by organizations headquartered in developed countries have similar levels of interconnectedness prior to the GDPR. As we will show in a series of graphs later in this section, the parallel pre-trends needed for the difference-in-difference approach are visually apparent for the treatment and control groups. Networks owned by organizations headquartered in non-EEA non-OECD countries or territories are excluded from the control group as networks in developing countries might behave differently in their interconnection behavior from networks in more developed countries prior to the GDPR. Appendix Table A1 presents complete lists of countries and territories that are in the EEA (treatment group), are not in the EEA but in the OECD (control group), and are neither in the EEA nor in the OECD (excluded).

For variables where the units of observations are network or country pairs, it is reasonable to believe either interconnecting party's affiliation with the EEA may impact the agreement, and if both parties are affiliated with the EEA, the effect may be different from that if only one party is affiliated with the EEA. As this is the case, we need to hold fixed the EEA membership status (or OECD status) of the counterparty of interconnection while we compare the outcomes for networks or countries in the EEA (treatment group) and for networks or countries in the OECD but not in the EEA (control group).

We therefore construct three subsamples for each variable based on counterparties: (a) the counterparties are in the EEA, (b) the counterparties are in the OECD but not in the EEA, (c) the counterparties are not in the EEA and not in the OECD. Within each subsample, we then keep only observations where networks or countries are in the EEA (treatment group) or are in the OECD but not in the EEA (control group) and compare their outcomes. As a result, subsample (a) allows us to compare network or country pairs that are EEA–EEA versus non-EEA OECD–EEA. Subsample (b) allows us to compare network or country pairs that are EEA–non-EEA OECD versus non-EEA OECD–non-EEA OECD. Subsample (c) allows us to compare network or country pairs that are EEA–non-EEA non-OECD versus non-EEA OECD–non-EEA non-

OECD. Note one observation can contribute to multiple subsamples, for example a country pair France–US can contribute to both the control group in (a) and the treatment group in (b). Appendix Figure A1 illustrates visually the construction of the three subsamples.

For the rest of this section, we present regression specifications and results for each variable. We first present results for outcomes on the interconnection decisions: the number of agreements for each network ($numAgNtwrk_{kt}$), the number of agreements for each country pair ($numAg_{ijt}$), and the number of agreements for each country pair divided into provider-to-customer, peer-to-peer and customer-to-provider agreements ($numProvAg_{ijt}$, $numPeerAg_{ijt}$, $numCustAg_{ijt}$). We then present results on the lower IP-address level: the number of IP-address-level interconnection points between a network pair, conditional on the two networks having an agreement ($numAgIP_{klt}$). Last but not least, we present additional measures on the growth and interconnectedness of networks: the number of networks owned by organizations headquartered in a country ($numNtwrk_{it}$) and the number of networks a network can reach purely through its customers ($NtwrkCustCone_{kt}$).

5.1 The Number of Agreements by Networks

Figure 3 shows a comparison of the average log number of agreements by networks in the EEA countries and in the non-EEA OECD countries. We observe visually apparent parallel trends between the two groups prior to the approval of the GDPR, between the approval and implementation of the GDPR, as well as after the implementation of the GDPR. We then run the following regression,

$$log(numAgNtwrk_{kt}) = \beta_0 + \beta_1 POST_{e,kt} \times EEA_{kt} + \beta_2 POST_{a,kt} \times EEA_{kt} + \gamma_k D_k + \lambda_t D_t + \varepsilon_{kt}. \tag{9}$$

We take the log of $numAgNtwrk_{kt}$ to reflect estimated effects in percentage changes. EEA_{kt} is an indicator variable equal to 1 if network k is owned by an EEA country, and equal to 0 if network k is owned by a non-EEA OECD country. A dummy D_k for each network k and a dummy D_t for each month t are included. The difference-in-difference effect is identified by the coefficients on the interaction terms $POST_{e,kt} \times EEA_{kt}$ and $POST_{a,kt} \times EEA_{kt}$. The results are shown in Table 3, and in no case are significantly different from zero, confirming our visual impression.

5.2 The Number of Agreements between Countries

While the above subsection looks at the number of agreements associated with individual networks, in this subsection we look at the number of agreements between country pairs. Figure 4 shows a comparison of the total number of agreements in the EEA countries and in the non-EEA OECD countries, holding fixed the counterparties. We make a few observations. First, EEA countries have more agreements with EEA countries than with non-EEA OECD countries, and vice versa for non-EEA OECD countries. Second, despite the differences in levels, EEA countries and non-EEA OECD countries exhibit remarkable parallel trends in setting up agreements with counterparties that are EEA countries, non-EEA OECD countries, and non-EEA non-OECD countries throughout the sample period. Third, agreements with developing countries or territories have a lot more noise in measurement compared to agreements within EEA and OECD countries.

We then run the following regression on each of the three subsamples,

$$log(numAg_{ijt}+1) = \beta_0 + \beta_1 POST_{e,ijt} \times EEA_{ijt} + \beta_2 POST_{a,ijt} \times EEA_{ijt} + \gamma_{ij} D_{ij} + \lambda_t D_t + \varepsilon_{ijt}. \tag{10}$$

We take the log of $numAg_{ijt}$ to reflect estimated effects in percentage changes and add one to account for zero values when we take the log. EEA_{ijt} is an indicator variable equal to 1 if the country pair ij is in the treatment group for the subsample, and equal to 0 if the country pair ij is in the control group for the subsample. A dummy D_{ij} for each country pair ij and a dummy D_t for each month t are included. The difference-in-difference effect is identified by the coefficients on the interaction terms $POST_{e,ijt} \times EEA_{ijt}$ and $POST_{a,ijt} \times EEA_{ijt}$. The results are shown in Table 4. The main effect, based on the coefficient on $POST_{e,ijt} \times EEA_{ijt}$, is not significantly different from zero across the three subsamples. The only significant result in this table comes from the coefficient on $POST_{a,ijt} \times EEA_{ijt}$ for the non-EEA non-OECD counterparty subsample and we test the robustness of this result. Table 4 clusters standard error by country pair. Alternatively, one might expect the interconnection decisions of one particular country to other countries to have correlated errors. This may be especially true for interconnection decisions from an EEA or OECD country to developing countries based on the EEA/OECD networks' global interconnection strategy to remote and low demand areas. Therefore, we cluster standard error by EEA and OECD countries in the country pairs for the non-EEA non-OECD counterparty subsample as a robustness test, resulting in 43 clusters as compared to 6,751 clusters in Column 3 of Table 4. The coefficient on $POST_{a,ijt} \times EEA_{ijt}$ is no longer significant and is therefore likely a spurious result.

5.3 The Number of Agreements between Countries by Agreement Type

We then further break down the number of agreements between country pairs to provider-to-customer, peer-to-peer, and customer-to-provider types. Figure 5 shows a comparison of the total number of agreements in the EEA countries and in the non-EEA OECD countries, by agreement type. We still observe EEA countries and non-EEA OECD countries have remarkable parallel trends by agreement type throughout the sample period. Based on visual evidence, the GDPR does not have heterogeneous effects on different types of agreements.

We then run the following regression on each agreement type for each of the three counterparty subsamples,

$$log(numTypeAg_{ijt}+1) = \beta_0 + \beta_1 POST_{e,ijt} \times EEA_{ijt} + \beta_2 POST_{a,ijt} \times EEA_{ijt} + \gamma_{ij} D_{ij} + \lambda_t D_t + \varepsilon_{ijt}. \tag{11}$$

We take the log of $numTypeAg_{ijt}$ to reflect estimated effects in percentage changes and add one to account for zero values when we take the log. $numTypeAg_{ijt}$ refers to $numProvAg_{ijt}$, $numPeerAg_{ijt}$, or $numCustAg_{ijt}$. EEA_{ijt} is an indicator variable equal to 1 if the country pair ij is in the treatment group for the subsample, and equal to 0 if the country pair ij is in the control group for the subsample. A dummy D_{ij} for each country pair ij and a dummy D_t for each month t are included. The difference-in-difference effect is identified by the coefficients on the interaction terms $POST_{e,ijt} \times EEA_{ijt}$ and $POST_{a,ijt} \times EEA_{ijt}$. The results are shown in Table 5.

We see a few significant results in the non-EEA non-OECD counterparty subsample. As previously, once we cluster standard error by EEA and OECD countries in the country pairs for the non-EEA non-OECD counterparty subsample as a robustness test, the significance of these results disappear. We also note these results, though sometimes significant, lack systematic patterns and are economically small in magnitude.[28]

5.4 The Number of IP-Address-Level Interconnection Points per Agreement

Our earlier results suggest the GDPR so far does not change whether agreements are made and what types of agreements are made between networks. One hypothesis for the absence of behavior change is that setting up an agreement is such a substantial decision that changes in usage and bargaining friction due to the GDPR are small in comparison. Networks may only change the capacity associated with each interconnection in response to lower usage instead of cancelling an agreement altogether. If that is the case, we are unlikely to observe effects of the GDPR on the extensive margin. The GDPR's impact may be on how dense the two networks' interconnection is. Motivated by this consideration, we examine how the GDPR affects the number of IP-address-level interconnection points two networks have, conditional on them having an agreement.

Figure 6 shows a comparison of the average log number of interconnection points per agreement in the EEA countries and in the non-EEA OECD countries, holding fixed the interconnec-

[28]To illustrate how economically small the implied effect based on the coefficients is, we take the coefficient -0.038 on $POST_e \times EEA$ from column (2) of Table 5 as an example. The dependent variable for the regression in column (2) is $log(numPeerAg_{ijt}+1)$. It has a mean of 0.109 and an SD of 0.544. Therefore, being in the treatment group post GDPR effective date has an effect which is a tiny fraction of one standard deviation of the outcome.

tion counterparties. We make a few observations. First, EEA countries have more observed interconnection points per agreement with non-EEA OECD countries than with EEA countries, and vice versa for non-EEA OECD countries. Second, observed interconnection points with developing countries have a lot of noise in our measurement while observed interconnection points among EEA and OECD countries are quite precisely measured, reflecting the large number of vantage points inside developed countries. When interconnection points are well-measured, we observe that, despite the differences in levels, EEA countries and non-EEA OECD countries still exhibit remarkable parallel trends in terms of the number of interconnection points per agreement throughout the sample period.

We then run the following regression on each of the three subsamples,

$$log(numAgIP_{klt}) = \beta_0 + \beta_1 POST_{e,klt} \times EEA_{klt} + \gamma_{kl} D_{kl} + \lambda_t D_t + \varepsilon_{klt}. \tag{12}$$

We take the log of $numAgIP_{klt}$ to reflect estimated effects in percentage changes. EEA_{klt} is an indicator variable equal to 1 if the network pair kl is in the treatment group for the subsample, and equal to 0 if the network pair kl is in the control group for the subsample. A dummy D_{kl} for each network pair kl and a dummy D_t for each week t are included. The difference-in-difference effect is identified by the coefficient on the interaction term $POST_{e,klt} \times EEA_{klt}$. Given this particular data source only started in December 2015, close to the GDPR approval date, we do not include $POST_{a,klt} \times EEA_{kljt}$. The results are shown in Table 6 and are in no case significantly different from zero. We include agreements present for at least 150 weeks for our regressions in Table 6. Given we are studying the intensive margin, alternatively we keep only agreements present for all of 169 weeks between December 2015 and June 2019. Doing so substantially reduces the sample size and the results are similar to those in Table 6.

In addition to interconnection behavior of networks, we study how the GDPR might have impacted the growth in the number of networks in Europe and the sizes of the customer cones of these networks.

5.5 The Number of Networks

Figure 7 shows a comparison of the average log number of networks per country in the EEA countries and in the non-EEA OECD countries. Again, we observe visually apparent parallel trends between the two groups prior to the approval of the GDPR, between the approval and implementation of the GDPR, as well as after the implementation of the GDPR. We then run the following regression,

$$log(numNtwrk_{it}) = \beta_0 + \beta_1 POST_{e,it} \times EEA_{it} + \beta_2 POST_{a,it} \times EEA_{it} + \gamma_i D_i + \lambda_t D_t + \varepsilon_{it}. \quad (13)$$

We take the log of $numNtwrk_{it}$ to reflect estimated effects in percentage changes. EEA_{it} is an indicator variable equal to 1 if country i is an EEA country, and equal to 0 if country i is a non-EEA OECD country. A dummy D_i for each country i and a dummy D_t for each quarter t are included. The difference-in-difference effect is identified by the coefficients on the interaction terms $POST_{e,it} \times EEA_{it}$ and $POST_{a,it} \times EEA_{it}$. The coefficient on $POST_e \times EEA = -.002$ ($se = 0.017$, clustered by country) and the coefficient on $POST_a \times EEA = 0.016$ ($se = 0.024$, clustered by country). Both are insignificant at conventional levels of significance. This result suggests the GDPR does not impact the number of networks in EEA countries compared to non-EEA OECD countries.

5.6 Customer Cone of Networks

Figure 8 shows a comparison of the average log customer cone of networks in the EEA countries and in the non-EEA OECD countries. We observe visually apparent parallel trends between the two groups prior to the approval of the GDPR, between the approval and implementation of the GDPR, as well as after the implementation of the GDPR. We then run the following regression,

$$log(NtwrkCustCone_{kt}) = \beta_0 + \beta_1 POST_{e,kt} \times EEA_{kt} + \beta_2 POST_{a,kt} \times EEA_{kt} + \gamma_k D_k + \lambda_t D_t + \varepsilon_{kt}. \tag{14}$$

We take the log of $NtwrkCustCone_{kt}$ to reflect estimated effects in percentage changes. EEA_{kt} is an indicator variable equal to 1 if network k is owned by an EEA country, and equal to 0 if network k is owned a non-EEA OECD country. A dummy D_k for each network k and a dummy D_t for each month t are included. The difference-in-difference effect is identified by the coefficients on the interaction terms $POST_{e,it} \times EEA_{it}$ and $POST_{a,it} \times EEA_{it}$. The coefficient on $POST_e \times EEA = -.007^*$ ($se = 0.004$, clustered by country) and the coefficient on $POST_a \times EEA = 0.011^{***}$ ($se = 0.004$, clustered by country). Though both are significantly different from zero, their magnitudes are economically very small, suggesting the GDPR has little impact on the centrality of networks in EEA countries compared to non-EEA OECD countries.

6. Conclusion

The effectiveness of the Internet in creating economic surplus depends on efficient interconnections bilaterally negotiated by independently operated networks. Simple bargaining theory predicts that costly policy at the application layer could have unintended effects on negotiating interconnection agreements at the internet layer. A similar prediction emerges from news reports and editorial at the time the policy implemented, as well as from all empirical evidence of the impact at the application layer. In this paper, we investigate whether the approval and implementation of the GDPR affects the growth and interconnection of the Internet in Europe. Despite evidence that these stringent set of privacy regulations so far had significant effects at the application layer, we find no visible consequences at the infrastructure layer, rejecting this hypothesis. Across multiple measures, we estimate precise zeros effects of the GDPR. Occasionally we estimate statistically significant effects, which prove to be not robust.

A number of possible reasons could have contributed to this finding. First, the lack of discernible short-run effect at the internet layer could have arisen from slow investment and behavioral changes at the internet layer. This seems unlikely because renegotiations of interconnection agreements happen frequently and we observe continued growth across all network connections. It is also possible that despite the large behavioral changes at the application layer due to the GDPR, the effect is small compared to other considerations in negotiating interconnection agreements. That could happen if, for example, the regular growth in data due to growth in many applications overwhelms any short-run impact of the GDPR. In that case, network operators may rationally expect the long run effect of the GDPR to be small even at the application layer. Finally, we only observe the short run, so we cannot rule out that more gradual changes due to the GDPR may surface in the longer run, which is an open question. If we are able to observe a longer period of time, we will be able to use the data from additional periods and the same methodology to study the effect of the GDPR in the longer run.

Our results have immediate policy implications. As many countries are contemplating implementing their own versions of privacy and data protection regulations, there are concerns about whether such regulations may negatively impact the growth of the Internet, reduce technology firms' incentives in operating and innovating, reduce the use of the Internet in productivity enhancing activities, and reduce the economic surplus generated through the use of the Internet in the country and beyond. Our results suggest limited effects of such regulations on the internet layer. Said another way, our results suggest the costs are concentrated at the application layer.

Our results also speak to the debate on the allocation of rents generated through the successful commercialization of the Internet. The enormous rents associated with the exploitation of Web 2.0 and mobile web represent a large portion of the private returns to innovation in the 21st century. These rents have been overwhelmingly captured by players at the application layer, notably the "big tech" companies, while firms at the internet layer captures little of the rents. Our study is consistent with the view that the cost of the GDPR has been a shock to rents, and the costs have been borne by the application layer, paid out of the rents from innovation.

In addition to policy implications, our paper contributes by presenting data consisting of virtually all operating networks in the world and a large number of interconnection agreements among them across many years, which opens the possibility of investigating a range of economic- and policy-relevant questions about the Internet. Across the academic and policy arena, the lack of well-measured data describing the interconnectivity and data flow in the Internet has brought great attention, especially in issues such as net neutrality, international trade in digitally delivered goods, and market power of big technology firms whose data flows dominate global internet traffic. While the theoretical literature has dabbled at many of these issues, empirical literature is scant. Our data may be a small progress towards filling the data gap in growing needs to analyze internet-related economic and policy issues.

References

Acquisti, Alessandro, Leslie K. John, and George Loewenstein. 2013. ”What is Privacy Worth?” The Journal of Legal Studies 42, no. 2: 249-274.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. ”The Economics of Privacy.” Journal of Economic Literature 54, no. 2: 442-92.

Athey, Susan, Christian Catalini, and Catherine Tucker. 2017. “The Digital Privacy Paradox: Small Money, Small Costs, Small Talk.” National Bureau of Economic Research Working Paper No. w23488.

Aguiar, Luis, and Joel Waldfogel. 2014. “Digitization, Copyright, and the Welfare Effects of Music Trade.” Available at SSRN: https://ssrn.com/abstract=2603238.

Badasyan, Narine, and Subhadip Chakrabarti. 2008. “A Simple Game-Theoretic Analysis of Peering and Transit Contracting among Internet Service Providers.” Telecommunications Policy 32, no. 1: 4-18.

Besen, Stanley, Paul Milgrom, Bridger Mitchell, and Padmanabhan Srinagesh. 2001. “Advances in Routing Technologies and Internet Peering Agreements.” American Economic Review 91, no. 2: 292-296.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky. 1986. “The Nash Bargaining Solution in Economic Modelling.” The RAND Journal of Economics: 176-188.

Campbell, James, Avi Goldfarb, and Catherine Tucker. 2015. “Privacy Regulation and Market Structure.” Journal of Economics & Management Strategy 24, no. 1: 47-73.

Chang, Hyunseok, Sugih Jamin, and Walter Willinger. 2006. “To Peer or Not to Peer: Modeling the Evolution of the Internet’s AS-level Topology.” In Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications, pp. 1-12. IEEE, 2006.

Center for Applied Internet Data Analysis, The. 2015-2019a. “AS Relationships.” University of California, San Diego. http://www.caida.org/data/as-relationships/ (accessed June-July 2019).

Center for Applied Internet Data Analysis, The. 2015-2019b. “Inferred AS to Organization Mapping Dataset.” University of California, San Diego. http://www.caida.org/data/as-organizations/ (accessed June-July 2019).

Center for Applied Internet Data Analysis, The. 2015-2019c. “IPv4 Prefix-Probing Traceroute Dataset.” University of California, San Diego.

https://www.caida.org/data/active/ipv4_prefix_probing_dataset.xml (accessed June-July 2019).

Center for Applied Internet Data Analysis, The. 2015-2019d. "Macroscopic Internet Topology Data Kit (ITDK)." University of California, San Diego.

http://www.caida.org/data/internet-topology-data-kit/ (accessed June-July 2019).

Center for Applied Internet Data Analysis, The. 2013. "Routeviews Prefix-to-AS mappings (pfx2as) for IPv4 and IPv6." University of California, San Diego.

http://data.caida.org/datasets/routing/routeviews-prefix2as/README.txt (accessed September 21, 2019).

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim. 2015. "Net Neutrality, Business Models, and Internet Interconnection." American Economic Journal: Microeconomics 7, no. 3: 104-41.

Degeling, Martin, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz.2018. "We Value Your Privacy... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy." arXiv preprint arXiv:1808.05096.

Dhamdhere, Amogh, and Constantine Dovrolis. 2010. "The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh." In Proceedings of the 6th International COnference, p. 21. ACM, 2010.

Dhamdhere, Amogh, Constantine Dovrolis, and Pierre Francois. 2010. "A Value-Based Framework for Internet Peering Agreements." In 2010 22nd International Teletraffic Congress (lTC 22), pp. 1-8. IEEE, 2010.

Di Battista, Giuseppe, Maurizio Patrignani, and Maurizio Pizzonia. 2003. "Computing the Types of the Relationships between Autonomous Systems." In IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428), vol. 1, pp. 156-165. IEEE.

D'Ignazio, Alessio, and Emanuele Giovannetti. 2009. "Asymmetry and Discrimination in Internet Peering: Evidence from the LINX." International Journal of Industrial Organization 27, no. 3: 441-448.

Dimitropoulos, Xenofontas, Dmitri Krioukov, Bradley Huffaker, and George Riley. 2005. "Inferring AS Relationships: Dead End or Lively Beginning?." In International Workshop on Experimental and Efficient Algorithms, pp. 113-125. Springer, Berlin, Heidelberg.

Dimitropoulos, Xenofontas, Dmitri Krioukov, Marina Fomenkov, Bradley Huffaker, Young Hyun, and George Riley. 2007. "AS Relationships: Inference and Validation." ACM SIGCOMM Computer Communication Review 37, no. 1: 29-40.

Economides, Nicholas. 2007. "Economics of the Internet." NYU Law and Economics Research Paper 07-07: 07-01.

Erlebach, Thomas, Alexander Hall, and Thomas Schank. 2002. "Classifying Customer-Provider Relationships in the Internet." TIK-Report 145.

Faratin, Peyman, David D. Clark, Steven Bauer, and William Lehr. 2007. “Complexity of Internet Interconnections: Technology, Incentives and Implications for Policy.”

Freund, Caroline, and Diana Weinhold. 2002. “The Internet and International Trade in Services.” American Economic Review 92, no. 2: 236-240.

Freund, Caroline L., and Diana Weinhold. 2004 “The Effect of the Internet on International Trade.” Journal of international economics 62, no. 1 (2004): 171-189.

Gao, Lixin. 2001. “On Inferring Autonomous System Relationships in the Internet.” IEEE/ACM Transactions on Networking (ToN) 9, no. 6: 733-745.

Godinho de Matos, Miguel, and Idris Adjerid. 2019. “The Impact of European Data Protection Regulation on Firm Targeting and Consumer Behavior.” Working paper.

Goldberg, Samuel, Garrett Johnson, and Scott Shriver. 2019. “Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes.” Available at SSRN 3421731.

Goldfarb, Avi, and Catherine Tucker. 2011a. “Online Display Advertising: Targeting and Obtrusiveness.” Marketing Science 30, no. 3: 389-404.

Goldfarb, Avi, and Catherine Tucker. 2011b. “Privacy Regulation and Online Advertising.” Management Science 57, no. 1: 57-71.

Goldfarb, Avi, and Catherine Tucker. 2012. “Privacy and Innovation.” Innovation Policy and the Economy 12, no. 1: 65-90.

Gyarmati, Laszlo, Nikolaos Laoutaris, Kostas Sdrolias, Pablo Rodriguez, and Costas Courcoubetis. 2014. “From Advertising Profits to Bandwidth Prices-A Quantitative Methodology for Negotiating Premium Peering.” arXiv preprint arXiv:1404.4208.

Hermalin, Benjamin E., and Michael L. Katz. 2006. ”Your Network or Mine? The Economics of Routing Rules.” The RAND Journal of Economics 37, no. 3: 692-719.

Internet Engineering Task Force. 1989. “RFC 1122: Requirements for Internet Hosts – Communication Layers.” R. Braden, Editor. https://tools.ietf.org/html/rfc1122.

Iordanou, Costas, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris. 2018. “Tracing Cross Border Web Tracking.” In Proceedings of the Internet Measurement Conference 2018, pp. 329-342. ACM, 2018.

Jahn, Eric, and Jens Prüfer. 2008. “Interconnection and Competition among Asymmetric Networks in the Internet Backbone Market.” Information Economics and Policy 20, no. 3: 243-256.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman. 2018. “The Short-Run Effects of GDPR on Technology Venture Investment.” National Bureau of Economic Research Working Paper No. w25248.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman. 2019. "GDPR and the Localness of Venture Investment." Available at SSRN 3436535.

Johnson, Garrett A., and Scott K. Shriver. 2019. "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR." Available at SSRN 3477686.

Hjort, Jonas, and Jonas Poulsen. 2019. "The Arrival of Fast Internet and Employment in Africa." American Economic Review 109, no. 3: 1032-79.

Huston, Geoff. 1990. ISP survival guide: strategies for running a competitive ISP. John Wiley & Sons, Inc.

Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole. 2001. "Interconnection and Access in Telecom and the Internet." In AEA Papers and Proceedings, vol. 91, no. 2, pp. 287-91.

Libert, Timothy, Lucas Graves, and Rasmus Kleis Nielsen. 2018. "Changes in Third-Party Content on European News Websites after GDPR." Reuters Institute for the Study of Journalism Reports: Factsheet.

Lodhi, Aemen, Amogh Dhamdhere, and Constantine Dovrolis. 2012a. "GENESIS: An Agent-Based Model of Interdomain Network Formation, Traffic Flow and Economics." In 2012 Proceedings IEEE INFOCOM, pp. 1197-1205.

Lodhi, Aemen, Amogh Dhamdhere, and Constantine Dovrolis. 2012b. "Analysis of Peering Strategy Adoption by Transit Providers in the Internet." In 2012 Proceedings IEEE INFOCOM Workshops, pp. 177-177.

Lodhi, Aemen, Amogh Dhamdhere, and Constantine Dovrolis. 2013. "GENESIS-CBA: An Agent-Based Model of Peer Evaluation and Selection in the Internet Interdomain Network." Complex Adaptive Systems Modeling 1, no. 1: 10.

Lodhi, Aemen, Natalie Larson, Amogh Dhamdhere, and Constantine Dovrolis. 2014a. "Using PeeringDB to Understand the Peering Ecosystem." ACM SIGCOMM Computer Communication Review 44, no. 2: 20-27.

Lodhi, Aemen, Amogh Dhamdhere, and Constantine Dovrolis. 2014b. "Open Peering by Internet Transit Providers: Peer Preference or Peer Pressure?" In IEEE INFOCOM 2014-IEEE Conference on Computer Communications, pp. 2562-2570.

Luckie, Matthew, Bradley Huffaker, Amogh Dhamdhere, and Vasileios Giotsas. 2013. "AS Relationships, Customer Cones, and Validation." In Proceedings of the 2013 conference on Internet measurement conference, pp. 243-256. ACM, 2013.

Ma, Richard TB, Dah Ming Chiu, John CS Lui, Vishal Misra, and Dan Rubenstein. 2010. "On Cooperative Settlement between Content, Transit, and Eyeball Internet Service Providers." IEEE/ACM Transactions on networking 19, no. 3: 802-815.

Ma, Richard TB, and Vishal Misra. 2013. “The Public Option: a Nonregulatory Alternative to Network Neutrality.” IEEE/ACM Transactions on Networking 21, no. 6: 1866-1879.

Ma, Richard TB. 2017. “Pay or Perish: The Economics of Premium Peering.” IEEE Journal on Selected Areas in Communications 35, no. 2: 353-366.

Ma, Richard TB, Jingjing Wang, and Dah Ming Chiu. 2017. “Paid Prioritization and its Impact on Net Neutrality.” IEEE Journal on Selected Areas in Communications 35, no. 2: 367-379.

Marder, Alexander, Matthew Luckie, Amogh Dhamdhere, Bradley Huffaker, and Jonathan M. Smith. 2018. “Pushing the Boundaries with bdrmapIT: Mapping Router Ownership at Internet Scale.” In Proceedings of the Internet Measurement Conference 2018, pp. 56-69. ACM, 2018.

Meltzer, Joshua. 2014. “The Importance of the Internet and Transatlantic Data Flows for US and EU Trade and Investment.” Global Economy and Development Working Paper 79.

Miller, Amalia R., and Catherine E. Tucker. 2011. “Can Health Care Information Technology Save Babies?.” Journal of Political Economy 119, no. 2: 289-324.

Mohan, Jayashree, Melissa Wasserman, and Vijay Chidambaram. 2019. “Analyzing GDPR Compliance Through the Lens of Privacy Policy.” arXiv preprint arXiv:1906.12038.

Nicholson, Jessica R., and Giulia McHenry. 2016. “Measuring Cross-Border Data Flows: Data, Literature, and Considerations.” US Department of Commerce. https://www.ntia.doc.gov/other-publication/2016/measuring-cross-border-data-flows-data-literature-and-considerations.

Norton, William B. 2010. “Internet Transit Prices - Historical and Projected.” http://drpeering.net/white-papers/Internet-Transit-Pricing-Historical-And-Projected.php/

Norton, William B. 2014. “The 2014 Internet Peering Playbook: Connecting to the Core of the Internet.” DrPeering Press.

Posner, Richard A. 1981. “The Economics of Privacy.” The American Economic Review 71, no. 2: 405-409.

Shiller, Benjamin, Joel Waldfogel, and Johnny Ryan. 2018. “The Effect of Ad Blocking on Website Traffic and Quality.” The RAND Journal of Economics 49, no. 1: 43-63.

Suksomboon, Kalika, Panita Pongpaibool, and Chaodit Aswakul. 2008. “An Equilibrium Policy for Providing End-to-End Service Level Agreements in Interdomain Network.” In 2008 IEEE Wireless Communications and Networking Conference, pp. 2963-2968.

Subramanian, Lakshminarayanan, Sharad Agarwal, Jennifer Rexford, and Randy H. Katz. 2002. “Characterizing the Internet Hierarchy from Multiple Vantage Points.” In Proc. IEEE Infocom 2002.

Tan, Yong, I. Robert Chiang, and Vijay S. Mookerjee. 2006. “An Economic Analysis of Interconnection Arrangements between Internet Backbone Providers.” Operations research 54, no. 4: 776-788.

Tang, Jing, and Richard TB Ma. 2019. “Regulating Monopolistic ISPs without Neutrality.” IEEE Journal on Selected Areas in Communications.

The Internet Society. 2015. “Policy Brief: Internet Interconnection.”
https://www.internetsociety.org/policybriefs/internetinterconnection/.

Xia, Jianhong, and Lixin Gao. 2004. “On the Evaluation of AS Relationship Inferences [Internet reachability/traffic flow applications].” In IEEE Global Telecommunications Conference, 2004. GLOBECOM’04., vol. 3, pp. 1373-1377. IEEE.

US International Trade Commission. 2014. “Digital Trade in the US and Global Economies, Part 2.” Available at: https://www.strtrade.com/media/publication/7238_pub4485.pdf.

Weller, Dennis, and Bill Woodcock. 2013. “Internet Traffic Exchange Market Development and Policy Challenges.” OECD Digital Economy Papers. https://doi.org/10.1787/5k918gpt130q-en.

Figure and Tables

Figure 1: Four layers of the Internet

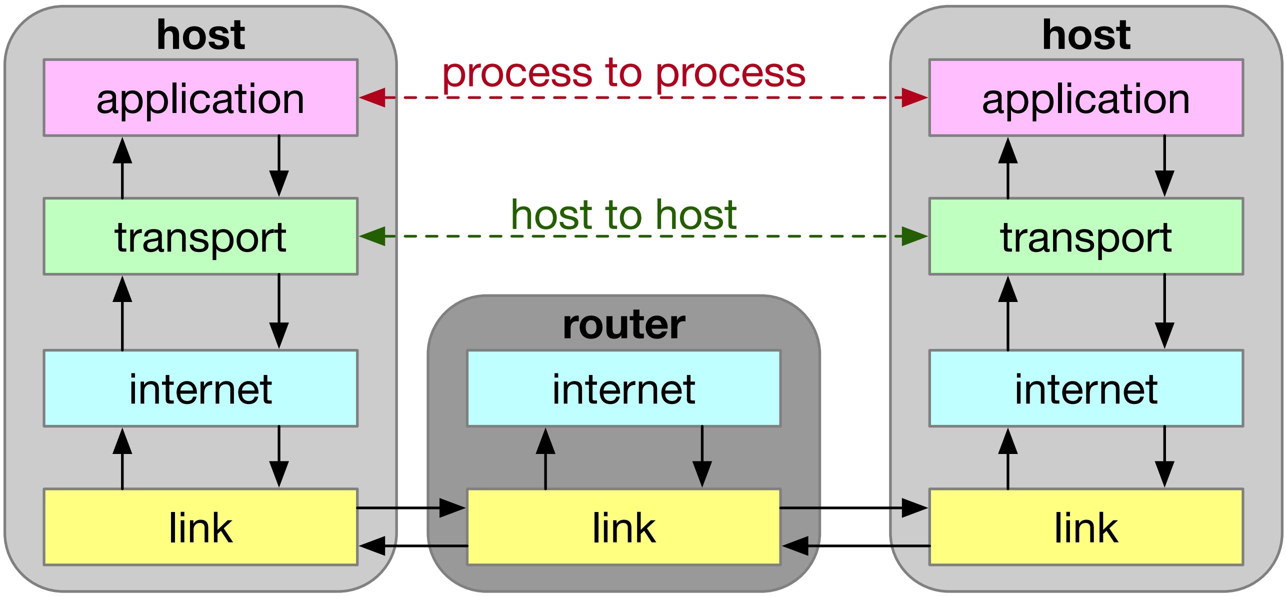
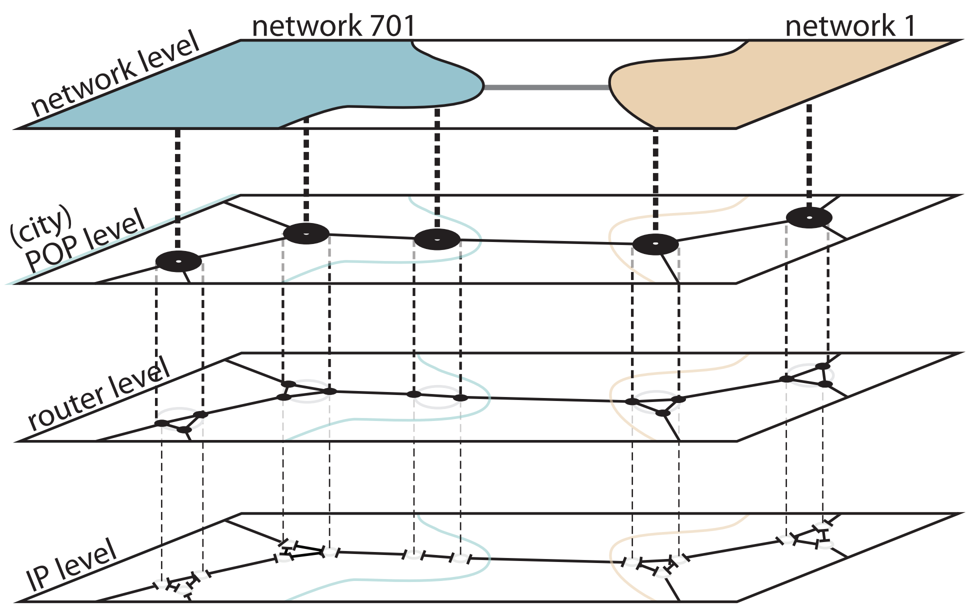
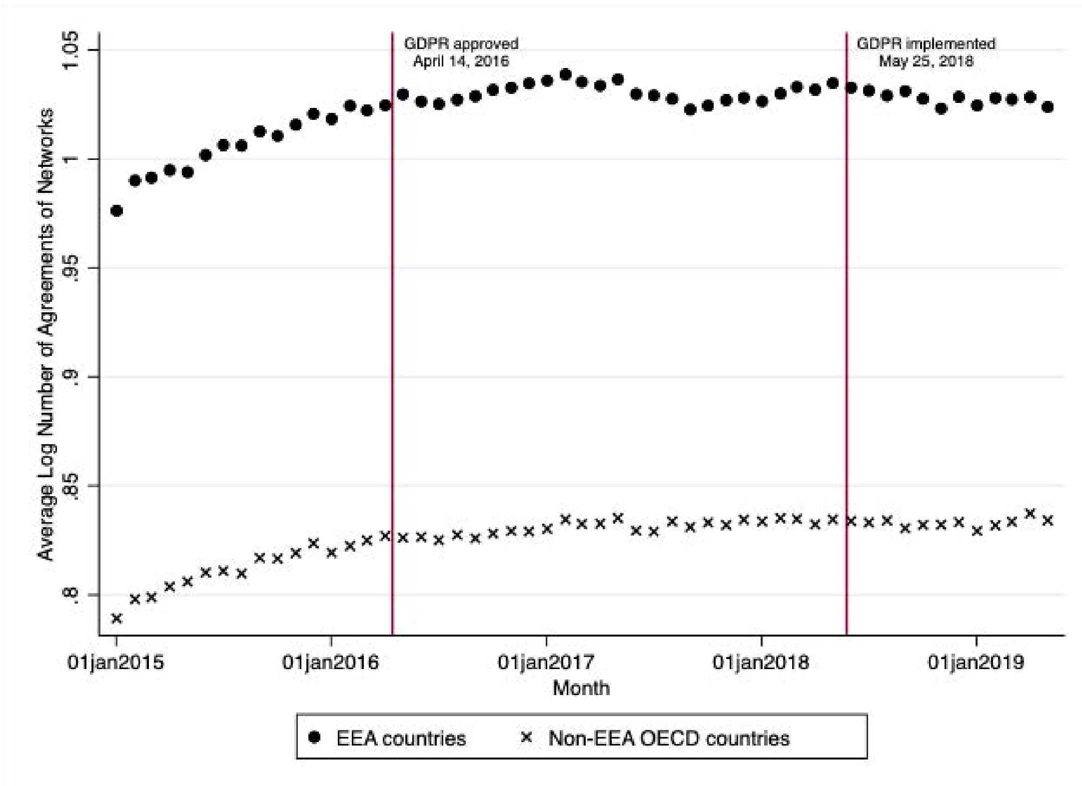


Figure 2: Data collection at the internet layer



Notes: Our network-level interconnection agreements extracted from routing tables correspond to the topmost level in this figure. Our IP-address-level interconnection points for each agreement extracted from active probes correspond to the bottom level in this figure. Geolocating points of presence (PoP) and mapping routers to networks are challenging and open questions, therefore we do not use data on the middle levels.

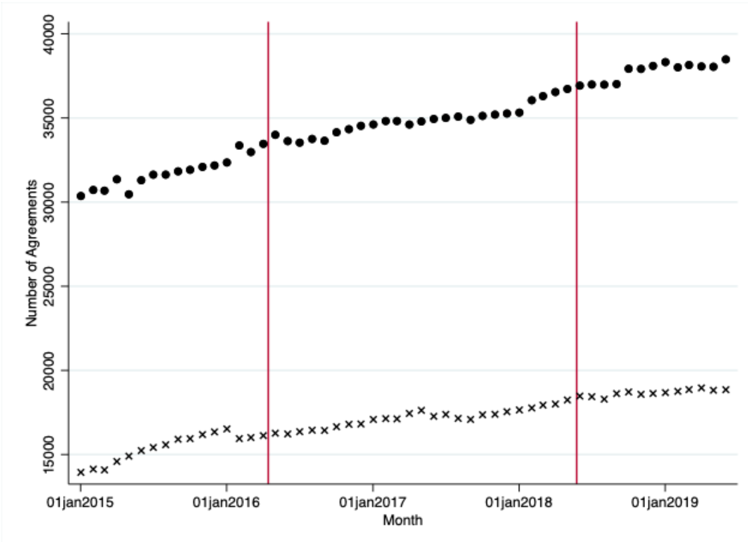
Figure 3: Average log number of interconnection agreements by networks in EEA and non-EEA OECD countries



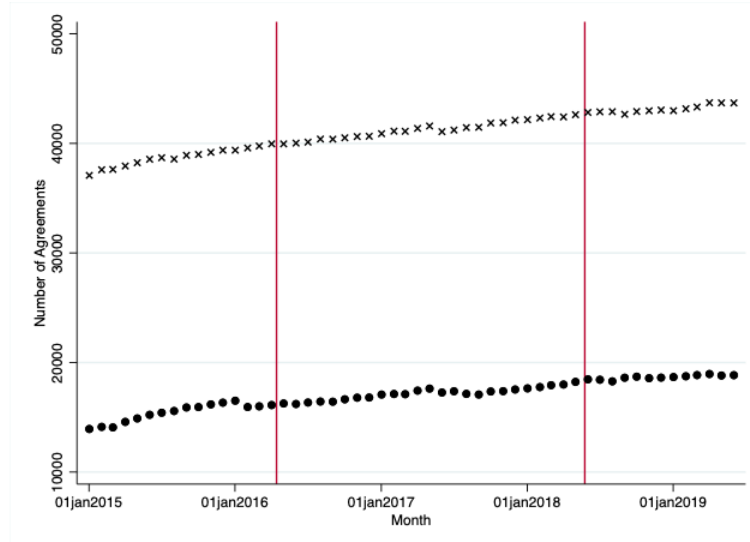
Notes: The dots represent $\overline{log(numAgNtwrk_{k \in EEA,t})}$, the log number of agreements averaged among networks owned by EEA countries. The crosses represent $\overline{log(numAgNtwrk_{k \in OECD \wedge k \notin EEA,t})}$, the log number of agreements averaged among networks owned by non-EEA OECD countries. Non-EEA and non-OECD countries' networks are not included in taking the averages. Only networks present throughout Jan 2015 – June 2019 are used to take the averages. The first red vertical line represents 14 April 2016, the approval date of the GDPR. The second red vertical line represents 25 May 2018, the implementation date of the GDPR.

Figure 4: Number of interconnection agreements by EEA and non-EEA OECD countries by counterparty

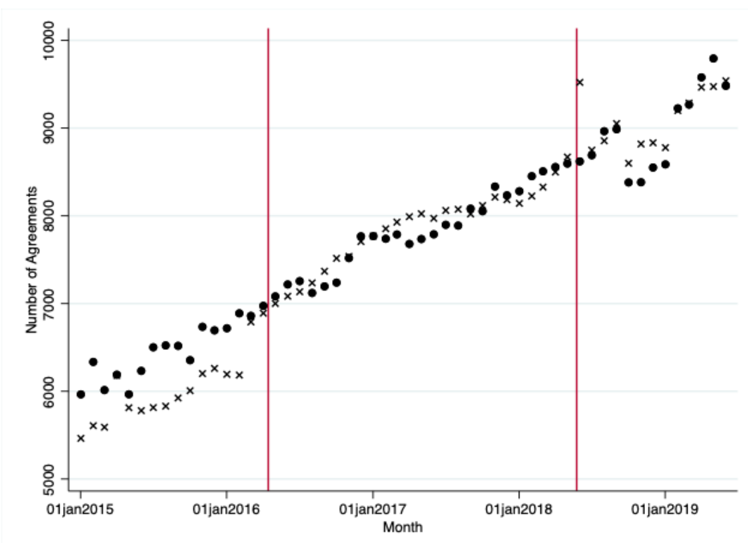
(a) Counterparties are EEA countries



(b) Counterparties are non-EEA OECD countries



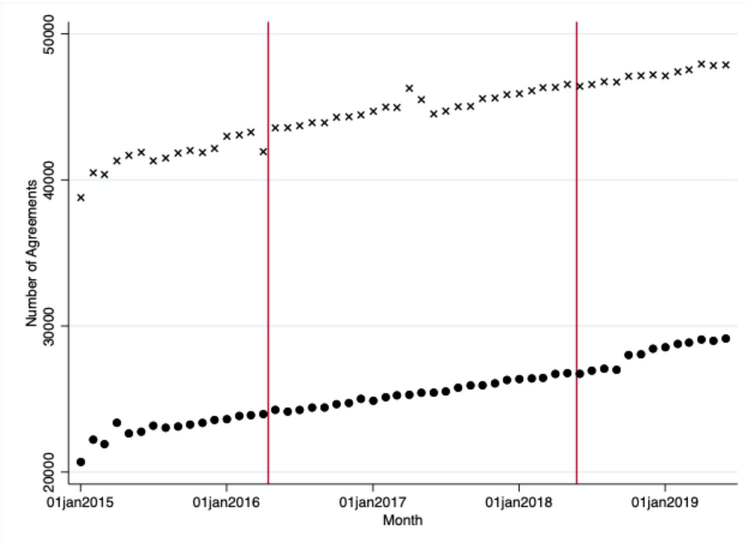
(c) Counterparties are non-EEA non-OECD countries



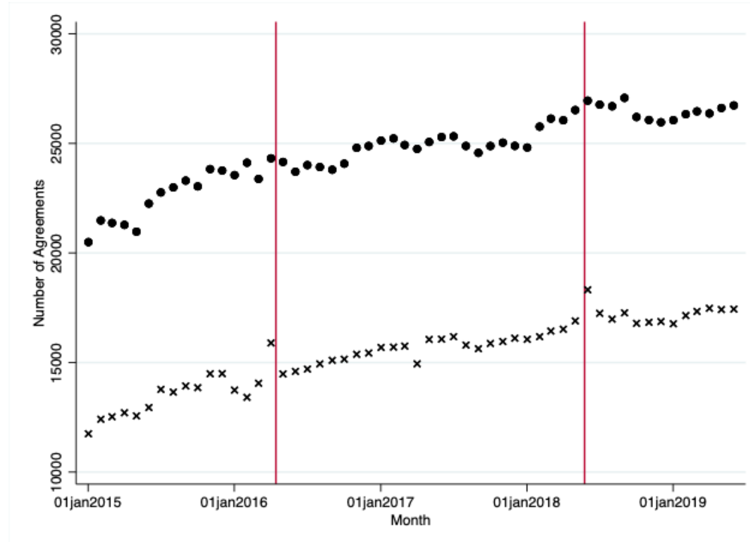
Notes: The dots represent $\sum_{i \in EEA,t} numAg_{ijt}$, the total number of agreements by networks owned by EEA countries when the counterparties are networks owned by (a) EEA countries, (b) non-EEA OECD countries, (c) non-EEA non-OECD countries. The crosses represent $\sum_{i \in OECD \wedge i \notin EEA,t} numAg_{ijt}$, the total number of agreements by networks owned by non-EEA OECD countries when the counterparties are networks owned by (a) EEA countries, (b) non-EEA OECD countries, (c) non-EEA non-OECD countries. The agreements of networks owned by non-EEA non-OECD countries when the counterparties are also networks owned by non-EEA non-OECD countries are not included in calculating these sums.

Figure 5: Number of agreements by EEA and non-EEA OECD countries by agreement type

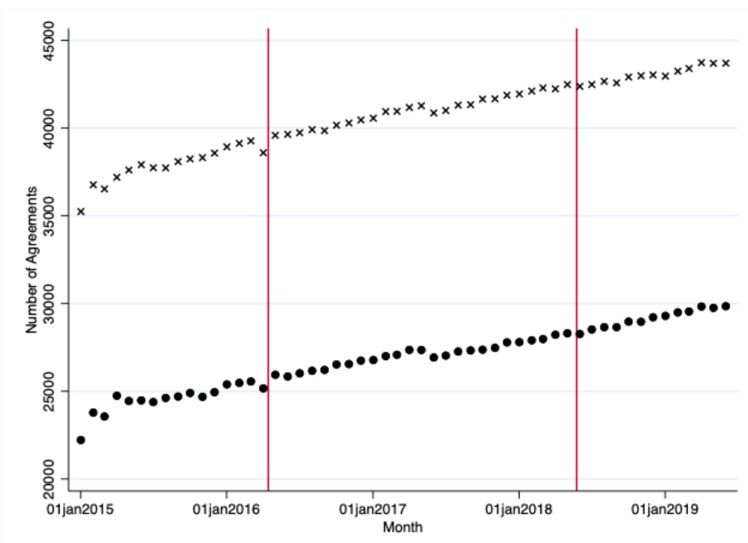
(a) Provider-to-customer (p2c) agreements



(b) Peering (p2p) agreements



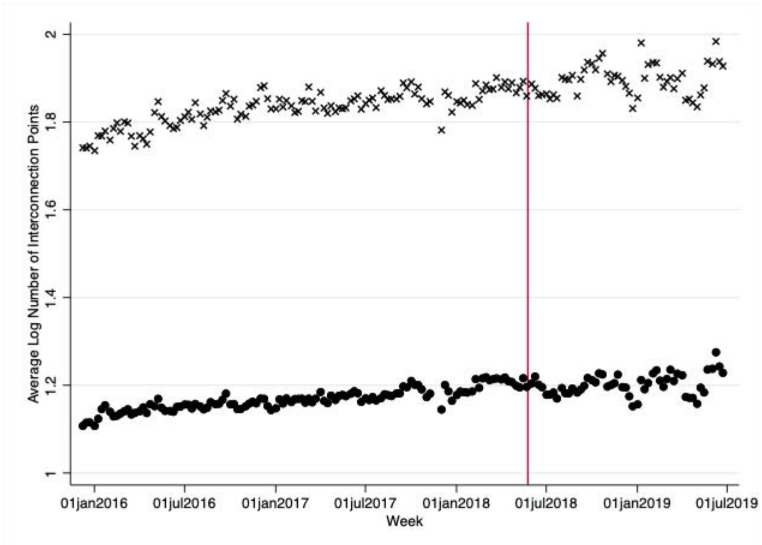
(c) Customer-to-provider (c2p) agreements



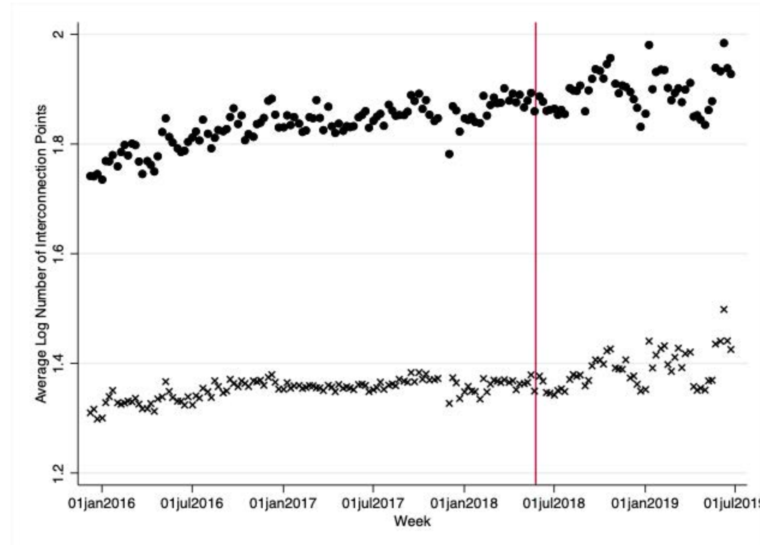
Notes: (a) The dots represent $\sum_{i\in EEA,t} numProvAg_{ijt}$, the total number of agreements by networks owned by EEA countries where these networks are providers. The crosses represent $\sum_{i\in OECD\wedge i\notin EEA,t} numProvAg_{ijt}$, the total number of agreements by networks owned by non-EEA OECD countries where these networks are providers.(b) The dots represent $\sum_{i\in EEA,t} numPeerAg_{ijt}$. The crosses represent $\sum_{i\in OECD\wedge i\notin EEA,t} numPeerAg_{ijt}$. Peering agreements between EEA countries and non-EEA OECD countries contribute to counts in both series. (c) The dots represent $\sum_{i\in EEA,t} numCustAg_{ijt}$. The crosses represent $\sum_{i\in OECD\wedge i\notin EEA,t} numCustAg_{ijt}$. Agreements between networks owned by non-EEA non-OECD countries are not included in calculating these sums.

Figure 6: Average log number of IP-address-level interconnection points per agreement by EEA and non-EEA OECD countries by counterparty

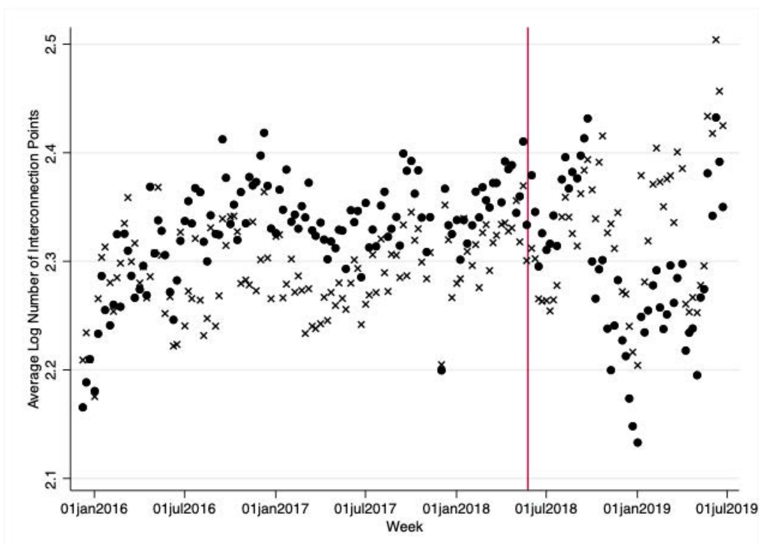
(a) Counterparties are EEA countries



(b) Counterparties are non-EEA OECD countries

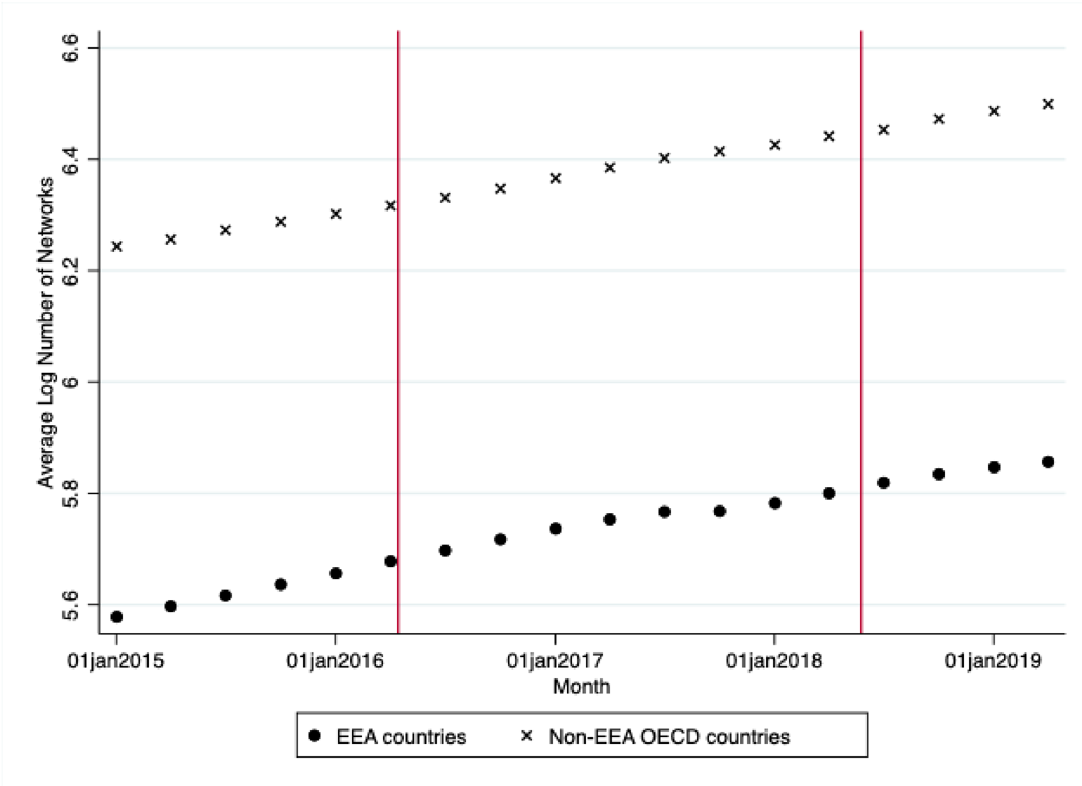


(c) Counterparties are non-EEA non-OECD countries



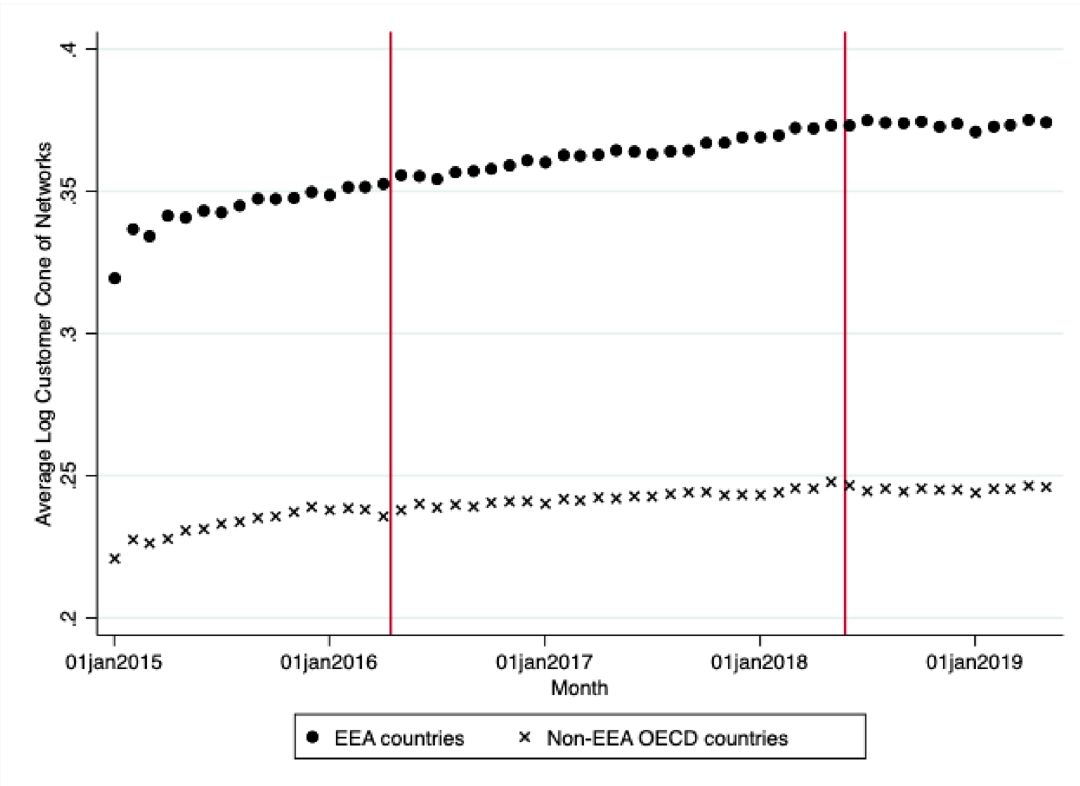
Notes: The dots represent $\overline{log(numAgIP_{i \in EEA\, jt})}$, the log number of IP-address-level interconnection points averaged among agreements by networks owned by EEA countries when the counterparties are networks owned by (a) EEA countries, (b) non-EEA OECD countries, (c) non-EEA non-OECD countries. The crosses represent $\overline{log(numAgIP_{i \in OECD \wedge \notin EEA\, jt})}$, the log number of IP-address-level interconnection points averaged among agreement by networks owned by non-EEA OECD countries when the counterparties are networks owned by (a) EEA countries, (b) non-EEA OECD countries, (c) non-EEA non-OECD countries. Only agreements present throughout Dec 2015 – June 2019 are used to take the averages.

Figure 7: Average log number of networks in EEA and non-EEA OECD countries



Notes: The dots represent $\overline{log(numNtwrk_{i \in EEA,t})}$, the log number of networks averaged among EEA countries. The crosses represent $\overline{log(numNtwrk_{i \in OECD \wedge i \notin EEA,t})}$, the log number of networks averaged among non-EEA OECD countries. Non-EEA and non-OECD countries' networks are not included in taking the averages. Regression including quarter and country fixed effects has the coefficient on $POST_e \times EEA = -.002$ ($se = 0.017$, clustered by country) and the coefficient on $POST_a \times EEA = 0.016$ ($se = 0.024$, clustered by country). Both are insignificant at conventional levels of significance.

Figure 8: Average log customer cone of networks in EEA and non-EEA OECD countries



Notes: The dots represent $\overline{log(NtwrkCustCone_{k \in EEA,t})}$, the log customer cone averaged among networks owned by EEA countries. The crosses represent $\overline{log(NtwrkCustCone_{k \in OECD \wedge i \notin EEA,t})}$, the log customer cone averaged among networks owned by non-EEA OECD countries. Non-EEA and non-OECD countries' networks are not included in taking the averages. Only networks present throughout Jan 2015 – June 2019 are used to take the averages. Regression including month and network fixed effects has the coefficient on $POST_e \times EEA = -.007^{*}$ ($se = 0.004$, clustered by country) and the coefficient on $POST_a \times EEA = 0.011^{***}$ ($se = 0.004$, clustered by country). Though both are significantly different from zero, their magnitudes are economically small.

Table 1: Description of variables

Variable	Unit of Observation	Frequency	Description	Source	Additional Notes
$numAgNtwrk_{kt}$	$ntwrk_k$	monthly	The number of interconnection agreements network k has.	AS Relationships	
$numAg_{ijt}$	$ctry_i$–$ctry_j$	monthly	The number of interconnection agreements between pairs of networks owned by the countries i and j.	AS Relationships	Agreements are available on network–network level and are aggregated to country–country level.
$numProvAg_{ijt}$	$ctry_i$–$ctry_j$	monthly	The number of interconnection agreements where country i's network is a provider to country j's network.	AS Relationships	Same as above.
$numPeerAg_{ijt}$	$ctry_i$–$ctry_j$	monthly	The number of interconnection agreements where country i's network and country j's network are peers.	AS Relationships	Same as above.
$numCustAg_{ijt}$	$ctry_i$–$ctry_j$	monthly	The number of interconnection agreements where country i's network is a customer of country j's network.	AS Relationships	Same as above. Value is identical to $numProvAg_{jit}$ for $ctry_j$–$ctry_i$.
$numAgIP_{klt}$	$ntwrk_k$–$ntwrk_l$	weekly	The number of IP-address-level interconnection points between network k and network l, given k and l have an agreement.	IPv4 Prefix-Probing	Data is available daily and is aggregated to weekly.
$numNtwrk_{it}$	$ctry_i$	quarterly	The number of networks country i owns.	AS Organizations	
$NtwrkCustCone_{kt}$	$ntwrk_k$	monthly	The number of networks network k can reach through its customer connections alone.	AS Relationships	A measure of a network's importance in Internet routing.

Notes: In the computer science field, a network is referred to as an *Autonomous System* (AS). All data sources listed here are available through CAIDA's webpage http://www.caida.org/data/overview/.

Table 2: Summary statistics

Variable	Observations	Mean	SD	Min	Max
Panel A: unrectangularized variables					
numAgNtwrk_{kt}	2,909,695	5.4	55.9	1	8,391
numAg_{ijt}	119,071	64.4	759.0	1	33,497
numProvAg_{ijt}	121,369	44.8	681.8	1	31,485
numPeerAg_{ijt}	62,241	30.8	140.1	1	4,155
numCustAg_{ijt}	121,369	44.8	681.8	1	31,485
numAgIP_{klt}	19,413,597	9.8	144.8	1	172,481
numNtwrk_{it}	3,597	357.3	1754.3	1	24,887
$\text{NtwrkCustCone}_{kt}$	2,909,695	7.8	263.3	1	37,061
Panel B: rectangularized variables					
numAg_{ijt}	1,085,400	7.1	252.2	0	33,497
numProvAg_{ijt}	2,160,000	2.5	161.9	0	31,485
numPeerAg_{ijt}	1,085,400	1.8	34.3	0	4,155
numCustAg_{ijt}	2,160,000	2.5	161.9	0	31,485

Notes: Panel A presents the variables with the appropriate levels of aggregation from the raw data. For numAg_{ijt}, numProvAg_{ijt}, numPeerAg_{ijt}, numCustAg_{ijt}, we also rectangularize the variables by filling in zero values for country pairs and dates with no observed agreements from our raw data and present the rectangularized variables in Panel B.

Table 3: The GDPR's impact on the number of agreements by networks

	(1)	(2)	(3)
$POST_e \times EEA$	-0.004	-0.002	
	(0.007)	(0.007)	
$POST_a \times EEA$	0.006		0.005
	(0.007)		(0.008)
Group dummies	networks	networks	networks
Time dummies	months	months	months
Clusters	43	43	43
R^2	0.933	0.933	0.933
Observations	1,275,236	1,275,236	1,275,236

Notes: The dependent variable is $log(numAgNtwrk_{kt})$. Only networks owned by EEA or OECD countries and present throughout Jan 2015 – June 2019 are used for regressions. All regressions include month dummies and network dummies. All regressions cluster standard error by country of ownership of network. Standard errors are in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 4: The GDPR's impact on the number of agreements by EEA and non-EEA OECD countries, by counterparty

	EEA	Non-EEA OECD	Non-EEA Non-OECD
	(1)	(2)	(3)
$POST_e \times EEA$	−0.003	−0.009	0.003
	(0.016)	(0.029)	(0.005)
$POST_a \times EEA$	0.011	−0.007	0.017^{***}
	(0.017)	(0.024)	(0.005)
Group dummies	country pairs	country pairs	country pairs
Time dummies	months	months	months
Clusters	880	418	6,751
R^2	0.987	0.991	0.948
Observations	47,520	22,572	364,554

Notes: The dependent variable is $log(numAg_{ijt}+1)$. The variable $numAg_{ijt}$ is rectangularized as described in Table 2 and we add one when we take the log to account for zero values. Column (1) includes observations when one party is a network owned by an EEA or non-EEA OECD country and the counterparty is a network owned by an EEA country. Column (2) includes observations when one party is a network owned by an EEA or non-EEA OECD country and the counterparty is a network owned by a non-EEA OECD country. Column (3) includes observations when one party is a network owned by an EEA or non-EEA OECD country and the counterparty is a network owned by a non-EEA non-OECD country. All regressions include month dummies and country pair dummies. All regressions cluster standard error by country pair. Standard errors are in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 5: The GDPR's impact on the number of agreements by EEA and non-EEA OECD countries, by counterparty and agreement type

	Counterparty is EEA			Counterparty is non-EEA OECD			Counterparty is non-EEA non-OECD		
	Provider	Peer	Customer	Provider	Peer	Customer	Provider	Peer	Customer
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$POST_e \times EEA$	−0.007	-0.038^{**}	0.005	−0.032	−0.040	−0.020	0.002	−0.007	-0.013^{***}
	(0.012)	(0.016)	(0.013)	(0.023)	(0.042)	(0.022)	(0.003)	(0.004)	(0.004)
$POST_a \times EEA$	0.027^{*}	0.002	−0.021	0.021	0.031	−0.025	0.011^{***}	0.007^{*}	-0.006^{**}
	(0.016)	(0.017)	(0.014)	(0.026)	(0.035)	(0.027)	(0.004)	(0.004)	(0.003)
Group dummies	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs
Time dummies	months	months	months	months	months	months	months	months	months
Clusters	1,376	880	1,376	473	418	473	6,751	6,751	6,751
R^2	0.978	0.980	0.977	0.984	0.984	0.985	0.941	0.930	0.925
Observations	74,304	47,520	74,304	25,542	22,572	25,542	364,554	364,554	364,554

Notes: The dependent variable is $log(numProvAg_{ijt} + 1)$ for columns (1), (4), (7), $log(numPeerAg_{ijt} + 1)$ for columns (2), (5), (8), and $log(numCustAg_{ijt} + 1)$ for columns (3), (6), (9). The dependent variables are rectangularized as described in Table 2 and we add one when we take the log to account for zero values. Columns (1), (2), (3) include observations when the treatment/control party is a network owned by an EEA/non-EEA OECD country and is the provider, peer, or customer to the counterparty network owned by an EEA country. Columns (4), (5), (6) include observations when the treatment/control party is a network owned by an EEA/non-EEA OECD country and is the provider, peer, or customer to the counterparty network owned by a non-EEA OECD country. Columns (7), (8), (9) include observations when the treatment/control party is a network owned by an EEA/non-EEA OECD country and is the provider, peer, or customer to the counterparty network owned by a non-EEA non-OECD country. All regressions include month dummies and country pair dummies. All regressions cluster standard error by country pair. Standard errors are in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Table 6: The GDPR's impact on the number of IP-address-level interconnection points per agreement by EEA and non-EEA OECD countries, by counterparty

	EEA	Non-EEA OECD	Non-EEA Non-OECD
	(1)	(2)	(3)
$POST_e \times EEA$	-0.032	0.039	0.003
	(0.024)	(0.023)	(0.049)
Group dummies	network pairs	network pairs	network pairs
Time dummies	weeks	weeks	weeks
Clusters	307	128	522
R^2	0.867	0.871	0.827
Observations	1,886,031	2,593,805	494,374

Notes: The dependent variable is $log(numAgIP_{ijt})$. Column (1) includes observations when one party of the agreement is a network owned by an EEA or non-EEA OECD country and the counterparty is a network owned by an EEA country. Column (2) includes observations when one party of the agreement is a network owned by an EEA or non-EEA OECD country and the counterparty is a network owned by a non-EEA OECD country. Column (3) includes observations when one party of the agreement is a network owned by an EEA or non-EEA OECD country and the counterparty is a network owned by a non-EEA non-OECD country. Only agreements present for at least 150 weeks are used. The GDPR approval date Apr 2016 is close to the sample starting date Dec 2015, so $POST_a \times EEA$ is not included in the regressions. All regressions include week dummies and network pair dummies. All regressions cluster standard error by country pair. Standard errors are in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

Data Appendix

In this section, we provide additional information about our data sources and data collection techniques. Our data comes from various data sources collected and compiled by the Center of Applied Internet Data Analysis (CAIDA) at the University of California, San Diego. Since 1998, CAIDA has been studying interconnectivity of the Internet by actively probing the Internet using its monitors placed at various vantage points around the world. Its current flagship active measurement infrastructure, Archipelago, collects interconnectivity data on the IP-address-level from more than 200 monitors located on 6 continents in over 60 countries. A list of current Archipelago monitor locations can be found at
https://www.caida.org/projects/ark/locations/.

CAIDA also collaborates with many organizations and compiles data collected from their monitors. Most notably, it collaborates with the Route Views Project at the University of Oregon and the Réseaux IP Européens Network Coordination Centre (RIPE NCC) in Europe to collect routing tables for network-level paths. A list of Route Views monitors can be found at http://www.routeviews.org/routeviews/index.php/collectors/. A list of RIPENCC monitors can be found at
https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data.

Moreover, CAIDA gathers records of network registration information from the world's five regional Internet registries (RIRs), allowing us to identify countries (or territories) of organizations that own individual networks. The dataset is available through the link:
https://www.caida.org/data/as-organizations/. The five RIRs are:

- The African Network Information Center (AFRINIC)
- The American Registry for Internet Numbers (ARIN)
- The Asia-Pacific Network Information Center (APNIC)
- The Latin America and Caribbean Network Information Center (LACNIC)
- The Réseaux IP Européens Network Coordination Centre (RIPE NCC)

Our main data on the network-level interconnection agreements comes from the routing tables, while our IP-address-level interconnection points for each agreement come from the active probes. The data extraction process is explained in the Data section in the main text.

A number of key variables in this study come from a dataset called *AS Relationships*, as in the computer science field, an independently operated network connected to the Internet is referred to as an Autonomous System (AS). This dataset is available through the link
http://www.caida.org/data/as-relationships/.

To construct the AS Relationships dataset, CAIDA collects BGP tables from its partner monitors placed at various vantage points across the Internet and peered directly with networks' BGP routers, typically major ones with large numbers of routes stored, at Internet exchange points. Network-to-network connection agreements are then extracted from routing paths announced in these BGP tables. Then the agreements are annotated with inferred agreement types. The inference algorithm draws from Gao (2001), Subramanian et al. (2002), Di Battista et al. (2003), Erlebach et al. (2002), Xia and Gao (2004), Dimitropoulos et al. (2007a) and Dimitropoulos et al. (2007b).

Our IP-address-level interconnection points within each agreement come from the dataset *IPv4 Prefix-Probing*. This dataset is available through the link https://www.caida.org/data/active/ipv4_prefix_probing_dataset.xml.

To keep visibility consistent throughout our sample periods, we extract agreements only from a set of monitors that operated throughout our sample periods, January 2015–June 2019 for AS Relationships and December 2015–June 2019 for IPv4 Prefix-Probing. Moreover, we dropped all of the affected interconnections due to configuration changes in three RIPENCC monitors in October 2018. To make these sample restrictions, we use nonpublic versions of the datasets which include monitor identifiers for each observation of interconnection.

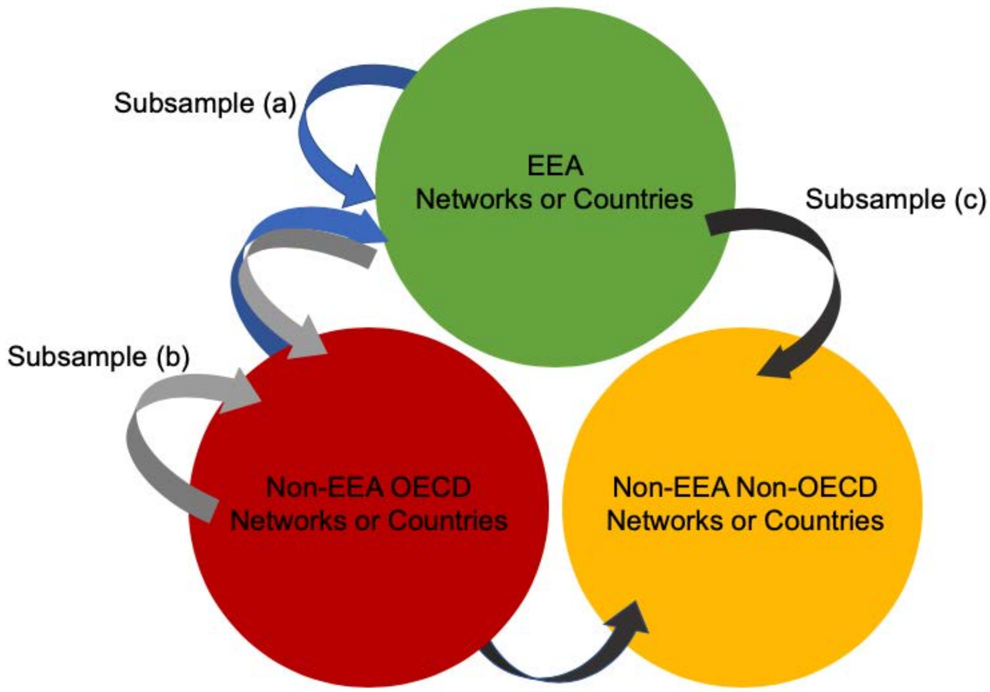
We drop networks owned by a number of small island countries, Andorra, Central African Republic, Eritrea, North Korea and Vatican City from our sample due to these countries' very small overall number of connections with the rest of the Internet. Our EEA subsample includes networks owned by organizations headquartered in the 31 EEA member countries as well as networks owned by EU-wide organizations. For networks owned by EU-wide organizations, their countries of origin are shown as "EU" in network registration records. We include these networks in the EEA subsample for the purpose of our empirical analysis. The resulting total number of countries and territories in our sample is 200. A complete list of countries and territories in our sample is presented in Table A1.

Table A1: List of Countries and Territories

EEA	Non-EEA OECD	Non-EEA Non-OECD				
Austria	Australia	Afghanistan	Congo	India	Mozambique	Somalia
Belgium	Canada	Albania	Congo (DR)	Indonesia	Myanmar	South Africa
Bulgaria	Chile	Algeria	Costa Rica	Iran	Namibia	Sri Lanka
Croatia	Israel	Angola	Cuba	Iraq	Nepal	Sudan
Cyprus	Japan	Argentina	Djibouti	Isle of Man	New Caledonia	Suriname
Czech Rep.	Korea (Rep.)	Armenia	Dominica	Jamaica	Nicaragua	Swaziland
Denmark	Mexico	Aruba	Dominican Rep.	Jersey	Niger	Syria
Estonia	New Zealand	Azerbaijan	Ecuador	Jordan	Nigeria	Taiwan
Finland	Switzerland	Bahamas	Egypt	Kazakhstan	Oman	Tajikistan
France	Turkey	Bahrain	El Salvador	Kenya	Pakistan	Tanzania
Germany	USA	Bangladesh	Eq. Guinea	Kuwait	Palestine, State of	Thailand
Greece		Barbados	Ethiopia	Kyrgyzstan	Panama	Timor-Leste
Hungary		Belarus	Faroe Is.	Laos	Papua New Guinea	Togo
Iceland		Belize	Fiji	Lebanon	Paraguay	Tonga
Ireland		Benin	Fr. Guiana	Lesotho	Peru	Trinidad and Tobago
Italy		Bermuda	Fr. Polynesia	Liberia	Philippines	Tunisia
Latvia		Bhutan	Gabon	Libya	Puerto Rico	Turkmenistan
Liechtenstein		Bolivia	Gambia	Macao	Qatar	Uganda
Lithuania		BiH	Georgia	Macedonia	Reunion	Ukraine
Luxembourg		Botswana	Ghana	Madagascar	Russia	UAE
Malta		Brazil	Gibraltar	Malawi	Rwanda	Uruguay
Netherlands		Brunei	Greenland	Malaysia	St. Kitts and Nevis	Uzbekistan
Norway		Burkina Faso	Grenada	Maldives	St. Lucia	Vanuatu
Poland		Burundi	Guadeloupe	Mali	St. Vincent	Venezuela
Portugal		Côte d'Ivoire	Guam	Marshall Is.	San Marino	Viet Nam
Romania		Cabo Verde	Guatemala	Mauritius	Saudi Arabia	Virgin Is.
Slovakia		Cambodia	Guernsey	Micronesia	Senegal	Yemen
Slovenia		Cameroon	Guinea	Moldova	Serbia	Zambia
Spain		Cayman Is.	Guyana	Monaco	Seychelles	Zimbabwe
Sweden		Chad	Haiti	Mongolia	Sierra Leone	
UK		China	Honduras	Montenegro	Singapore	
EU		Colombia	Hong Kong	Morocco	Solomon Is.	

Note: The list of EEA countries and territories includes the 31 EEA member countries and the EU, as some networks are owned by EU-wide organizations. The list of non-EEA OECD countries and territories includes 11 countries. The list of non-EEA non-OECD countries and territories includes 157 countries and territories. There were no changes to EEA or OECD membership status during our study period.

Figure A1: Three subsamples for the analysis on the network pair or country pair level



Notes: Subsample (a) fixes EEA networks or countries as interconnection counterparties. Subsample (b) fixes non-EEA OECD networks or countries as interconnection counterparties. Subsample (c) fixes non-EEA non-OECD networks or countries as interconnection counterparties. Interconnections between EEA countries and non-EEA OECD countries contribute to both subsample (a) and subsample (b).