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, Revised April 2020

The authors thank Tim Bresnahan, Roderick Fanou, Samuel Goldberg, Avi Goldfarb, Ginger Zhe Jin, Garrett Johnson, Stephen Strowes, the editors and an anonymous referee for helpful suggestions. We thank Dan Andersen for technical assistance. 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, Revised April 2020
JEL No. L00,L51,L86

ABSTRACT

The Internet comprises thousands of independently operated networks, interconnected using bilaterally negotiated data exchange agreements. The European Union (EU)'s General Data Protection Regulation (GDPR) imposes strict restrictions on handling of personal data of European Economic Area (EEA) residents. A close examination of the text of the law suggests significant cost to application firms. Available empirical evidence confirms reduction in data usage in the EEA relative to other markets. We investigate whether this decline in derived demand for data exchange impacts investment in interconnection by networks in the EEA relative to networks in non-EEA OECD countries. Our data consists of a large sample of interconnection agreements between networks globally in 2015–2019. All evidence estimates precisely zero effects: the number of observed agreements, 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 there is no visible short run consequence of the GDPR at the internet layer.

Ran Zhuo
Department of Economics
Harvard University
Cambridge, MA 02138
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
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
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 network operators voluntarily exchange data via bilaterally-negotiated agreements (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] The European Union (EU)'s General Data Protection Regulation (GDPR) serves as a landmark privacy law, regulating the collection, processing, and transfers of consumers' personal data that occur along with these transactions. Since the GDPR's approval in April 2016 and implementation in May 2018, it has inspired a wave of privacy regulation in countries such as Brazil, India, Japan, South Korean and the US (Goldberg et al. (2019)). The unprecedented scale and scope of the GDPR makes it the most important privacy policy since the commercialization of the internet in the 1990s and many hypothesize it would fundamentally change the internet's operation and the digital economy.

This paper investigates whether the GDPR affected investment in the interconnection and growth of the Internet at the internet layer. Consider this layer as analogous to the postal network. When consumers and providers of online content and services send each other "letters" containing digital data, networks ("post offices") deliver the mail.[2] The GDPR restricts how and where content and service providers can collect, store, share, and monetize personal information contained in the "mail," bringing increased cost and complexity for these application firms, which

[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. We thank the editors for suggesting this simple and insightful analogy as an accessible introduction to the internet layer.

may in turn impact the demand for "mail". We investigate whether the networks ("post offices") reduced investments in their interconnection post-GDPR in response to the decline in the demand for "mail" services.

A growing literature measures the impact of privacy regulations on investment in applications, while little research studies how such laws affect the internet layer. We hypothesize that the GDPR shapes investment in applications by influencing user participation.[3] Participation reacts to rules limiting the collection, use, storage, and disposal of personal data, and limiting the resale and (re)disclosure of user data.[4] If visitors value these privacy protections, then the GDPR may generate visits to online sites complying with the GDPR, and those visitors may engage more with the sites.[5] The GDPR simultaneously reduces the value of those visits because it lowers the effectiveness of targeted advertising and targeted sales.[6] Thus, it could decrease the ability of firms to monetize user participation and engagement. These effects operate in opposite directions.

We also hypothesize that the GDPR raises the operational costs of an online business that collects personal information.[7] The prospective enforcement of rules also raises the expectations of fines and ongoing negotiation, and these prospective costs apply to firms with the largest online presence. The GDPR also generates additional costs associated with uncertainty about how enforcement operates. The increase in operational costs reduces investment in applications. It also leads to lower valuations for entrepreneurial startups in online commerce.

We sort through these hypotheses with limited available evidence. Some evidence suggests traf-

[3]Miller and Tucker (2011, 2018) postulate a similar trade-off between participation and the costs of supplying services in the context of medical services.

[4]We explain the legal details of the GDPR in Section 3 with a close examination of the text of the law. We thank an anonymous referee for suggesting use this approach to tighten our discussion.

[5]Empirical evidence on this hypothesis, however, is extremely lacking. Moreover, many provisions of the GDPR are motivated by views that these are intrinsic rights, and do not account for their consequences for online commerce. See Hoofnagle et al. (2019)).

[6]This is similar to Goldfarb and Tucker (2011). Goldberg et al. (2019) and Aridor et al. (2020) also hypothesize this effect. We further discuss the cost associated with behavioral changes of content and service providers in Section 3.

[7]We offer a detailed discussion about the various requirements for compliance and their cost in Section 3.

fic diminished after the implementation of the GDPR.[8] Some evidence also suggests the negative effects on investment outweigh the positive, leading to lower overall investment in applications.[9] We offer a more thorough discussion on the costs and benefits of the GDPR to consumers and providers of online content and services in Section 3 with a close examination of the text of the law, sorting the available academic literature into the taxonomy of costs and benefits.

Most relevant to the question we study, we further expect a decline in applications to lead to a decline in traffic and connectivity at the internet layer. This effect operates through changing the derived demand for data exchange between networks when the supply of applications falls. A simple bilateral bargaining model between networks, such as one in Besen et al. (2001), formalizes this intuition.[10]

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.

We then dive into the empirical analysis. 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

[8] Goldberg et al. (2019) find some evidence of a decline in traffic at existing firms, while Johnson & Shriver (2019) and Peukert et al. (2020) find evidence of a shift in traffic to the largest firms.

[9] This evidence points towards a reduction in entrepreneurial ventures and market share of smaller firms in Jia, Jin and Wagman (2019, 2020).

[10] We present the derivations and comparative statics in Appendix A for interested readers.

number of networks that are customers to a given networks. By combing the topology we can infer the number of interconnection points between pairs of networks with interconnection agreements on the level of IP addresses, the numerical labels assigned to unique devices connected to the Internet. We collect the datasets used in this paper 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/.[11] We explain the collection of the data and the construction of the variables in Section 4 and provide additional details in Appendix B.

Using this data, we present descriptives which 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,[12] though the levels of interconnectedness differ.

We treat the GDPR's April 2016 approval and May 2018 enforcement as two cutoff dates for periods post policy treatment. We offer several reasons for this assumption and discuss them in detail in Section 5.

We then use a difference-in-differences approach, contrasting interconnection activities by networks owned by organizations headquartered in the EEA (treatment group) and networks owned by organizations headquartered in other countries (control group) before and after the approval and implementation of the policy. Given the wide territorial scope of the GDPR, we find it important to discuss whether it is ever possible to have a reasonable control group.[13] We motivate our choice of control group in several different ways and discuss them thoroughly in Section 5. We also acknowledge the limitations of our empirical approach in the same section.

Contrasting changes in EEA networks' interconnection behavior before and after April 2016

[11] For more information about the data sources used in this paper, please see the data appendix (Appendix B).
[12] Please see Appendix Table B1 for lists of EEA countries and non-EEA OECD countries.
[13] We are very grateful to the editors and an anonymous referee for highlighting this key issue.

and May 2018 relative to non-EEA 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. Networks' affiliation with the EEA also does not affect the observed numbers of IP-address-level interconnection points between each pair of interconnecting networks. We also find economically small effects of the GDPR on the 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 EEA networks' interconnecting behavior, rejecting the hypothesis that the negative impact of the GDPR at the application layer has thus far propagated to the internet layer. In Section 6, we present these results. We discuss several robustness checks to our main results in Section 6.7.

Our paper has an obvious policy implication: 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 conclusion section of the paper, we discuss a number of possible reasons for this result.

Our paper also contributes by presenting data of unprecedented scale and scope. While a theoretical literature tackles questions 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 represents 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. We think future works should keep tackling the issue of unmet data needs.

We organize the rest of the paper as follows. Section 2 provides the background in network interconnection. Section 3 provides the background in the GDPR. Section 4 describes the data. Section 5 explains justifications for our empirical strategy. Section 6 presents results across a number of measures of the impact of the GDPR on interconnection. Section 7 concludes.

2. Internet Interconnection

In Section 1, we use an analogy to the postal network to introduce the internet layer. We offer a more precise definition in this section. The section explains the technicalities associated with the four layers of the internet, including the internet layer, and describes the demanders and suppliers associated with each layer and the flow of payment. We also discuss the contractual and institutional foundations behind interconnection, and argue networks can respond to policy changes by quickly changing the number or specifications of interconnection agreements.

The Internet was designed with four layers of data exchange in mind: application, transport, internet, and link.[14] Each layer uses a specific set of protocols, shared state, and provides a con-

[14]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.

nection 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 exchange takes place between and across each layer.

As shown in Figure 1, a consumer's personal computer (PC) or smart phone has applications like web browsers, webpages and gaming platforms working at the application layer, and an operating system handling the transport, internet, and link layers. Consumers use their applications to connect, using lower layers, to other applications hosted on other devices remotely. The application layer is the layer where personal data is most relevant. Consumers may pay service and content providers directly, and/or provide their engagement and personal information to these application firms who resell it on to advertisers targeting content, services, and ads. Metrics of engagement include ad views, ad clicks, and purchases resulting from referrals.

When applications connect, data exchange happens between the consumer and content/service providers. Application layer communication relies on lower layers of internet infrastructure and communication protocols. The transport layer makes 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 maintains global routing state, routing data packets to their destination address by selecting the next closest router. At this layer, the Internet can be conceptualized as a collection of different networks, each with its own set of routers and routing policies. Routers connect multiple networks and forward data packets destined either for their own networks or other networks. In Figure 1, the internet layer is visualized to facilitate moving data from the consumer's network to intermediary ISPs (transport networks A and B) then to the service/content network, and the service/content network may send data back the same route. Below the internet layer layer, the link layer forwards data packets to immediately

adjacent (the "next hop") routers.

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. Networks can be operated by consumers' and service/content firms' Internet Service Providers (ISPs), or service/content firms themselves. Consumers and service/content firms pay ISPs to connect their networks to each other. ISPs in turn pay each other where necessary to complete or enhance reachability to the rest of the Internet.

Network operators typically use a mix of agreements with different interconnection counterparties. As described by the Internet Society (The Internet Society, 2015), we can broadly classify these agreements as 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 customer network 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 network, lowering the average cost of traffic delivery. If the peers have similar negotiating power, they form a settlement-free agreement. Under an imbalance, the weaker network pays the other under a paid peering agreement.

Transition provider networks typically price p2c and c2p agreements as a metered service outside of the residential market on a per-megabit-per-second (Mbps) basis. Transit providers compete vigorously, resulting in a strong declining trend for the prices for transition from 1998 until

present.[15] 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, networks usually renegotiate even multiple-year agreements yearly. Other than the potential legal cost of breaking a multiple-year agreement, customers incur little cost to switch to a different provider for substantial potential gains of bringing the unit transit prices in line with the current market price. Networks also commonly 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. Potential peers typically negotiate p2p agreements on a case-by-case basis. Traffic volume often represents "a key determinant" of 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, it incurs additional cost 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.[16]

[15]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).

[16]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 two networks form an interconnection agreement, 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 networks can establish or terminate interconnections reasonably quickly in response to policy changes.

3 The GDPR

In this section, we provide an overview of the GDPR. We closely examine the text of the law, specifically on the key definitions, the responsibilities of application firms, the rights of consumers, and the mechanisms of enforcement. We discuss where networks fit into the regulatory framework and how the law may impact them, sorting empirical evidence into a taxonomy of potential costs and benefitss of the law. Combining the legal provisions and the empirical evidence with a discussion on the reaction of the popular media to the regulation, we find it easy to hypothesize significant negative impact of the regulation on networks' investment, with networks located in the EEA harder hit compared to their counterparts in other countries. We need sound empirical

evidence to support or refute this hypothesis.

Approved on April 14, 2016 and effective on May 25, 2018, the GDPR applies to most application firms and networks with EEA end customers because it applies to any organization that "processes personal data" of EEA consumers.[17] The GDPR defines "personal data" broadly, as any information that might identify a consumer ("data subject").[18] It also defines "processing" broadly, as any operation that is performed on personal data, whether or not by automated means.[19] The GDPR places the burden of responsibilities on organizations that determine the purposes and the means of processing of personal data (data "controllers"), while organizations that process personal data on behalf of controllers (data "processors") also have to comply with a considerable portion of the GDPR.[20] Many application firms fall within the meaning of the GDPR as both data controllers and processors, while networks, routing data on behalf of application firms, fall within the meaning of the GDPR as data processors. The GDPR has a wide territorial scope. Even when a firm has no physical presence in the EEA, the GDPR applies if it is apparent that the firm envisages offering services to consumers located in the EEA.[21]

The GDPR may generate significant compliance cost for application firms as it specifies stringent responsibilities of data controllers (and sometimes data processors). Firms need to fulfill major obligations such as keeping detailed, account-like records of their processing activities,[22] incorporating protection into the technical design for the services with data protection "by design" and "by default,"[23] (for organizations that process personal data on a large scale or process sen-

[17]The GDPR was incorporated into the EEA Agreement on July 6, 2018, so its scope covers both EU member states and non-EU EEA member states. The nationality of the consumer is not relevant, the relevant criterion is whether the person is located in the EEA (Hoofnagle et al. (2019)).

[18]GDPR Art. 4(1). Under this definition, not only a person's name and physical addresses, but also IP addresses, cookies, and similar data are personal data.

[19]GDPR Art. 4(2).

[20]GDPR Art. 4(7), (8). In principle, if data processors violate the GDPR, the data controller will be considered responsible and liable (Hoofnagle et al. (2019)).

[21]GDPR Recital 23.

[22]GDPR Art. 30.

[23]GDPR Art. 25(1), (2). Data protection "by design" refers to measures such as pseudonymisation, which are

sitive personal data) appointing a fully independent Data Protection Officer (DPO),[24] developing "Data Protection Impact Assessments" (DPIA) for high-risk processing activities,[25] and notifying the Data Protection Authority of a personal data breach within 72 hours after its discovery[26] (Hoofnagle et al. (2019)). Ernest & Young predicts the world's 500 largest companies would spend $ 7.8 billion to comply with the GDPR.[27] While a report by DataGrail, a privacy management platform, estimates 74% of small- and mid-sized organizations would spend more than $100,000 and 20% of them would spent more than $1 million.[28]

The GDPR may also generate costly behavioral changes to application firms, hampering their ability to monetize user participation and engagement, due to the law's legal basis for processing personal data and many high level principles. Under the GDPR, consent must be freely given, specific, informed, unambiguous, and revocable[29] (Hoofnagle et al. (2019)), preventing firms from using long and inaccessible consent processes to obtain personal data. The purpose limitation principle specifies that personal data shall be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes."[30] This would limit application firms' ability to repurpose data in unanticipated ways. The data minimization principle specifies that personal data should be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed" and shall be "kept in a form which

designed to implement data-protection principles. Data protection "by default" means only personal data which are necessary for each specific purpose of the processing are processed and personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

[24]GDPR Art. 37-39. According to GDPR Art. 37(1)(b), companies involved in internet marketing will have to have DPOs as they are deemed to process personal data for "systematic monitoring of data subjects on a large scale" (Hoofnagle et al. (2019)).

[25]GDPR Art. 35(1). High-risk processing activities would include automated processing or profiling that leads to decisions that significantly affect people and sensitive data are processed on a large scale (Hoofnagle et al. (2019)).

[26]GDPR Art. 34.

[27]Kahn, Jeremy, Stephanie Bodoni & Stefan Nicola. 2018. "It'll Cost Billions for Companies to Comply With Europe's New Data Law." Bloomberg Businessweek.

[28]Lindsey, Nicole. 2019. "Understanding the GDPR Cost of Continuous Compliance." CPO Magazine.

[29]GDPR Art 7(3).

[30]GDPR Art. 5(1)(b).

permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed."[31] This accounting comes with increased cost and complexity, and by intent reduces window for monetization.

Available evidence from the empirical literature, though limited, supports the hypothesis that the GDPR generates behavioral changes costly to application firms. Godinho de Matos & Adjerid (2019) studied the effectiveness of a campaign for obtaining GDPR-compliant consent for personal marketing and found such practices effective, though at additional cost to the firm to elicit such consent. Goldberg et al. (2019) found a 10% decrease in recorded e-commerce sales for a sample of EU firms after the GDPR's enforcement. Johnson & Shriver (2019) found that the week after the GDPR's enforcement, website use of web technology vendor 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%. Peukert et al. (2020) found similar effects and the magnitude of change was particularly large for websites with EU-specific top-level domains.[32] Aridor et al. (2020) found a 12.5% drop in observable consumers to a data analytics intermediary after the GDPR's enforcement and that resulted in declines in revenue from targeted ads for European travel platforms compared to their non-European counterparts.[33]

Moreover, under the GDPR, application firms face prospective cost associated with fines and on-going negotiation, as well as cost associated with uncertainty about how enforcement operates. Less serious violations can trigger administrative fines up to ten million euros or up to 2% of the

[31] GDPR Art. 5(1)(c), (e).

[32] A top-level domain is the last segment of a domain name. Common top-level domains include .com, .org and .us.

[33] Additional empirical evidence suggests content and service providers indeed alter their behavior significantly following the implementation of the GDPR. Libert, Graves & Nielsen (2018) found the GDPR has led to a 22% decrease in third-party cookies on a set of EU news sites (third-party cookies are information stored in browsers used for tracking and advertising, sent from sites other than the one the user is currently visiting). Degeling et al. (2018) and Mohan et al. (2019) found extensive updates to websites' and cloud services' privacy policies. 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.

total worldwide annual turnover of the preceding financial year, whichever is higher, while more serious violations, such as violations to the purpose limitation principle or data minimization principle, can trigger administrative fines of up to 20 million euros or up to 4% of the total worldwide annual turnover.[34] The GDPR gives Data Protection Authorities broad powers to investigate, intervene and bring legal proceedings. Hoofnagle et al. (2019) stress the presumption of on-going communications between the Data Protection Authorities and data controllers over the unresolved features of the rules. On-going negotiation generates cost especially for firms with the largest online presence. Topics with on-going negotiation encompass questions about requirements to store data locally, the precise rules of reselling data, and so on. Hoofnagle et al. (2019) also argue that the rules, especially those in the recitals, were written with open-ended features to provide regulators with the flexibility to respond to unexpected and unanticipated issues. This brings additional cost to application firms due to uncertainty about how enforcement operates.

In addition, the GDPR may impact the derived demand for data flow between networks by restricting international transfers of data. To prevent data controllers from simply moving personal data to a "data haven" with fewer or no restrictions, the GDPR only allows transfers of personal data outside the EEA when the destination country or organization upholds privacy protection to a comparable level of that specified in the GDPR.[35] Application firms either bear the significant cost of achieving GDPR-level personal data protection even outside the EEA or choose to reduce the amount of data they transfer outside the EEA.

We expect the costs associated with the GDPR to hit harder on application firms located within the EEA than application firms located outside the EEA. The GDPR makes all EEA consumers more costly to serve, but a non-EEA application firm faces different costs for non-EEA consumers,

[34] GDPR Art. 83(4), (5).

[35] GDPR Art. 45-47. Adequacy status are evaluated by the European Commission. The US in general does not achieve adequacy. US-based firms may choose to participate the EU-US Privacy Shield, which requires firms to commit to a GDPR-like level of protection.

and may not comply with the GDPR for their non-EEA consumers to lower their cost outside of the EEA. Non-EEA application firms may also choose not to comply, cutting out EEA consumers all together. We discuss in more detail the various evidence for noncompliance of non-EEA application firms and for differential impact of the GDPR at the application layer in Section 5, as support for our empirical strategy.

As opposed to the cost to application firms, consumers receive many data-related rights under the GDPR, which may boost their participation. The GDPR specifies seven rights for consumers: the right 1) to access, 2) to data portability, 3) to rectify data, 4) to stop processing, 5) to object, 6) to erase data, and 7) to resist profiling and computerized decision making processes[36] (Hoofnagle et al. (2019)). If consumers value these privacy protections, then the GDPR may generate more use of the content and services that comply with the law. Empirical evidence on this, however, is lacking. To the best of our knowledge, we are not aware of academic works that have studied whether consumers actually increase their demand for online content or services in response to better privacy protection.[37] In contrast, Goldberg et al. (2019) found a large and significant 10% decline of recorded page views, visits and orders for a set of EU e-copmmerce firms after the GDPR became effective, suggesting the possibility of less user engagement with less personalized ad targeting and recommendations.

Given the direct compliance cost, the indirect cost due to behavioral changes, prospective cost and cost due to uncertainty, the GDPR may reduce application firms' revenue, chance of survival, and investment into new products. Adding together the apparent lack of demand response from consumers to better privacy protection, one may expect the negative effects on investment to out-

[36]GDPR Art. 4(3), 8, 16, 17, 20, 21(1), 22.

[37]The only available evidence are anecdotal, which suggest consumers feel no better off under the GDPR. See for examples:

Olenick, Doug. 2019. "Consumers Feel Privacy is No Safer under GDPR." SC Media.

Tesseras, Lucy. 2018. "GDPR Three Months On: Most Consumers Feel no Better Off." MarketingWeek.

weigh the positive, leading to lower overall investment in applications. The effect on investment may be especially pronounced for application firms located in the EEA.[38] Empirical evidence, though limited, supports this notion. Jia, Jin & Wagman (2019, 2020) show that 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.

Given the reasons outlined above, it is easy to hypothesize that the overall negative effect on investment at the application layer may propagate to the internet layer by decreasing the derived demand for data flow between networks, with EEA networks harder hit. This hypothesis coincides with the alarmist and negative picture the popular media and numerous public "White Papers" paint for the broad impact of the regulation.[39] 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, though views expressed in the popular media are neither supported by systematic data collection, nor informed by a census of experience, and most of them stress the costs in unspecific terms. We need sound empirical works to support or refute the hypothesized impact and the uninformed speculation in the media.

[38] We discuss in greater detail the extent of non-compliance of non-EEA application firms and evidence for differential impact in Section 5 where we provide justifications for our choice of control group.

[39] A sampling from (the most credible) news sources gives a good sense of the range of concerns voiced at the time GDPR became binding. See for examples:

Bershidsky, Leonid. 2018. "Europe's Privacy Rules Are Having Unintended Consequences." Bloomberg.

Cool, Alison. 2018. "Europe's Data Protection Law Is a Big, Confusing Mess." New York Times.

Downes, Larry. 2018. "GDPR and the End of the Internet's Grand Bargain." Harvard Business Review.

Hern, Alex. 2018. "Facebook Moves 1.5bn Users out of Reach of New European Privacy Law." Guardian.

Kostov, Nick & Sam Schechner. 2019. "GDPR Has Been a Boon for Google and Facebook." Wall Street Journal.

Satariano, Adam. 2018. "G.D.P.R., a New Privacy Law, Makes Europe World's Leading Tech Watchdog." New York Times.

Trentmann, Nina. 2018. "Companies Worry That Spending on GDPR May Not Be Over." Wall Street Journal.

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 we collect data 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.[40]

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 (Appendix B).

As shown in Table 1, a number of our key variables come from a dataset called *AS Relation-*

[40]We present a complete list of countries and territories in our sample in Appendix Table B1.

ships. 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. (2005) and Dimitropoulos et al. (2007). It achieved over 98% accuracy of agreement type inference via direct validation with a set of network operators (Luckie, et al., 2013). The algorithm succeeded in inferring 96% of the agreement types in our sample.

We compute the AS Relationships dataset 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 In-

ternet 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 (a prefix is a block of IP addresses) 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.[41] 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 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

[41] 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%.

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., 2014). We do not know how much traffic exchange happens 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. Empirical Strategy

In this section, we provide justifications for key assumptions in our empirical strategy. Our empirical strategy, in short, constitutes using a difference-in-differences approach that compares interconnection activities by networks owned by organizations headquartered in the EEA (treatment group) and networks owned by organizations headquartered in non-EEA OECD countries (control group) before and after the GDPR approval date (April 2016) and implementation date (May 2018). The validity of our approach hinges on the validity of our assumptions that any potential policy effect did not set in before the approval of the GDPR and that our control group is reasonable. We therefore focus our discussion on the justifications for these assumptions. We also

acknowledge the limitations of our empirical approach in this section.

We offer several justifications for our assumption about the timing of the effect: 1) a robustness check using an alternative cutoff date, 2) conversations with network operators on their decision horizon, and 3) empirical evidence at the application layer supporting stark cutoff dates.

We first discuss a robustness check that uses December 2015 as an alternative cutoff date to study whether networks responded to the law prior to its approval. Examining the timeline of the creation of the law, we think December 2015 is the earliest possible date for firms to respond to the future law. Consultation for the law began as early as 2009 and the European Commission published a proposal text in 2012. In 2013, the European Parliament adopted a compromised text, based on almost 4,000 proposed amendments. In 2015, the Council of the European Union published its proposal for the GDPR and started negotiations with the European Parliament. The Parliament and Council reached agreement on the text of the GDPR in December 2015 (Hoofnagle et al. (2019)). Given the intensity of negotiation and the amount of changes the proposal went through, we think it was unlikely for firms to respond before the text of the law was fixed. Most of our variables are available well before December 2015, allowing us to use December 2015 as an alternative cutoff date and test whether networks responded in anticipation of the law. We discuss results from this robustness check in our results section.

Moreover, through conversations with network operators, we learn network operators respond to real-time changes in actual data flows at the application layer, rather than respond to potential changes on the longer time horizon. This is due to the fact that networks can establish and terminate interconnection agreements relatively quickly and networks have incentives to plan in very short terms given the ever decreasing per Gbps cost of data exchange, which we have discussed in Section 2. As networks respond to changes in derived demand for data exchange at the application layer, we think they were unlikely to respond before changes happened at the application layer.

Empirical evidence at the application layer supports the notion that changes at the application layer happened after the GDPR effective date of May 2018. Jia, Jin, & Wagman (2019) discuss that, within the two years between the GDPR's approval and effective dates, many organizations chose to roll out their compliance strategy only days and weeks before the effective date. Goldberg et al. (2019) show large declines in page views, visits, orders and revenue from EU consumers at a set of e-commerce sites relative to a control group within four weeks after the policy implementation date.[42] Johnson & Shriver (2019) show a cliff-like decline in websites' use of web technology within thirty days after the policy implementation date.[43] Aridor et al. (2020) similarly show an immediate effect on consumer opt-out behavior and firm revenue.[44] We think such evidence of large and quick responses at the application layer helps to justify our choice of stark cutoff dates as well.

Now we move to discuss our choice of control group. Our treatment group consists of networks owned by organizations headquartered in EEA countries. We choose networks owned by organizations headquartered in non-EEA OECD countries as the control group. We think 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 results section, the parallel pre-trends needed for the difference-in-differences approach are visually apparent for the treatment and control groups across all outcome variables of interest. We exclude networks in non-EEA non-OECD countries and territories from the control group for worries that networks in developing countries might behave differently from networks in more developed countries prior to the GDPR. Appendix Table B1 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

[42] See their Figure 2 for details.
[43] See their Figure 1 for details.
[44] See their Figures 3–6 for details.

EEA nor in the OECD (excluded).

We are well aware of the concern that, given the GDPR's global ambition and wide territorial scope, application firms in non-EEA OECD countries also need to incur substantial cost to comply with the law if they want to serve EEA consumers. This would in turn change the derived demand for data exchange associated with non-EEA OECD networks and bias our results towards zero given our choice of control group. We discuss thoroughly how we address this concern in five ways: 1) a robustness check using a first differences approach, 2) a discussion of the extent of compliance of non-EEA application firms, 3) a reorganization of our results based on our confidence of the validity of the control group across our eight outcome measures and various subsample breakdowns, 4) interpretation of our results as differential impact and empirical evidence of differential impact at the application layer, and 5) a straightforward acknowledgement of problems with our control group for some of our outcome measures and subsamples.

First, we note that we can perform a robustness check to our main difference-in-differences approach by simply first differencing our outcome variables within EEA subsamples and studying whether the approval or the implementation of the GDPR impacted the rate of interconnection growth within EEA countries or networks. As we will show in a series of figures in our results section, the growth of interconnection within EEA itself was remarkably smooth and steady. This observation motivates us to test first differences as an additional piece of evidence corroborating our main difference-in-differences estimates. We discuss results from this robustness check in our results section.

We then discuss the extent of compliance to the GDPR among non-EEA application firms. We hypothesize that non-EEA consumers are unlikely to enjoy similar protection as EEA consumers following the GDPR, even if these firms choose to comply. These firms may also choose not to comply, cutting out EEA consumers all together. We note that, as compliance to the GDPR can

be extremely costly (see Section 3 for a detailed discussion), there is incentive for non-EEA application firms to limit compliance to EEA consumers. While some non-EEA application firms allegedly improved privacy protection for non-EEA consumers following the GDPR, the degree of protection non-EEA consumers enjoyed fell far short from their EEA counterparts.[45] In fact, the California Consumer Privacy Act (CCPA) was strongly motivated by the goal to bring privacy protection of Californian residents on par with that of EEA residents under the GDPR as firms did not voluntarily do so. Peukert et al. (2020) suggest websites catering to non-EU audiences decreased their use of third-party web technology vendors following the GDPR. However, they found the magnitude of change for those websites at 2.2–3.6 percentage points in their most reliable specifications to be far smaller than the magnitude of change for websites catering to EU audience at 7.1–9.3 percentage points.[46] A lot of anecdotal evidence also suggests many content and service providers located outside the EEA simply stopped serving EEA consumers.[47] When EEA consumers were not blocked, they could be offered a very stripped down version of the content.[48]

Our above discussion helps us to identify the outcomes and subsamples that are less likely to be plagued by the bias towards zero across our eight different outcome measures and various subsample breakdowns. We present our results in Section 6 with this consideration, showing first

[45]The Associated Press investigated Facebook's claim on global GDPR compliance and found its implementation of many GDPR provisions were vague for non-EEA consumers. While Facebook did not publicize the fact, the Associated Press also found users in six Asian countries did not get the protection through manual checks. See Jesdanun, Anick. 2018. "How Google, Facebook will adapt to Europe's New Privacy Law." The Associated Press.

[46]We believe their estimates for websites catering to non-EU audience are overestimates. They explored four different definitions for "websites that cater to EU audience": 1) the website has a top-level domain that is specific to a country in the EU (for examples, .de or .fr); 2) the website appears on Alexa's rank for any country in the EU; 3) the website returns content in any of the official languages of member countries of the EU, except English; 4) the website is visited by users in Germany but not users in the US in Nielsen clickstream data. In each of the four cases, "websites that cater to non-EU audience" are defined as the websites that did not meet the criterion. We note that all four definitions would misclassify a large number of European-based English language sites with common top-level domains such as .com, .org and .net as catering to non-EU audience.

[47]For example, Joseph O'Connor, a web developer, compiled an list of 1,361 websites (mostly US-based news sites) that blocked visitors from the EU after the GDPR effective date. See O'Connor, Joseph. 2018-2019. "Websites Not Available in the European Union after GDPR." https://data.verifiedjoseph.com/dataset/websites-not-available-eu-gdpr. As of March 2019, the last time the list was updated, 1,129 websites on the list remained blocked.

[48]Sentance, Rebecca. 2018. "GDPR: Which Websites are Blocking Visitors from the EU?" Econsultancy.

the measures we are the most confident about. We discuss the rationale for our confidence in those cases in Section 6.

Moreover, we are confident that our estimates, interpreted as the differential effects between EEA networks and non-EEA OECD networks, will be nonzero if the policy has any effect at the internet layer at all. A number of empirical works have found significant and often large differential effects of privacy regulation at the application layer using a difference-in-differences approach and control groups similar to ours. These differential changes in derived demand for data exchange motivate our expectation of differential changes at the internet layer, if the policy does affect the internet layer.

The pioneering paper in this literature is Goldfarb & Tucker (2011), which found the EU's 2002 e-Privacy Directive reduced online display ad effectiveness in the EU relative to other countries. Their data consisted of survey responses of purchasing intent from online ad campaigns in which survey takers were randomly exposed to display ads. They used the implementation dates of the law in different EU countries as the policy cutoff date and constructed a triple interaction term that interacted exposure to the ads, EU membership status, and post policy treatment. Their control group of non-EU countries consisted of Australia, Brazil, Canada, Mexico, and the US.

Works specifically focused on the GDPR include Jia, Jin & Wagman (2019), which used EU-based technology ventures as their treatment group and US-based ventures as their primary control group. They used ventures based in the rest of the world[49] as an alternative control group. They found EU ventures after the implementation of the GDPR, relative to their counterparts in the US or the rest of the world, declined by as much as 26.1% in the number of monthly venture deals and between 11.1%–16.1% in the aggregate investment. Jia, Jin & Wagman (2020) used

[49]Countries in the rest of the world in their data were predominantly Australia, Canada, China, Israel, India, Japan, Russia, and South Korea.

the same design and found the impact of the GDPR on EU venture investment relative to their US counterparts was larger when ventures and lead investors were not in the same state or union. Using the same approach, Aridor et al. (2020) found the GDPR resulted in a 12.5% drop in trackable consumers on European travel platforms as compared to their non-European counterparts and resulted in declines in revenue for those firms following the implementation of the GDPR. They used travel platforms in United States, Canada, and Russia as the control group.

In addition to all of above, We acknowledge the limitation of our difference-in-differences research design that it is not able to estimate the absolute effect of the GDPR at the internet layer. We interpret our findings as the differential effect of the policy on EEA networks relative to non-EEA OECD networks. We also acknowledge that the differential impact for some of our subsamples and outcome measures are more likely to tend towards zero while others less so. We discuss this in more detail when we present results by outcome measure and subsample in our next section.

6. Results

In this section, we present regression specifications and regression results specific to each outcome variable. We also discuss results from various robustness checks at the end of this section.

6.1 The Number of Agreements between Countries

In this subsection, we study the outcome variable $numAg_{ijt}$, the number of interconnection agreements between pairs of networks owned by the countries i and j. As the unit of observations is a country pair, we need to hold fixed the EEA membership status (or OECD status) of the counterparty of interconnection while we compare the outcomes for EEA countries (treatment group) and for countries in the OECD but not in the EEA (control group).

We therefore construct three subsamples based on counterparties: (a) the counterparties are non-EEA OECD countries, (b) the counterparties are non-EEA non-OECD countries, and (c) the counterparties are EEA countries. 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 country pairs that are EEA–non-EEA OECD versus non-EEA OECD–non-EEA OECD. Subsample (b) allows us to compare country pairs that are EEA–non-EEA non-OECD versus non-EEA OECD–non-EEA non-OECD. Subsample (c) allows us to compare network or country pairs that are EEA–EEA versus non-EEA OECD–EEA. Note one observation can contribute to multiple subsamples, for example a country pair France–US can contribute to both the treatment group in (a) and the control group in (c). Appendix Figure B1 illustrates visually the construction of the three subsamples.

We note that a bias towards zero is less likely to impact regression results for subsamples (a) and (b) than results for subsample (c). The control group of either subsample (a) or subsample (b) does not involve EEA countries. As we discussed in the previous section, we believe non-EEA application layer firms are far less likely to comply with the GDPR in markets outside the EEA or change their behavior in those markets due to the regulation. Their derived demand for data exchange at the internet layer from and to those markets therefore should change little. We are concerned that results for subsample (c) may be biased towards zero as non-EEA application firms need to comply with the GDPR in EEA markets or they may exit those markets, whichever would reduce derived demand for data exchange at the internet layer.

Figure 3 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, despite the differences in levels, EEA countries and non-EEA OECD countries exhibit remarkable

parallel trends in setting up agreements with counterparties that are non-EEA OECD countries, non-EEA non-OECD countries, and EEA countries throughout the sample period. Second, agreements with developing countries or territories have a lot more noise in measurement compared to agreements within OECD countries or EEA countries.

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

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

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. $POST_{e,ijt}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,ijt}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. 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-differences 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 3. 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 3 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 2 of Table 3. The coefficient on $POST_{a,ijt} \times EEA_{ijt}$ is no longer significant and is therefore likely a spurious result.

6.2 The Number of Agreements between Countries by Agreement Type

In this subsection, we further break down the number of agreements between country pairs to provider-to-customer, peer-to-peer, and customer-to-provider types. As before, we prioritize results for subsamples where interconnection counterparties are non-EEA OECD countries or non-EEA non-OECD countries.

Figure 4 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,

$$log(numTypeAg_{ijt}+1) = \beta_0 + \beta_1 POST_{e,ijt} \times EEA_{ijt} + \beta_2 POST_{a,ijt} \times EEA_{ijt} + \gamma_{ij} D_{ij} + \lambda_t D_t + \varepsilon_{ijt}. \quad (2)$$

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 $numCustAg_{ijt}$. $POST_{e,ijt}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,ijt}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. 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-differences effect is identified by the coefficients on the interaction terms $POST_{e,ijt} \times EEA_{ijt}$ and $POST_{a,ijt} \times EEA_{ijt}$.

We show the results in Table 4. 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.[50]

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

Our earlier results suggest the GDPR did not change whether agreements were made and what types of agreements were 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.

[50]To illustrate how economically small the implied effect based on the coefficients is, we take for example the coefficient -0.038 on $POST_e \times EEA$ from column (8) of Table 4, the largest significant result in the table. The dependent variable for the regression in column (8) 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.

Motivated by this consideration, we examine how the GDPR affected the number of IP-address-level interconnection points two networks had, conditional on them having an agreement.

The outcome variable we study in this section is $numAgIP_{ijt}$, the number of IP-address-level interconnection points between network k and network l, given k and l have an agreement. As the unit of observations is a network pair, we hold fixed the EEA membership status (or OECD status) of the counterparty of interconnection while we compare the outcomes for EEA countries (treatment group) and for countries in the OECD but not in the EEA (control group).

As previously, we construct three subsamples based on counterparties: (a) the counterparties are in non-EEA OECD countries, (b) the counterparties are in non-EEA non-OECD countries, and (c) the counterparties are in EEA countries. 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.

The control group of either subsample (a) or subsample (b) does not involve networks in EEA countries. As we discussed before, we believe non-EEA application layer firms are far less likely to comply with the GDPR in markets outside the EEA or change their behavior in those markets due to the regulation. Their derived demand for data exchange at the internet layer from and to those markets therefore should change little. We are concerned the results for subsample (c) may be biased towards zero as non-EEA application firms need to comply with the GDPR in EEA markets or they may exit those markets, whichever would reduce derived demand for data exchange at the internet layer.

Figure 5 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 interconnection counterparties. We first note that 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,

$$log(numAgIP_{klt}) = \beta_0 + \beta_1 POST_{e,klt} \times EEA_{klt} + \gamma_{kl} D_{kl} + \lambda_t D_t + \varepsilon_{klt}. \tag{3}$$

We take the log of $numAgIP_{klt}$ to reflect estimated effects in percentage changes. $POST_{e,klt}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,klt}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. 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-differences 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 5 and are in no case significantly different from zero. We include agreements present for at least 150 weeks for our regressions in Table 5. Given we study 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 5.

In addition to interconnection behavior between pairs of countries or networks, we study how the GDPR might have impacted the number of agreements per network, the number of networks

per country and the sizes of the customer cones of each networks. For these results, we stress that our estimates are differential impact between EEA networks and non-EEA OECD networks. The results are biased towards zero if one wants to interpret them as absolute effects of the policy.

6.4 The Number of Agreements by Networks

Figure 6 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,

$$log(numAgNtwrk_{kt}) = \beta_0 + \beta_1 POST_{e,kt} \times EEA_{kt} + \beta_2 POST_{a,kt} \times EEA_{kt} + \gamma_k D_k + \lambda_t D_t + \varepsilon_{kt}. \quad (4)$$

We take the log of $numAgNtwrk_{kt}$ to reflect estimated effects in percentage changes. $POST_{e,kt}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,kt}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. 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. We include a dummy D_k for each network k and a dummy D_t for each month t. The difference-in-differences effect is identified by the coefficients on the interaction terms $POST_{e,kt} \times EEA_{kt}$ and $POST_{a,kt} \times EEA_{kt}$.

We show the results below Figure 6. The coefficient on $POST_e \times EEA = -.004$ ($se = 0.007$, clustered by network) and the coefficient on $POST_a \times EEA = 0.006$ ($se = 0.007$, clustered by network). Both are insignificant at conventional levels of significance. This result suggests the GDPR does not have differential impact on the number of interconnection agreements on EEA

networks relative to their non-EEA OECD counterparts.

6.5 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,

$$log(numNtwrk_{it}) = \beta_0 + \beta_1 POST_{e,it} \times EEA_{it} + \beta_2 POST_{a,it} \times EEA_{it} + \gamma_i D_i + \lambda_t D_t + \varepsilon_{it}. \quad (5)$$

We take the log of $numNtwrk_{it}$ to reflect estimated effects in percentage changes. $POST_{e,it}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,it}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. 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. We include a dummy D_i for each country i and a dummy D_t for each quarter t. The difference-in-differences effect is identified by the coefficients on the interaction terms $POST_{e,it} \times EEA_{it}$ and $POST_{a,it} \times EEA_{it}$.

We show the results below Figure 7. 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 differentially impact the number of networks in EEA countries compared to non-EEA OECD countries.

6.6 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,

$$log(NtwrkCustCone_{kt}) = \beta_0 + \beta_1 POST_{e,kt} \times EEA_{kt} + \beta_2 POST_{a,kt} \times EEA_{kt} + \gamma_k D_k + \lambda_t D_t + \varepsilon_{kt}. \tag{6}$$

We take the log of $NtwrkCustCone_{kt}$ to reflect estimated effects in percentage changes. $POST_{e,kt}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,kt}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. 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. We include a dummy D_k for each network k and a dummy D_t for each month t. The difference-in-differences effect is identified by the coefficients on the interaction terms $POST_{e,it} \times EEA_{it}$ and $POST_{a,it} \times EEA_{it}$.

We show the results shown below Figure 8. 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.

6.7 Robustness Checks

In this subsection, we discuss a few robustness checks. We first replace the logged outcome variables in all of our specifications with their original unlogged values. We present the results in Appendix C.1. As shown, all of the results are qualitatively similar to results with logged outcome variables. This alleviates the concern that our precise zero estimates are driven by taking the log of the outcome variables.

We then perform a robustness check by redefining $POST_a$ to equal to 1 if the observation is made after December 2015. As discussed in our empirical strategy section, it is possible that networks invested in anticipation of the enforcement of the GDPR even before the law was approved. We think the earliest possible date for the firms to respond in anticipation is December 2015, the time when the text of the law was fixed. The results from this robustness check for our various outcomes are almost identical to our main results presented in Tables 3–5 and Figures 6–8. This alleviates the concern that our main specifications did not capture possible effects due to anticipation.

Lastly, we perform a first differences regression for each outcome using only the subsample consisting of the treatment group in our main regression. Specifically, we run the following regression:

$$log(outcome_{mt}) - log(outcome_{m,t-1}) = \beta_0 + \beta_1 POST_{e,mt} + \beta_2 POST_{a,mt} + \gamma_m D_m + \varepsilon_{mt}, \qquad (7)$$

where m is the unit of observation of the outcome variable of interest. m can take the network subscript k, country subscript i, network pair subscript kl, or country pair subscript ij. $POST_{e,mt}$ is an indicator variable equal to 1 if time t is after the GDPR effective date. $POST_{a,mt}$ is an indicator variable equal to 1 if time t is after the GDPR approval date. A dummy D_m for each unit of

observation m is included. We do not include a time dummy as such a dummy would absorb any effect of the GDPR. The effect on first differences is identified by the coefficients on the terms $POST_{e,mt}$ and $POST_{a,mt}$. We show these results in Appendix C.2. As shown, the results are very precisely estimated and in all cases do not exceed 0.6 percentage points. These results suggest that the rate of growth in interconnection in the EEA after the GDPR did not change when compared to that before the GDPR and provide further support for our main results.

7. Conclusion

The effectiveness of the Internet in creating economic surplus depends on efficient interconnections bilaterally negotiated by independently operated networks. 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 the GDPR so far had significant effects at the application layer on European firms, we find no visible consequences at the infrastructure layer. Across multiple measures, we estimate precise zeros effects of the GDPR on EEA networks relative to their non-EEA OECD counterparts. Occasionally we estimate statistically significant effects, which prove to be not robust. Our robustness checks suggest that our results are not driven by taking the log of outcome variables or our choice of April 2016 as the first policy cutoff date. Using the first differences approach as an additional robustness check, we show that EEA networks had similar growth rates before and after the GDPR approval or implementation, therefore our main results are not driven by our choice of control group.

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. Video traffic includes Netflix and Youtube, and these were less affected by GDPR, and could grow even as other traffic drops. 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 much 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.

We also note that current empirical works, including this paper, study the impact of the GDPR on suppliers of internet services and content at various layers of the Internet. Empirical evidence on consumers' responses to privacy regulation is extremely lacking. As policy makers strive to enhance consumer welfare through better privacy protection while trying to minimize such laws' impact of the digital economy, evaluating the laws' effect on consumers is an important direction for future research to allow for the overall welfare analysis.

In addition to policy implications, our paper presents 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. Future works should keep tackling the issue of unmet data needs.

References

Aridor, Guy, Yeon-Koo Che, and Tobias Salz. 2020. "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR." Available at SSRN 3522845.

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.

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.

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.

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

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. "Consumer Consent and Firm Targeting after GDPR: The Case of a Large Telecom Provider." 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. 2011. "Privacy Regulation and Online Advertising." Management Science 57, no. 1: 57-71.

Hoofnagle, Chris Jay, Bart van der Sloot, and Frederik Zuiderveen Borgesius. 2019. "The European Union General Data Protection Regulation: What It Is and What It Means." Information & Communications Technology Law 28, no. 1: 65-98.

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.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman. 2019. "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. 2020. "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.

Jordan, Scott. 2019. "The California Consumer Privacy Act and Impact for Network Measurement and Research." Slides.

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

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.

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.

Miller, Amalia R., and Catherine Tucker. 2018. "Privacy Protection, Personalized Medicine, and Genetic Testing." Management Science 64, no. 10: 4648-4668.

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. 2014. "The 2014 Internet Peering Playbook: Connecting to the Core of the Internet." DrPeering Press.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer. 2020. "European Privacy Law and Global Markets for Data." CEPR Discussion Paper 14475.

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

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

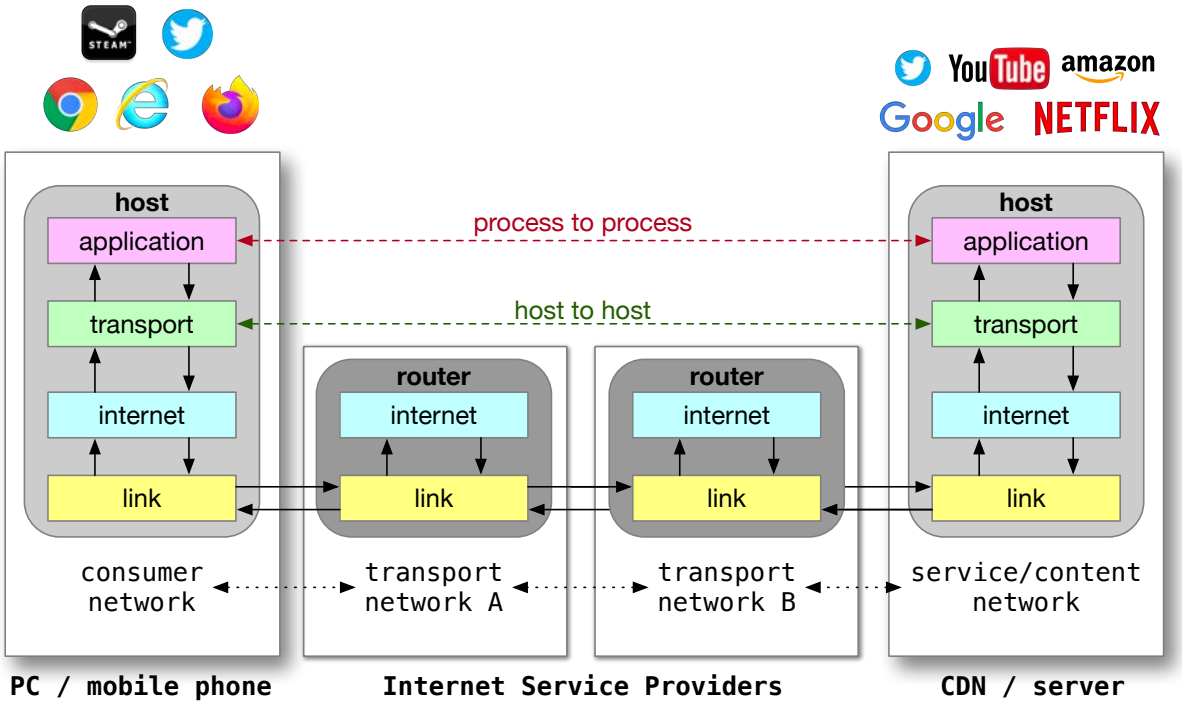
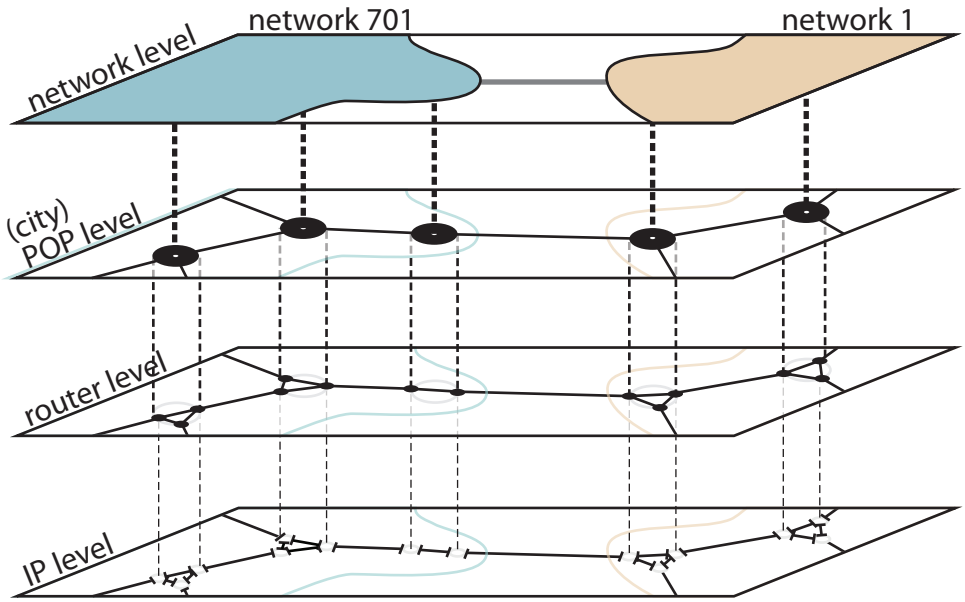


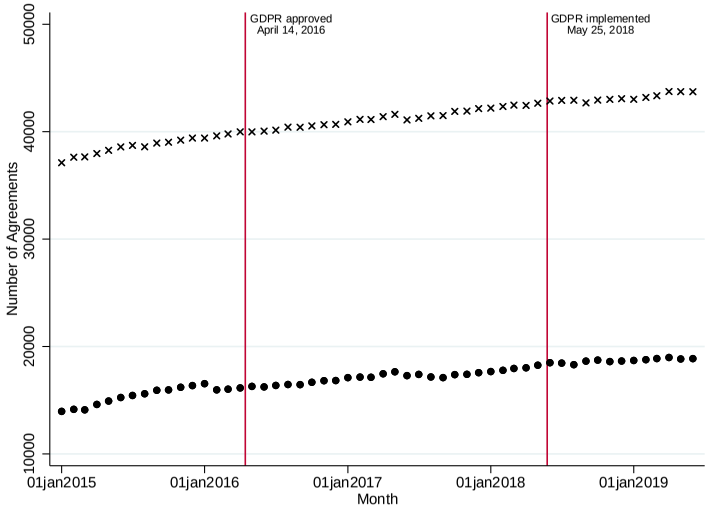
Figure 2: Data collection at the internet layer



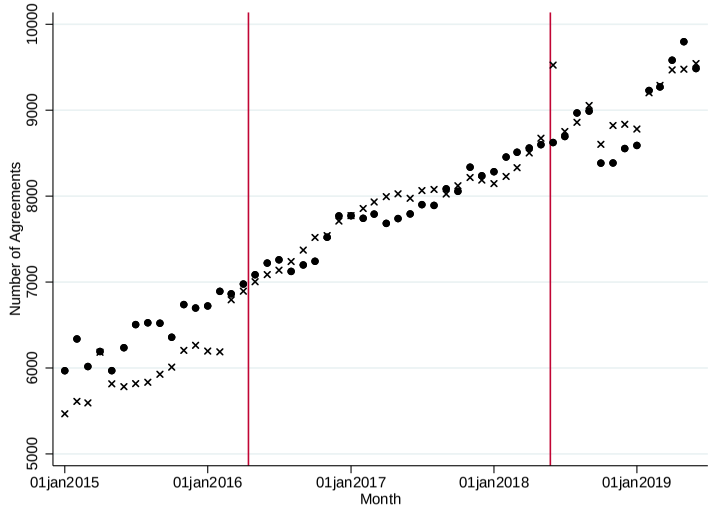
Notes: Our network-level interconnection agreements extracted from routing tables correspond to the topmost level in this figure. Our IP-address-level interconnection points for each agreement extracted from active probes correspond to the bottom level in this figure. Geolocating points of presence (PoP)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

Figure 3: Number of interconnection agreements by EEA and non-EEA OECD countries by counterparty

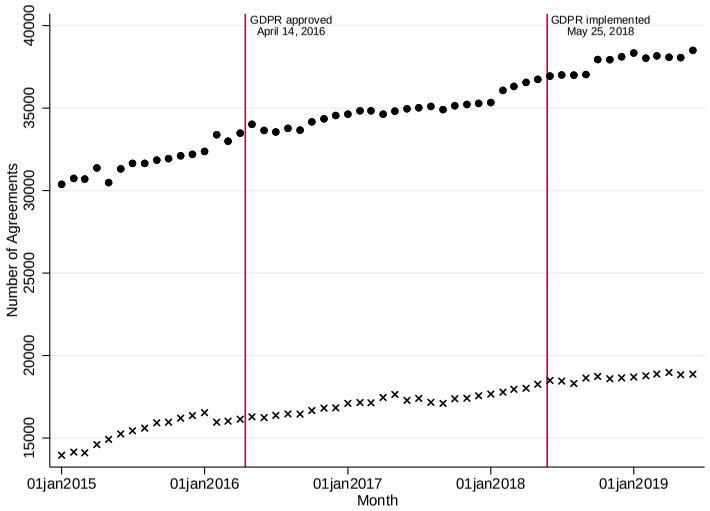
(a) Counterparties are non-EEA OECD countries



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



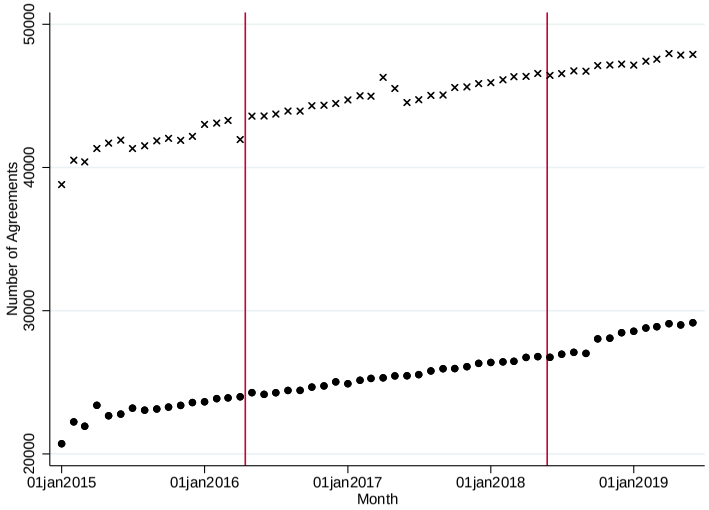
(c) Counterparties are EEA countries



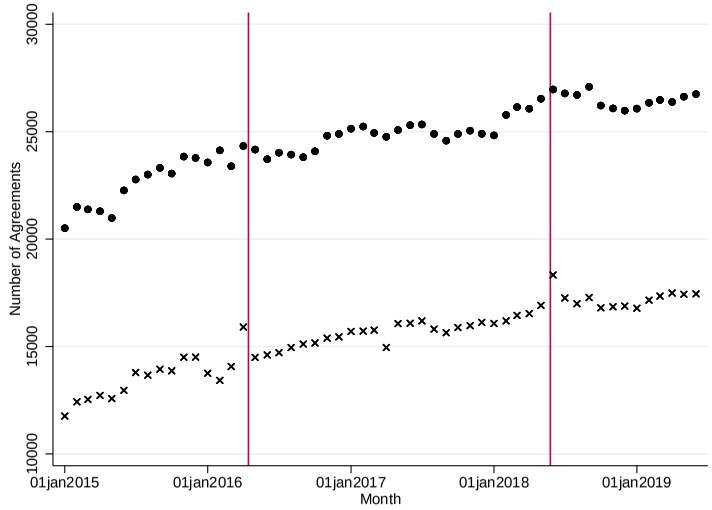
Notes: The dots represent $\sum_{i \in EEA,t} numAg_{ijt}$, the total number of agreements by networks owned by EEA countries when the counterparties are networks owned by (a) non-EEA OECD countries, (b) non-EEA non-OECD countries, (c) EEA countries. The crosses represent $\sum_{i \in OECD \wedge i \notin EEA,t} numAg_{ijt}$, the total number of agreements by networks owned by non-EEA OECD countries when the counterparties are networks owned by (a) non-EEA OECD countries, (b) non-EEA non-OECD countries, (c) EEA 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 4: Number of agreements by EEA and non-EEA OECD countries by agreement type

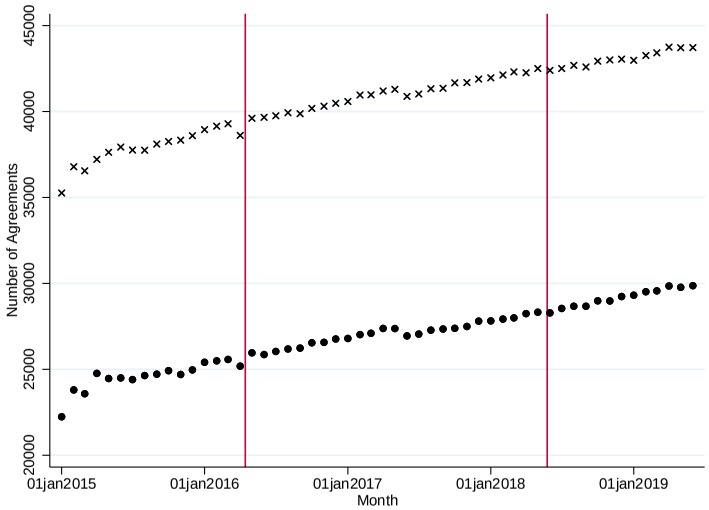
(a) Provider-to-customer (p2c) agreements



(b) Peering (p2p) agreements



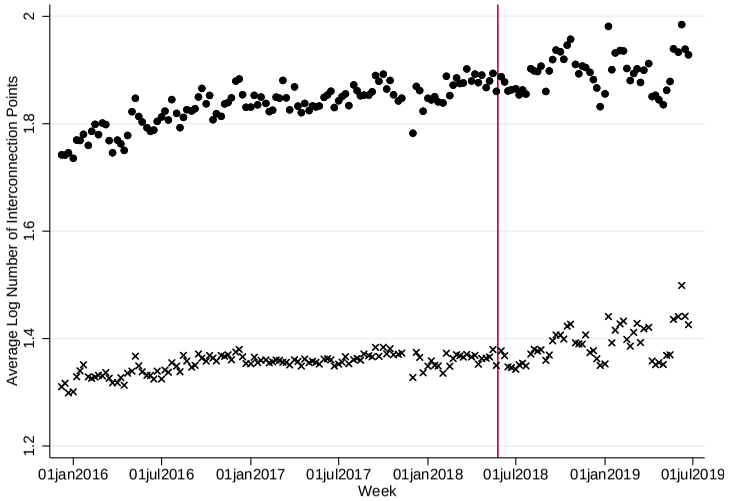
(c) Customer-to-provider (c2p) agreements



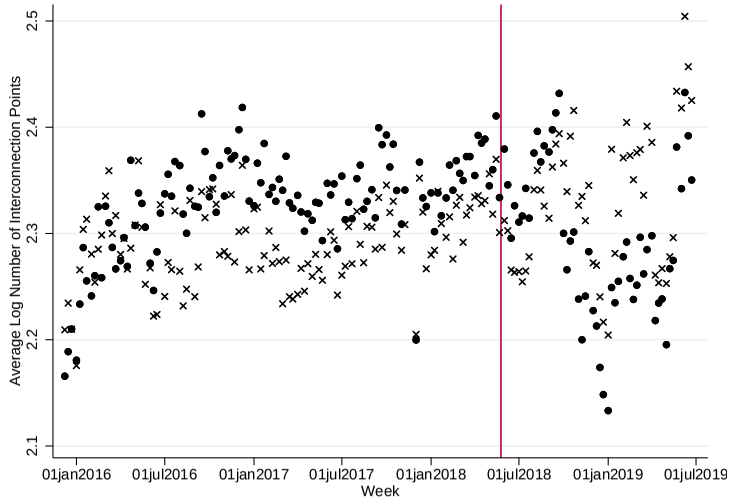
Notes: (a) The dots represent $\sum_{i \in EEA,t} numProvAg_{ijt}$, the total number of agreements by networks owned by EEA countries where these networks are providers. The crosses represent $\sum_{i \in OECD \wedge i \notin EEA,t} numProvAg_{ijt}$, the total number of agreements by networks owned by 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 5: Average log number of IP-address-level interconnection points per agreement by EEA and non-EEA OECD countries by counterparty

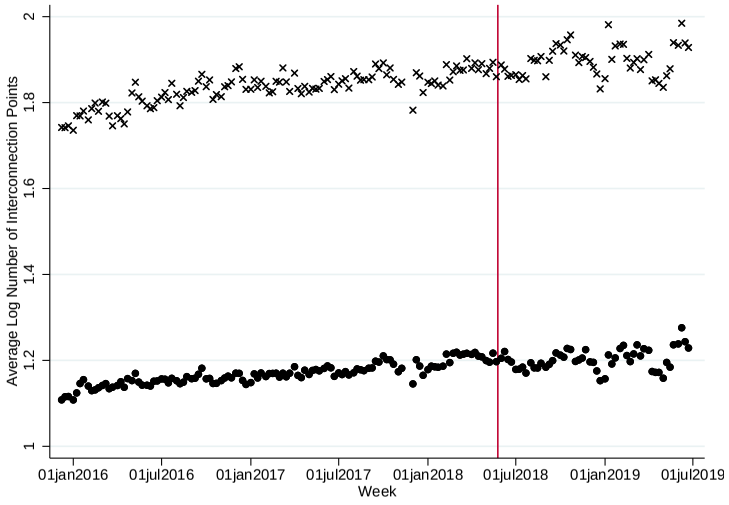
(a) Counterparties are non-EEA OECD countries



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

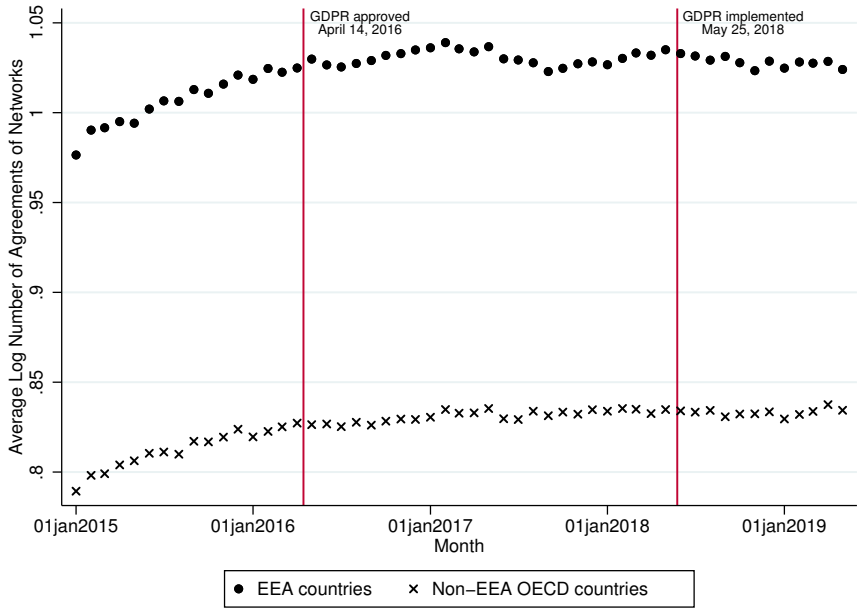


(c) Counterparties are EEA countries



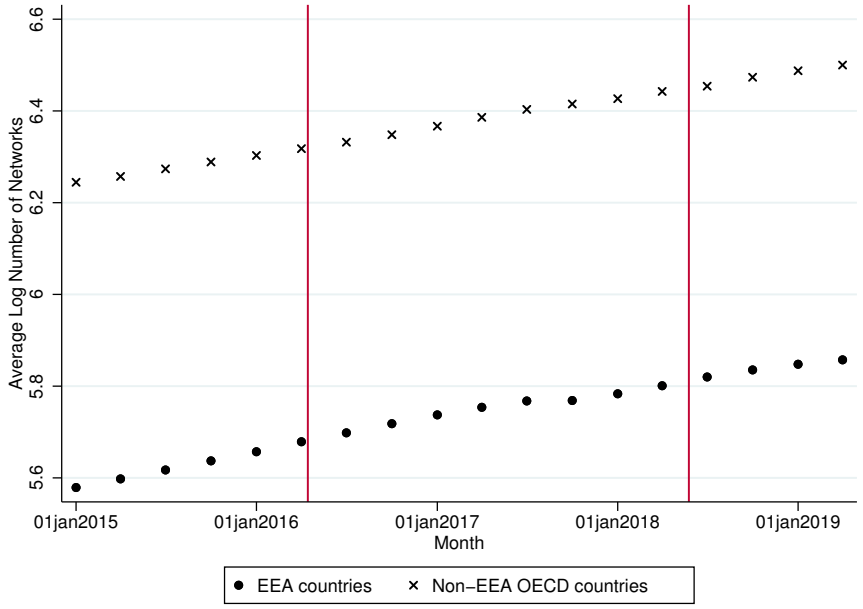
Notes: The dots represent $\overline{log(numAgIP_{i\in EEA\,jt})}$, the log number of IP-address-level interconnection points averaged among agreements by networks owned by EEA countries when the counterparties are networks owned by (a) non-EEA OECD countries, (b) non-EEA non-OECD countries, (c) EEA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

Figure 6: Average log number of interconnection agreements by networks in EEA and non-EEA OECD countries



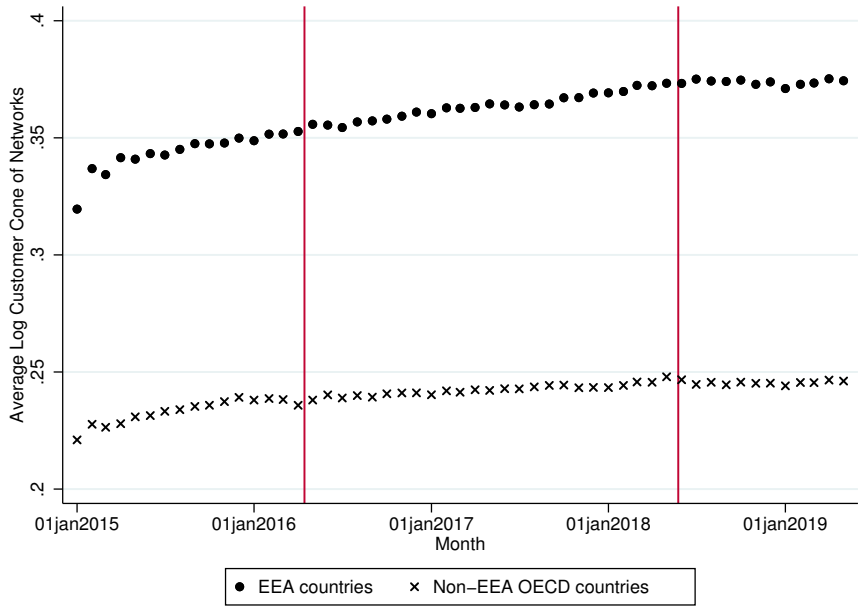
Notes: The dots represent $\overline{log(numAgNtwrk_{k\in EEA,t})}$, the log number of agreements averaged among networks owned by EEA countries. The crosses represent $\overline{log(numAgNtwrk_{k\in OECD\wedge k\notin EEA,t})}$, the log number of agreements averaged among networks owned by non-EEA OECD 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. Regression including month and network fixed effects has the coefficient on $POST_e \times EEA = -.004$ ($se = 0.007$, clustered by country) and the coefficient on $POST_a \times EEA = 0.006$ ($se = 0.007$, clustered by country). Both are insignificant at conventional levels of significance.

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
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.
numAgNtwrk_{kt}	ntwrk_k	monthly	The number of interconnection agreements network k has.	AS Relationships	
numNtwrk_{it}	ctry_i	quarterly	The number of networks country i owns.	AS Organizations	
$\text{NtwrkCustCone}_{kt}$	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					
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
numAgNtwrk_{kt}	2,909,695	5.4	55.9	1	8,391
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 EEA and non-EEA OECD countries, by counterparty

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

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. $POST_e$ 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. 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 a non-EEA OECD 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 non-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 an EEA 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 4: The GDPR's impact on the number of agreements by EEA and non-EEA OECD countries, by counterparty and agreement type

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

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. $POST_e$ 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. 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 a non-EEA OECD 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 non-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 an EEA 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 IP-address-level interconnection points per agreement by EEA and non-EEA OECD countries, by counterparty

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

Notes: The dependent variable is $log(numAgIP_{ijt})$. $POST_e$ 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. 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 a non-EEA OECD 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 non-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 an EEA 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.