Appendices to "The Impact of the General Data Protection Regulation on Internet Interconnection"

Ran Zhuo, Bradley Huffaker, kc claffy, Shane Greenstein

A. Bilateral Bargaining Model

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

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

For transit providers and content delivery networks, customers would be other networks depending on them to connect to other parts of the Internet. For governments, private companies and universities, their customers are just themselves. Assume that from O_1's perspective, forming a peer-to-peer interconnection with O_2 would allow O_1 to reach mass M_2 more efficiently. O_1 can in term generate revenue from its customers due to improved service. Assume also that O_2 would reach mass M_1 more efficiently under the peer-to-peer agreement. If O_1 is the provider, forming a provider-to-customer link with O_2 would allow O_1 to reach M_2 more efficiently. While from O_2's perspective, forming a customer-to-provider link with O_1 would allow O_2 to access I_1.[1]

[1]Note M_1, M_2, I_1 and I_2 are specific to the negotiation between O_1 and O_2. If O_1 and O_2 form a p2c agreement

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

Assume any disruption to data exchange between O_1 and O_2 is only sustained during bargaining[2] and customers do not change their networks during bargaining or in response to the bargaining outcome. We also hold fixed the interconnection agreements between either of O_1, O_2 and all other networks. Assume these other agreements allow each network to access mass G_1 and G_2. O_1 and O_2 can either form an agreement with one of the three agreement types: (a) a p2c agreement where O_1 is a provider to O_2, (b) a p2p agreement, (c) a c2p agreement where O_1 is a customer to O_2, or take the outside option (d) no agreement. In practice, as the relative masses and bargaining power of the two networks strongly influence the type of agreement formed[3], we first assume networks compare one of (a), (b), (c) with the outside option (d), rather than comparing all of the four options simultaneously, and derive comparative statics. We then discuss potential substitutions between agreement types.

where O_1 is the provider, under a negotiation between O_1 and another network O_3, M_2 becomes part of O_1's combined value of content and users M_1'.

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

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

A.1 Peer-to-Peer Agreements (p2p)

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

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

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

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

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

and for network O_2 is

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

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

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

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

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

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

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

It is also easy to show $\frac{d[h(M_1) - h(M_2)]}{dM_1} > 0$. Together, these derivatives imply two changes when M_1 decreases: (1) Gains from agreement fall. If the gains fall below zero, the agreement between O_1 and O_2 breaks. (2) O_1 receives a reduced amount of transfer from O_2, though we do not observe transfers in our data. Using similar derivations, an increase in τ would also imply higher chance of termination of the interconnection agreement, though it does not have an effect on the transfers.

A.2 Provider-to-Customer Agreements (p2c)

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

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

The negotiated net payment from O_2 to O_1 is

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

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

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

A.3 Substitution between Agreement Types

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

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

Suppose O_1 is a EU network and the GDPR negatively impacts both M_1 and I_1. $\frac{dg_\Delta}{dM_1} > 0$, implying that a decrease in M_1, holding all else fixed, would make it more likely for the two networks to enter a p2c agreement. However, $\frac{dg_\Delta}{dI_1} < 0$, implying that a decrease in I_1, holding all else fixed, would make it more likely for the two networks to enter a p2p agreement. The overall effect is unclear and depends on the relative changes to the masses M_1 and I_1 offered, their prices and O_2's revenue function.[4] Suppose instead O_2 is a EU network and the GDPR negatively impacts

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

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

In summary, this simple model formalizes the intuition that negative impacts of the GDPR on the application layer negatively impact European networks' bargaining positions. European networks would have fewer agreements of all three types and receive a reduced amount of transfers. The effect of the GDPR on the potential substitutions between agreement types is unclear.

B. Data Appendix

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

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

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

- The African Network Information Center (AFRINIC)

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

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

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

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

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

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

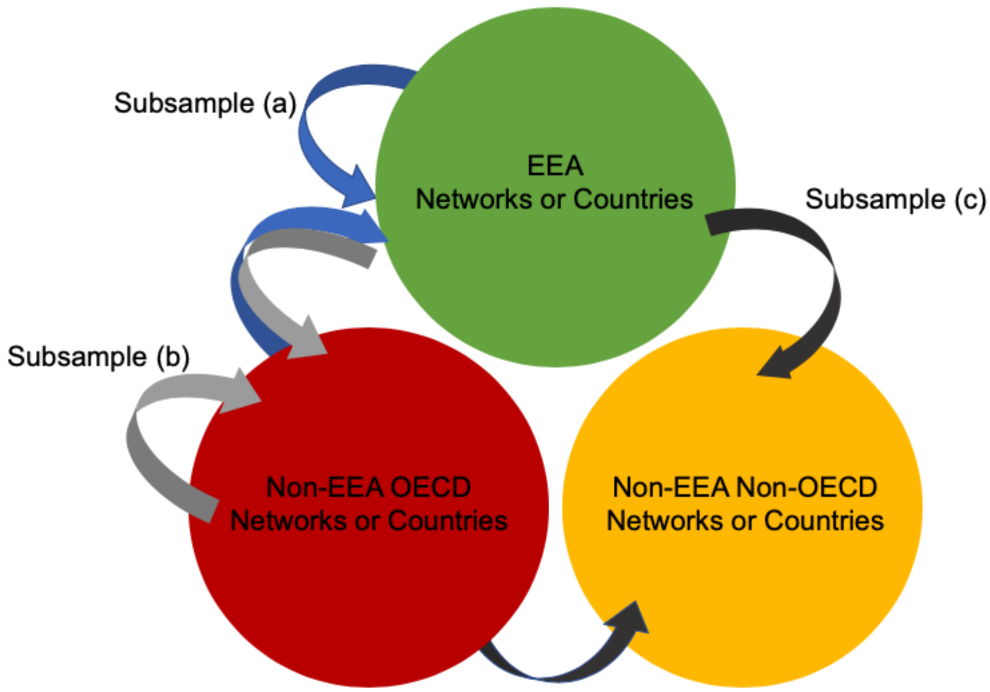
We drop networks owned by a number of small island countries, Andorra, Central African Republic, Eritrea, North Korea and Vatican City from our sample due to these countries' very small overall number of connections with the rest of the Internet. Our EEA subsample includes networks owned by organizations headquartered in the 31 EEA member countries as well as networks owned by EU-wide organizations. For networks owned by EU-wide organizations, their countries of origin are shown as "EU" in network registration records. We include these networks in the EEA subsample for the purpose of our empirical analysis. The resulting total number of countries and territories in our sample is 200. A complete list of countries and territories in our sample is presented in Table B1.

Table B1: List of Countries and Territories

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

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

Figure B1: Three subsamples for the analysis on the network pair or country pair level



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

C. Robustness Checks

C.1 Results on Levels

Table C1: 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)
$POST_e \times EEA$	−24.00	−0.48	−1.22
	(15.12)	(0.30)	(1.84)
$POST_a \times EEA$	−33.79	−0.79	0.93
	(27.52)	(0.01)	(0.02)
Group dummies	country pairs	country pairs	country pairs
Time dummies	months	months	months
Clusters	418	6,751	880
R^2	0.999	0.903	0.990
Observations	22,572	364,554	47,520

Notes: The dependent variable is $numAg_{ijt}$. 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 C2: 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)
$POST_e \times EEA$	−12.06	−2.76	−11.99	−0.03	−0.32	-0.10^{***}	0.77	−0.53	0.84
	(7.84)	(2.19)	(7.83)	(0.10)	(0.20)	(0.04)	(0.59)	(0.97)	(0.64)
$POST_a \times EEA$	−19.2	−1.68	−18.94	-0.19^{**}	−0.57	-0.114^{***}	0.29	−0.31	0.587
	(14.51)	(2.59)	(14.51)	(0.08)	(0.49)	(0.04)	(0.75)	(1.50)	(0.67)
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	0.999	0.987	0.999	0.878	0.840	0.954	0.993	0.971	0.993
Observations	25,542	22,572	25,542	364,554	364,554	364,554	74,304	47,520	74,304

Notes: The dependent variable is $numProvAg_{ijt}$ for columns (1), (4), (7), $numPeerAg_{ijt}$ for columns (2), (5), (8), and $numCustAg_{ijt}$ 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 C3: 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 (1)	Non-EEA Non-OECD (2)	EEA (3)
$POST_e \times EEA$	2.281	9.393	-1.734
	(2.326)	(8.956)	(2.650)
Group dummies	network pairs	network pairs	network pairs
Time dummies	weeks	weeks	weeks
Clusters	128	522	307
R^2	0.640	0.701	0.742
Observations	2,593,805	494,374	1,886,031

Notes: The dependent variable is $numAgIP_{ijt}$. $POST_e$ 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.

Table C4: The GDPR's impact on additional outcomes

	$numAgNtwrk_{kt}$	$numNtwrk_{it}$	$NtwrkCustCone_{kt}$
	(1)	(2)	(3)
$POST_e \times EEA$	0.039	−94.84	0.017
	(0.183)	(91.45)	(0.754)
$POST_a \times EEA$	0.167	−111.70	1.479
	(0.264)	(105.1)	(0.995)
Group dummies	networks	countries	networks
Time dummies	months	quarters	months
Clusters	43	43	43
R^2	0.963	0.998	0.978
Observations	1,275,236	1,275,236	1,275,236

Notes: The dependent variables are noted in column headers. Only observations from networks or countries in the EEA or OECD and present throughout Jan 2015 – June 2019 are used for regressions. All regressions include time dummies and group dummies. All regressions cluster standard error by country of ownership of network. Standard errors are in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.

C.2 First Differences

Table C5: The GDPR's impact on the number of agreements by EEA countries, by counterparty

	Non-EEA OECD (1)	Non-EEA Non-OECD (2)	EEA (3)
$POST_e$	-0.001	-0.000	-0.002
	(0.002)	(0.000)	(0.001)
$POST_a$	-0.002^{*}	-0.000^{*}	-0.003^{**}
	(0.001)	(0.000)	(0.001)
Group dummies	country pairs	country pairs	country pairs
Clusters	418	6,751	880
R^2	0.007	0.008	0.006
Observations	18,656	266,272	27,984

Notes: The dependent variable is $log(numAg_{ijt}+1) - log(numAg_{ij,t-1}+1)$. 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 time t is after the GDPR became effective. $POST_a$ is an indicator variable equal to 1 if time t is after the GDPR was approved. Column (1) includes observations when one party is a network owned by an EEA 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 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 and the counterparty is a network owned by an EEA country. All regressions include 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 C6: The GDPR's impact on the number of agreements by EEA 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)
$POST_e$	−0.002	0.001	−0.001	−0.000	−0.000	−0.000	−0.000	−0.003**	−0.000
	(0.001)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
$POST_a$	−0.001	−0.004***	0.004***	−0.000*	−0.000	−0.000	−0.001*	−0.006***	−0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Group dummies	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs	ctry pairs
Clusters	473	418	473	6,751	6,751	6,751	1,376	880	1,376
R^2	0.007	0.007	0.009	0.006	0.007	0.005	0.007	0.004	0.007
Observations	18,656	18,656	18,656	266,272	266,272	266,272	54,272	27,984	54,272

Notes: The dependent variable is $log(numProvAg_{ijt}+1) - log(numProvAg_{ij,t-1}+1)$ for columns (1), (4), (7), $log(numPeerAg_{ijt}+1) - log(numPeerAg_{ij,t-1}+1)$ for columns (2), (5), (8), and $log(numCustAg_{ijt}+1) - log(numCustAg_{ij,t-1}+1)$ 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 time t is after the GDPR became effective. $POST_a$ is an indicator variable equal to 1 if time t after the GDPR was approved. Columns (1), (2), (3) include observations when the treatment party is a network owned by an EEA 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 party is a network owned by an EEA 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 party is a network owned by an EEA and is the provider, peer, or customer to the counterparty network owned by an EEA country. All regressions include 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 C7: The GDPR's impact on the number of IP-address-level interconnection points per agreement by EEA countries, by counterparty

	Non-EEA OECD	Non-EEA Non-OECD	EEA
	(1)	(2)	(3)
$POST_e$	-0.001	-0.004^{**}	-0.002^{***}
	(0.001)	(0.002)	(0.000)
Group dummies	network pairs	network pairs	network pairs
Clusters	128	522	307
R^2	0.012	0.017	0.010
Observations	2,205,760	1,344,324	3,487,571

Notes: The dependent variable is $log(numAgIP_{ijt}+1) - log(numAgIP_{ij,t-1}+1)$. $POST_e$ is an indicator variable equal to 1 if time t is after the GDPR became effective. $POST_a$ is an indicator variable equal to 1 if time t after the GDPR was approved. Column (1) includes observations when one party of the agreement is a network owned by an EEA 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 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 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 $POST_a$ is not included in the regressions. All regressions include 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.

Table C8: The GDPR's impact on additional outcomes

	$numAgNtwrk_{kt}$	$numNtwrk_{it}$	$NtwrkCustCone_{kt}$
	(1)	(2)	(3)
$POST_e$	-0.001^{***}	-0.001	-0.001^{***}
	(0.000)	(0.002)	(0.000)
$POST_a$	-0.003^{***}	-0.005^{**}	-0.001^{***}
	(0.000)	(0.002)	(0.000)
Group dummies	networks	countries	networks
Clusters	43	43	43
R^2	0.007	0.218	0.008
Observations	451,316	510	451,316

Notes: The dependent variables are the first differences in the logged outcomes noted in column headers. Only observations from networks or countries in the EEA and present throughout Jan 2015 – June 2019 are used for regressions. All regressions include group dummies. All regressions cluster standard error by country of ownership of network. Standard errors are in parentheses. Significantly different from 0 in a two-tailed test at the *10% level, **5% level, ***1% level.