Unravelling the Economic and Political Drivers of Deep Packet Inspection

An empirical study of DPI use by broadband operators in 75 countries

Submitted to the GigaNet 7th Annual Symposium, November 5, 2012, Baku, Azerbaijan

Hadi Asghari
Delft University of Technology, Faculty of Technology, Policy and Management, NL
h.asghari @ tudelft.nl

Michel van Eeten
Delft University of Technology, Faculty of Technology, Policy and Management, NL
m.j.g.vaneeten @ tudelft.nl

Milton Mueller
Syracuse University, School of Information Studies, USA
mueller @ syr.edu

Abstract

The use of Deep Packet Inspection technology has been the focus of a growing amount of scholarly work due to its impact on sensitive policy issues. In this paper we look at the use of DPI for throttling or blocking peer to peer protocols by 288 broadband operators over three years, and correlate this with economic and political variables. Our empirical data shows that as of 2011, half of the studied ISPs are actively using DPI in their networks, although to varying degrees. We examine the role of seven economic and political drivers of DPI technology based on typical use-cases: bandwidth scarcity, network security, competition, surveillance, privacy protections, censorship and the strength of copyright industries. Performing bivariate analysis, we find that a few of these drivers are significantly correlated with the use of DPI.

1. Introduction

Deep Packet Inspection, a technology that enables Internet Service Providers to inspect in real time the contents of network traffic and use this information for routing decisions or data collection, has been the subject of heated policy debates. Traditionally, ISPs have been understood as "bit pipes" or "neutral conduits", passing Internet packets to and from their customers to the rest of the Internet irrespective of the contents of the packets. In reality, ISPs have always done more than being passive bit pipes, but Deep packet inspection (DPI) is nevertheless a more radical break from this model. It allows ISPs to throttle, prioritize or block certain types of traffic in real-time. Insight into the contents of traffic also allows more detailed profiling of users. These abilities change the traditional role and power of ISPs and can potentially disrupt Internet Governance. DPI impacts several sensitive policy issues, including network neutrality (Mueller and Asghari 2012), control of copyrighted material (Mueller, Kuehn, and Santoso 2012), security, censorship (Wagner 2012), privacy and intermediary liability.

This paper wishes to add empirical substance to this debate. This will be done by first briefly taking a look at the actual usage patterns of DPI by broadband operators across 75 countries for 2009-2011. The second step is to look at the driving forces behind DPI adoption. Given that DPI technology nowadays is capable of handling network loads of even the largest operators, the choice to deploy DPI is now driven by the market and political forces that an ISP faces, rather than by technical limitations. The incentives of ISPs play a significant role in the decision to use DPI - as they are the actors that need to eventually deploy the technology. For example, a key incentive to use DPI that many operators have openly acknowledged is bandwidth constraints. Another incentive is new revenue generation, illustrated by the intention of KPN Netherlands to charge customers fees for messages sent via the free WhatsApp messaging service (Preuschat 2011).[1] These incentives are balanced by other incentives, such as the need to maintain good reputation among customers. We call such incentives "market drivers". They are limited, or encouraged, by legal requirements and pressures from other actors that are following their own agendas - "political drivers". By combining the Glasnost data with data on ISPs and their environment, the paper can test whether certain market and political drivers of DPI mentioned in the literature are supported by empirical evidence. The initial ambition of this paper involved a third step to compare the relative importance of these drivers and determine which set has the strongest effect. We do touch upon this question, but will have to leave most of the exploration for future work.

The opportunity for this empirical work comes from a crowd-sourced test named Glasnost (Dischinger et al. 2010). Glasnost probes a user's Internet connection for signs of DPI deployment. It has already been run several hundred thousands of times by people all across the world. Using the recorded tests logs, we develop a "DPI score" for each Internet service provider. The steps involved in developing the DPI scores are outlined in the methodology section of this paper, building on our earlier work. This paper predominantly focuses on using the calculated DPI scores to study the issues outlined above.

[1] The plan was abandoned after the intervention of the Dutch Parliament.

2. Research Focus

We assume that the reader has a basic familiarity with the Deep Packet Inspection technology. DPI is a label for a collection of applications that detect and shape live traffic on a network. DPI recognizes patterns in and across TCP/IP packets, a data format standardized for transmitting information over an electronic network. The primary technical capability underlying DPI is the ability to recognize. DPI has been developed to detect, for example, applications, protocols, media content, viruses or data in a specific format, such as credit card numbers. Recognition supports two further capabilities: manipulation and notification. Manipulation is the ability to act on the detection, for example, by blocking, prioritizing or de-prioritizing, or otherwise regulate the flow of certain traffic. Notification concerns actions around the information that can be extracted from detection, for example, generating reports, alarms or billing incidents (Mueller 2011).

This research uses the Glasnost tests to measure the actual adoption of DPI by ISPs worldwide, and in empirically testing forces could that explain this adoption. The following research questions will be answered:

1. To what extent is DPI in active use by broadband ISPs[2] worldwide and how does this use evolve over time?
2. Which of the various economic and political drivers of DPI mentioned in the literature can be observed to have a significant correlation with the use of DPI?

The main contribution of this paper is in its attempt to answer the second research question. As far as we know, the hypothesized relationships have not yet been tested quantitatively on a global level.

3. Methodology

Answering the research questions involves the following general steps: (1) developing a DPI score for each ISP and country; (2) building a conceptual model with DPI score as a dependent variable and the drivers as independent variables; (3) Running statistical tests to identify relationships between the dependent and independent variables.

3.1 Measuring DPI use by operators

The raw data for measuring DPI use comes from a web-based test named Glasnost[3], developed by Dischinger et al. (2010). By running Glasnost, an interested user can determine whether or not her ISP is slowing down or blocking certain categories of Internet traffic, most importantly BitTorrent – a protocol used often for the exchange of media files on peer-to-peer networks. Using several upstream and downstream flows, Glasnost determines whether limitations are being imposed using traditional "port-based" methods, using Deep Packet Inspection, or not at all.

Glasnost is hosted by M-Lab, a research platform sponsored by several parties including Google. Access to the test logs are provided for free, although several stages of processing are required to make it usable for statistical work.

[2] Our definition of a broadband ISP is a network operator offering Internet access via cable, DSL, WiMax or fibre-to-the-home to retail customers.

[3] http://broadband.mpi-sws.org/transparency/bttest-mlab.php

Using the test logs turns out to be a rather laborious process. The Glasnost test logs record the underlying TCP-flow measurements for each test, but do not record the actual results (verdict of whether DPI is present or not) shown to the user after the test finishes. The logs need to be processed and analysed to obtain this verdict. This process is complicated by a large number of aborted tests, the presence of noisy measurements, and changes to the Glasnost server configuration over time regarding the test parameters. It also involves a number of corner cases with the results not always being clear-cut. The interested reader is referred to Asghari et al. (2012) for a detailed account.

After cleaning and processing the logs, we were left with approximately 262,000 "tests with verdicts" – logs that tell whether DPI based throttling was present or not – spanning from February 2009 till December 2011.

Each test has to be mapped to the operator and country of the host that ran it. The test logs contain only the IP address of the user; using GeoIP and ASN lookups, the country and "autonomous system" of the host can be determined. Autonomous Systems are connected groups of IP prefixes run by a particular operator. Determining the actual market entity responsible for an Autonomous System (AS) is not straight forward. Through a manual and labour-intensive procedure, the WHOIS entries for each AS were consulted and matched to market data purchased from TeleGeography.[4]

A DPI score for each operator is then created by dividing the number of tests indicating DPI use by the total number of tests run from that operator's network. We calculate separate scores for each year. To increase sample validity, and reduce the effects of false positives and negatives, we only include operators that have tests run from at least five different IP addresses on five separate days. The calculated score is a percentage. Taking into account a certain level of noise, we assume that a score under 9% indicates the absence of DPI. [5] Scores above 40% indicate very high use of DPI. Scores in between point to the use of DPI for throttling BitTorrent only at certain times or for certain customers.

Table 1 and 2 provide a summary of the number of logs processed and the number of observations in our final dataset.

[4] In many cases, one ASN maps perfectly to one Operator. But in some cases, multiple ASNs are_aggregated under one operator. (Extreme cases include RosTelecom, Comcast and AT&T Roadrunner, for which 10 or more ASNs belong to one operator.) In these cases, the data in these ASNs are combined. The opposite case also exists: some companies, such as UPC share one ASN across Europe (e.g. see AS6830), where in fact per country they have a different legal entity. For these cases, we split up the ASN over each country, and consequently mapped each section to their respective companies. A final complication arises when companies merge, resulting in autonomous systems changing owner, or consolidating.

[5] Event after throwing out noisy tests, Glasnost tests still have a measurement errors that can be as high as 16% (MPI 2011). Please see footnote 22 of the findings section for more details.

Table 1 – Counts of all Glasnost tests logs and those that have verdicts

Year	Glasnost test logs	# aborted or corrupt logs	# noisy tests	Tests w. verdicts (all countries)	Tests w. verdicts (select countries & ASNs)
2009	355,685	180,419	21,935	153,331	120,529
2010	203,232	114,721	17,884	70,627	55,882
2011	163,718	114,829	10,901	37,988	29,608
Total	**722,635**	409,969	50,720	261,946	**206,019**

Table 2 - Count of observations in final dataset

Country count	ASN count	Operator count	Years	Total num. of observations
75 selected (out of 207 in data)	694 selected (out of 8356 in data)	288	2009-2011	**787**

The number of countries used in this research has had to be limited to those that had (a) a sufficient number of Glasnost tests run from them, and (b) market data available. This makes 75 countries, a luckily diverse set that includes all OECD member states, the BRIC economies, and several countries from Eastern Europe, the Middle East, South-East Asia and Latin America. (The full list is presented in Section 4.)

Although DPI can have many different use cases, the metric we have captures just the use of DPI for P2P traffic management—a limitation imposed on us by the underlying data. While this limitation needs to be taken into account in interpreting the results, we believe that it provides a valid proxy for the wider use of DPI, including other use cases. Earlier research, for example by the Body of European Regulators of Electronic Communications (2012), has found that traffic management is the #1 application of DPI. And even when DPI is deployed for other uses cases, it is often also used for bandwidth monitoring. This is one explanation for observing correlations between different political forces and the DPI score, as we will see in the Section 4.

3.2 Explaining DPI use by ISPs

In this section we will build a conceptual model of what influences an ISP's decision of whether or not to deploy DPI. This model will serve as the basis for choosing independent variables. The presented model is theoretical, bringing together the existing literature and anecdotal evidence.

Implementation of DPI is not without costs. These costs include DPI equipment purchasing costs, operational costs, and possible reputational and legal risks involved with using DPI, given that the technology has certain privacy and fairness implications. Our model assumes that ISPs are economically rational actors which only deploy such a technology if the economic gains from its deployment outweigh the costs, or if the deployment is mandated by the state or the courts, or both. In all these situations, either the ISPs or the mandating parties have certain use cases in mind. A good starting point for building our model is thus to look at the different use cases of DPI.

After an extensive review of the literature, Bendrath (2009) identifies the following key functions:

- Network security: blocking malware and other dangerous traffic from reaching customers and service centres
- Bandwidth management: dealing with bandwidth scarcity; routing optimizations
- Government Surveillance: real-time monitoring of the Internet; lawful interception

- Content regulation: mandatory censorship of content considered a threat to the state or the public
- Copyright enforcement: push by the content industry for ISPs to detect and block the exchange of copyrighted content on peer-to-peer networks
- Ad-injection: analysis of traffic of consumers by ISPs and subsequently injecting ads into websites visited by those consumers

Based on the above uses cases, we can hypothesize a number of relations to exist. These will be now discussed and are grouped together in Table 3.

H1: Bandwidth scarcity coupled with high costs for bandwidth will drive ISPs towards the use of DPI. The second criterion is added to emphasize that even under conditions of scarcity, DPI must remain cheaper for the ISP then purchasing extra bandwidth capacity.[6]

H2: Network security problems[7] will drive ISPs towards the use of DPI. This hypothesis generates a complementary relationship: we would expect to see ISPs that have adopted DPI as a security solution end up with having higher (or at least equal) security performance to their peers at a later time period.

H3: States that have adopted high levels of surveillance of its population are more likely to push their ISPs to deploy DPI. The tendency of a state to use large-scale surveillance depends on the level of threat that it perceives, as well as the relative strength or weakness of civil rights protections in that country.

H4: States that have adopted high levels of censorship and control of political and social speech are more likely to push their ISPs to deploy DPI to effectively regulate online content

H5: The stronger the copyright and creative industries are in a country, the more likely the ISPs in that country are to deploy DPI. DPI can be used to curb the sharing of copyrighted material online. It has often been hypothesized that the power of copyright holders is a driving force behind the enforcement of copyrights, both by states and intermediaries such as ISPs. An oft-cited example is France's three-strikes legislation, which requires ISPs to cut off Internet access to customers, after the public authority HADOPI judges that they have been downloading infringing material for the third time and a judge confirms this sentence. Many observers have pointed to the relationship between the French government and the large media conglomerate Vivendi to explain why France has pioneered this type of legislation.

H6: Strong privacy protection regulation (as a legal barrier to the deployment of DPI) lowers the probability of DPI adoption by ISPs. Turning the final use case, ad-injection, into a hypothesis requires a bit more work. Firms will always seek extra sources of revenue, given that it is legally and socially acceptable and fits with their existing business practices. Thus instead of focusing on the

[6] Please note that although we state the hypotheses in casual terms - which is acceptable due to the existence of theory motivating each one, the statistical instruments we use actually only test for correspondence between the mentioned variables. This limitation is discussed in more detail in Section 5 of the paper.

[7] Examples of security problems include the spread of malware among an ISP's user-base, outbound spam and DDoS attacks originating from the ISP's address space.

positive push, it would be best to hypothesize the negative relations, i.e. the legal obstacles or customer sensitivity that act as barriers to using DPI for ad-injection.

Such negative relationships fall in-line with another pattern commonly seen with regards to DPI deployment. Kuehn and Mueller (2012) describe that typically ISPs initially start using DPI secretly; at some point, the issue is discovered and subsequently followed by public outcry and attention of the regulators. This in some cases results in the ISP abandoning the practice, as was the case with Comcast in the US. Thus, we have phrased the hypothesis to focus on privacy regulation as the driver.

H7: Competition in the Internet access market lowers the probability of DPI adoption by ISPs. This wording is chosen as consumer dissatisfaction with their ISP can only meaningfully manifest as a force if they have the ability to switch to a different operator.

Table 3 – The conceptual model of the drivers of DPI adoption

	Driver	Effect on DPI adoption
Market & internal drivers	Bandwidth scarcity & high costs of bandwidth	Positive
	Network security problems	Positive
	Competitive market	Negative
Political drivers	Widespread government surveillance	Positive
	Strong privacy & civil right protections	Negative
	Widespread censorship	Positive
	Powerful copyright lobbies	Positive

3.3 Empirical model

In order to test the relationships presented in Table 3, we will need to measure or otherwise quantify the expressed qualities. In a few of the cases, we have built these metrics ourselves, but in the majority of cases we use existing indicators as proxies for the measuring the quality. These indicators have been chosen from a review of the indicators available in high quality public datasets. The result of the variables and proxies are presented in the following table. Each variable is discussed in more detail in the Findings chapter of the paper.

Table 4 – Empirical model with selected indicators and proxies

	Variable	Indicator or proxy	Source
Market & Internal Drivers	Bandwidth scarcity & costs	International Internet bandwidth per Internet user ; (Lower values indicates bandwidth scarcity at the country)	ITU[8]
		Monthly Internet subscription fees for an entry level fixed broadband connection (cross-country comparable);	ITU
	Network security	Infected machines (spam bots) per subscriber	Own construct (spam data[9])
	Competition	Herfindahl–Hirschman Index, a measure of market concentration; (0=perfect competition, 10000=monopoly)	Own construct (TG data[10])

[8] International Telecommunication Union's World Telecom Indicators, *http://www.itu.int/ITU-D/ict/publications/world/world.html*.

[9] See Van Eeten et al. (2010) for details of how this metric is built.

Political Drivers	Surveillance & Privacy[11]	User data requests by government made to Google. per Internet user; (Internet user data is from WDI[12])	Google [13]
		Privacy index composed of constitutional & statutory protections, privacy enforcement and other safeguards	Privacy International[14]
	Censorship	Censorship of political topics (views opposing the government) online; scored between 0-4	ONI[15]
		Censorship of socially sensitive topics (e.g. sexuality, gambling and illegal drugs) online; scored between 0-4.	ONI
		Freedom of the press index, an annual survey of media independence that assesses the degree of print, broadcast, and internet freedom in every country	Freedom House[16]
	Surveillance, Privacy & Censor.	The Polity index examines concomitant qualities of democratic and autocratic authority in institutions. The score is a spectrum from total autocracies to democracies.	Polity-IV Project[17]
	Strength of copyright industry	Creative industry exports as percentage of services trade; (Higher percentages indicate a stronger copyright industry)	UNCTAD Statistics[18]
		Software piracy rate; based on the volume and value of unlicensed software installed on PCs in a given year; (High rates indicate a weak copyright industry)	BSA [19]

3.4 Statistical instruments

We use a straightforward statistical instrument to explore the relationships, namely the Spearman rank correlation. This is a non-parametric test of statistical dependence. We use it to test the existence and strength of (bivariate) relationships between the DPI scores and the different independent variables. Most of the variables in our dataset are not normally distributed and some are based on an ordinal scale. This makes the use of a non-parametric test necessary.[20]

Spearman's rank correlation returns a p-value and a correlation coefficient (rho). We are interested in both. The p-value is the probability that the observed coefficient might be a result arising by chance. We use the 0.05 significance level and accept two variables as correlated when $p<0.05$. The

[10] TeleGeography GlobalComms, *http://www.telegeography.com/research-services/globalcomms-database-service/index.html*.

[11] We have combined the two categories as the Privacy Index contains data on both as privacy safeguards and surveillance are to a certain extent two sides of the same equation and highly connected.

[12] Worldbank's World Development Indicators, *http://data.worldbank.org/data-catalog/world-development-indicators*.

[13] Google Transparency Report, *http://www.google.com/transparencyreport/userdatarequests/*.

[14] Privacy International's Surveillance Monitor 2007, *https://www.privacyinternational.org/reports/surveillance-monitor-2007-international-country-rankings*.

[15] OpenNet Initative, *http://opennet.net/research/data*.

[16] Freedom House's Freedom of the Press data, *http://www.freedomhouse.org/report-types/freedom-press*.

[17] Center for Systematic Peace's Polity IV project, *http://www.systemicpeace.org/polity/polity4.htm*.

[18] United Nations Conference on Trade and Development, *http://unctad.org/en/Pages/Statistics.aspx*.

[19] Business Software Alliance's Global Software Piracy Study, *http://portal.bsa.org/globalpiracy2011*.

[20] Spearman's rank correlation has a few other advantages in addition to being non-parametric. (i) it detects all monotonic correlations between two variables, not just linear ones (ii) it is much less sensitive to outliers and (iii) it is not disrupted by skewed variables, removing the need for variable transformations.

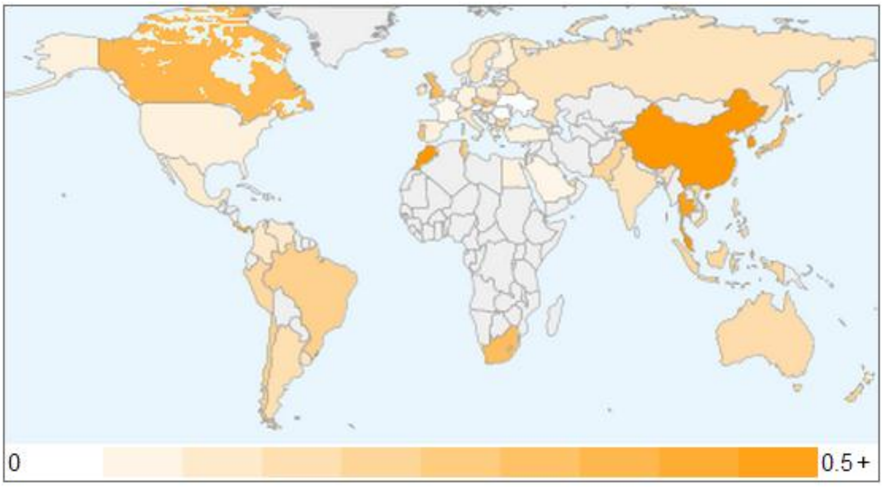
strength of the relationship is however determined by rho, and the sign of coefficient tells us whether relationship is negative or positive.

At this point, after calculating the DPI score, building an empirical model, and choosing our statistical instrument, we are ready to proceed to the findings.

4 Findings

4.1 Descriptive findings

A basic finding of this work is that Deep Packet Inspection is in wide use by ISPs across the world. In 2011, under a third of the countries in our dataset show no significant use of DPI by their ISPs. Of the remaining countries, half show some level of DPI deployment, and the other half pervasive use of the technology – meaning more than a few of their operators have deployed DPI, and in some cases use it to throttle nearly all traffic. Figure 1 visualizes these patterns.



Countries with pervasive DPI use:

Brazil, Canada, China, Costa Rica, Czech Republic, Hong Kong, Israel, Japan, Macau, Malaysia, Morocco, Panama, Portugal, Singapore, South Africa, Slovakia, South Korea, Thailand, Trinidad and Tobago, Tunisia, United Kingdom

Figure 1 - Heatmap of DPI usage among ISPs across 75 countries in 2011 (0.5 and higher are shown as one shade)

We see the widespread use across all three years, but with a surprising pattern: *a rise in DPI use in 2010 and a subsequent drop in 2011*.

Table 8 displays the DPI scores aggregated at the country level. These country scores are weighted averages of the operator scores in each country, taking into account the number of subscribers. (See footnote 24 for details on why weighted averages are used.)

Table 5 groups the countries in three categories: negligible, noticeable and pervasive DPI use[21]; the stated pattern is clearly visible. Table 6 looks at the individual operator scores, irrespective of country. [22] Again, the pattern clearly holds: in 2009, 52% of the ISPs were using DPI. This rises to 64% and then drops back to 48%.

[21] The boundary conditions have been chosen as follows: countries in the negligible group contain zero (or just one small) operators using DPI; the noticeable category have one or more operators using DPI for a select part of their users; the pervasive group have many operators doing DPI, or a few doing DPI at high levels, or both.

[22] The reason that DPI scores for ISPs up to 0.09 are considered as "No DPI" is related to the existence of *false positives* in the Glasnost data. These are cases where Glasnost detects speed differences in BitTorrent traffic, and concludes the use of DPI, but in fact this is not the case. Dischinger et al. (2010) give a detailed analysis of why this miscategorization happens , and MPI (2011) states that this can occur in up to 16% of the cases, i.e.,

Table 5 – Level of DPI usage in Countries (as deployed by ISPs) – summed by category and year

Year	Negligible use *country score<.09*	Noticeable use *0.09 ≤ country score ≤ 22*	Pervasive use *country score > 0.22*	Total
2009	22	26	28	75 countries
2010	17	31	27	75 countries
2011	22	27	21	70 countries

Table 6 – Grouping of ISPs based on whether they use DPI or not (see footnote 22 for score ranges)

Year	No DPI *DPI score<.09*	Unknown *.09≤DPI score≤.13*	Yes – Med. DPI use *.13≤DPI score<.40*	Yes - High DPI use *DPI score≥.40*	Total
2009	116 48%	35 --	68 28%	59 24%	*278*
2010	80 36%	49 --	85 39%	56 25%	*270*
2011	109 52%	31 --	68 33%	31 15%	*239*

How can we explain the peak in 2010? An explanation that comes to mind for the surge is that mass diffusion of DPI technologies took place in 2010. This could have been for a variety of reasons, such as the technology becoming more affordable, or increasing awareness of the possibilities and benefits among ISPs (and other interested actors). The subsequent drop can be explained by the negative pushback in some markets – be it by market forces or political pressure. Table 7 lists countries according to the rise or drop of DPI from 2009 to 2011.

Table 7 - DPI use trend by ISPs, 2009-2011

DPI Trend 2009-2011	Countries
Increasing	Australia, Denmark, Japan, Morocco, Peru, Philippines, Slovakia, Sweden, Trinidad and Tobago, United Kingdom
Decreasing	Argentina ,Austria , Dominican Rep., Hungary, Ireland, Italy, Lithuania, Malaysia, New Zealand, Poland, Puerto Rico, Romania, Saudi Arabia, Switzerland, UAE, Venezuela
Stable or missing/inconclusive	33 + 16 countries

The last observation is that DPI implementations seem to be moving towards "less aggressive" practices (moving from the Yes–High to the Yes-Medium category in Table 6). In other words, operators are using DPI, but for a smaller portion of their subscriber base or only during peak hours.

the false positive rate could be as high as 16%. With the help of qualitative data on the individual ISPs in the dataset – that is, by knowing which ISPs are in fact using DPI and which are not – we can fine tune the false positive rate. We have found that in practice ISPs with scores under 0.09 are not using DPI and those with scores above 0.13 are using it. Scores between 0.09 and 0.13 remain unclear, and for this reason are classified as having an unknown status. A final difference between the country and ISP scores should be noted: Country scores are averages, hence we classify countries as say having *negligent* or *noticeable* DPI use (not no or yes); while ISPs as we saw can be classified with clear *yes/no* values.

Table 8 - DPI scores 75 countries (288 operators) 2009-2011

Country	2009			2010		2011			BB Subs	BB Ops	Trend 2009-2011[23]
	DPI score[24]	DPI ops[25]	Tests	DPI score	Tests	DPI score	DPI ops	Tests			
Argentina	20.3%	3	1166	22.4%	400	15.1%	2	249	4.4 m	5	-
Australia	8.8%	2	1938	14.8%	1471	16.2%	5	743	5.4 m	6	+
Austria	20.6%	1	1181	10.4%	334	0.7%	0	88	2.2 m	3	-
Belarus	7.4%	0	27	5.9%	17	14.3%	1	7	1.3 m	1	?
Belgium	13.4%	1	484	19.8%	230	11.4%	1	104	3.5 m	3	=
Brazil	31.1%	5	14670	28.0%	3572	22.0%	5	2554	15.6 m	7	=
Bulgaria	3.8%	0	26	30.7%	26	20.6%	1	19	1.3 m	3	?
Canada	31.4%	4	7920	28.4%	3389	34.5%	4	2561	10.5 m	9	=
Chile	17.9%	4	825	20.5%	328	20.2%	2	250	2.0 m	5	=
China	57.2%	3	376	67.0%	149	71.3%	2	41	150 m	3	=
Colombia	13.0%	1	164	12.9%	90	9.4%	1	79	3.3 m	4	=
Costa Rica	15.2%	1	33	19.3%	21	30.0%	1	10	0.3 m	2	=
Croatia	7.1%	1	137	8.5%	105	7.2%	0*	23	0.9 m	3	=
Cyprus	43.3%	1	42	5.5%	20	0.0%	0	7	0.2 m	2	?
Czech Republic	27.3%	3	384	21.3%	184	24.3%	2*	45	2.3 m	4	=
Denmark	2.2%	0	333	4.8%	176	4.2%	1	46	2.2 m	3	+
Dominican Rep.	40.5%	1	74	6.7%	15	10.0%	0	20	0.4 m	1	-
Egypt	9.5%	1	85	5.9%	32	5.6%	0	18	1.8 m	2	?
Estonia	6.9%	0	102	9.8%	61	9.7%	0	23	0.4 m	2	=
Finland	6.0%	0	518	9.8%	285	6.6%	0	63	1.6 m	4	=
France	7.8%	0	2224	8.4%	867	3.0%	0	335	22.8 m	5	=
Germany	7.3%	2	2748	8.9%	1283	5.2%	1	327	27.3 m	9	=
Greece	6.4%	0	1371	13.8%	593	4.6%	0	339	2.6 m	5	=
Hong Kong	69.0%	4	2101	52.8%	253	44.4%	3*	104	2.5 m	4	=
Hungary	18.2%	2	1176	15.2%	540	8.4%	0	194	2.1 m	5	-
Iceland	61.1%	2	62	59.8%	29	18.5%	1	11	0.1 m	2	?
India	9.5%	3	1407	11.2%	1097	12.9%	2*	858	13.4 m	8	=
Indonesia	13.6%	2	143	28.3%	75	17.6%	2	88	2.1 m	3	=
Ireland	25.1%	4	797	19.2%	316	9.5%	2	92	1.1 m	6	-
Israel	61.4%	3	2253	56.1%	2503	30.7%	3	206	2.1 m	3	=
Italy	14.4%	3	6871	13.6%	2872	10.3%	2	711	13.4 m	5	-
Japan	27.9%	2	1187	20.1%	660	29.9%	3	303	35.2 m	7	+
Latvia	16.8%	2	71	14.6%	91	.	.	.	0.6 m	3	
Lithuania	31.5%	3	118	18.8%	54	10.7%	1	28	0.7 m	3	-
Luxembourg	0.0%	0	11	10.0%	10	.	.	.	0.2 m	1	
Macau	71.7%	1	92	78.3%	23	78.6%	1	14	0.1 m	1	=
Macedonia FYR	0.0%	0	15	3.1%	16	.	.	.	0.3 m	3	
Malaysia	84.8%	2	1143	77.0%	669	57.0%	1	325	2.7 m	2	-
Mexico	6.8%	2	464	11.1%	286	9.1%	2	195	12.3 m	5	=
Morocco	2.7%	0	37	20.0%	15	63.2%	1	68	0.6 m	1	+
Netherlands	8.4%	1	1962	6.9%	794	4.4%	1	163	6.4 m	6	=

[23] The trend indicates whether DPI use has increased (+), decreased (-), remained more or less the same (=) between 2009 and 2011, or is unknown (due to missing data or unclear op scores).

[24] Country level DPI scores are calculated using a weighted average of the operator scores in each country. The weights are based on the broadband market share of each operator. The reason that a weighted average is used is that otherwise the scores would be skewed towards that of the ISP whose users run the most tests. Glasnost is crowd-sourced, so it is quite likely that users that are suspicious their ISP will run the test more. Consider the case of Germany, in which one small ISP is doing DPI, while all the other are not. In such a case, unweighted scores would unfairly show a high total DPI.

[25] The remaining operators in that country could be not using DPI, inconclusive, or missing for that year

New Zealand	**34.2%**	5	*952*	**41.1%**	*535*	**19.9%**	3	*286*	1.2 m	6	-
Norway	**5.8%**	0	*487*	**7.6%**	*207*	**8.8%**	0	*92*	1.8 m	4	=
Pakistan	**22.3%**	1	*146*	**24.9%**	*120*	**19.2%**	2	*54*	1.8 m	4	?
Panama	**48.8%**	1	*41*	**37.8%**	*37*	**47.4%**	1	*19*	0.3 m	1	=
Peru	**11.6%**	0	*69*	**30.3%**	*33*	**18.2%**	1	*33*	1.2 m	1	+
Philippines	**12.6%**	2	*945*	**19.4%**	*519*	**16.2%**	3	*324*	4.6 m	5	+
Poland	**22.5%**	4	*2904*	**21.3%**	*840*	**8.0%**	1	*305*	5.5 m	5	-
Portugal	**39.8%**	3	*2329*	**32.9%**	*1765*	**27.5%**	3	*456*	2.2 m	5	=
Puerto Rico	**23.9%**	1	*132*	**23.1%**	*83*	**3.9%**	0	*37*	0.6 m	3	-
Romania	**15.3%**	1	*941*	**11.5%**	*422*	**0.8%**	0	*113*	3.3 m	4	-
Russia	**15.2%**	2	*540*	**19.4%**	*501*	**12.7%**	3	*138*	19.5 m	6	?
Saudi Arabia	**14.7%**	3	*92*	**17.9%**	*78*	**4.3%**	0	*23*	2.1 m	3	-
Serbia	**10.3%**	0	*160*	**16.0%**	*56*	**0.0%**	0	*35*	1.2 m	3	?
Singapore	**66.4%**	4	*1410*	**49.7%**	*755*	**36.4%**	3	*244*	1.3 m	4	=
Slovakia	**12.6%**	1	*527*	**5.8%**	*59*	**27.4%**	2	*29*	1.0 m	3	+
Slovenia	**13.9%**	2	*134*	**14.8%**	*73*	**.**	.	*.*	0.5 m	5	
South Africa	**31.9%**	2	*80*	**38.3%**	*282*	**31.1%**	2	*73*	1.1 m	4	=
South Korea	**97.5%**	3	*127*	**85.8%**	*123*	**74.7%**	3	*57*	18.7 m	3	=
Spain	**7.1%**	0	*2504*	**8.7%**	*1393*	**6.3%**	0	*637*	11.2 m	6	=
Sri Lanka	**8.9%**	1	*42*	**26.2%**	*37*	**.**	.	*.*	0.4 m	2	
Sweden	**9.7%**	1	*535*	**17.3%**	*324*	**11.6%**	2	*142*	3.0 m	4	+
Switzerland	**11.3%**	1	*316*	**7.5%**	*196*	**1.0%**	0	*34*	3.0 m	3	-
Taiwan	**11.0%**	0*	*2778*	**11.7%**	*660*	**16.5%**	1	*333*	5.6 m	1	=
Thailand	**56.1%**	3	*261*	**62.6%**	*532*	**44.7%**	3	*198*	3.9 m	3	=
Trinidad & Tobago	**6.0%**	0	*50*	**20.0%**	*74*	**33.9%**	2	*63*	0.2 m	2	+
Tunisia	**39.3%**	1	*61*	**13.2%**	*38*	**25.5%**	1	*51*	0.6 m	1	=
Turkey	**9.0%**	1	*99*	**14.3%**	*158*	**5.9%**	0	*59*	7.6 m	3	?
Ukraine	**7.4%**	0	*82*	**3.9%**	*51*	**0.0%**	0	*17*	4.5 m	5	=
U. Arab Emirates	**30.6%**	2	*223*	**26.2%**	*132*	**15.5%**	1	*79*	0.9 m	2	-
United Kingdom	**24.2%**	4	*8332*	**28.4%**	*4631*	**31.7%**	5	*2157*	20.4 m	7	+
United States	**5.7%**	2	*36218*	**6.4%**	*16994*	**6.2%**	0	*11981*	83 m	15	?
Uruguay	**15.4%**	1	*39*	**0.0%**	*10*	**11.8%**	0*	*17*	0.5 m	1	?
Venezuela	**14.7%**	2	*159*	**15.1%**	*153*	**10.8%**	1	*82*	1.8 m	2	-
Vietnam	**21.1%**	3	*76*	**13.9%**	*60*	**17.9%**	2*	*97*	4.5 m	3	=

4.2 Bivariate relationships

The results of the bivariate correlation tests between DPI score and the model's independent variables are summarized in Table 9 and explained in the proceeding paragraphs.

Table 9 – Results of bivariate correlations between the operator-DPI-score and several indiciators [26]

	Independent Variable	Spearman rank correlation	Expected relation	Observed relation	Hypothesis is:
Market & Internal Drivers	International Internet bandwidth per user	p=0.00 rho=-0.17 784 obs	Negative	Significant, negative	Accepted
	Broadband Internet monthly subscription fees	p=0.00 rho=-0.12 773 obs	Negative	Significant, negative	Accepted
	Infected machines per sub. (Op. security performance)	p=0.00 rho=0.12 783 obs	Negative*	Significant, positive	Rejected
	Herfindahl-Hirschman Index (Market concentration)	p=0.00 rho=0.13 781 obs	Positive*	Significant, positive	Accepted
Political Drivers	User data requests (by governments) per user	p=0.82 401 obs	Positive	Insignificant	Rejected
	Privacy Index	p=0.01 rhi=-0.10 595 obs	Negative	Significant, negative	Accepted
	Censorship of political topics on the web	p=0.00 rho=0.20 454 obs	Positive	Significant, positive	Accepted
	Censorship of social topics on the web	p=0.00 rho=0.30 454 obs	Positive	Significant, positive	Accepted
	Freedom of the press index	p=0.00 rho=0.20 775 obs	Positive*	Significant, positive	Accepted
	Polity index	p=0.00 rho=-0.21 753 obs	Negative	Significant, negative	Accepted
	Creative services exports (as % of services trade)	p=0.00 rho=-0.14 687 obs	Positive	Significant, negative	Rejected
	Software piracy rate	p=0.00 rho=0.15 778 obs	Negative*	Significant, positive	Rejected

** See footnote 26*

[26] An operationalization note regarding relations marked with stars: the encoding of several of the variables is such that it causes the expected relation direction to "flip". As an example, the *freedom of the press* index actually indicates *more* restrictions when it has *larger* values. Thus, the hypothesis that *press restrictions ≈ more censorship ≈ more DPI* would have an expected positive direction, even though the direct wording might suggest otherwise. The same holds for the other starred variables. The *network security performance* variable has an extra complexity attached due to the way its hypothesis is tested, which is explained in the text.

Market & internal drivers

***Hypothesis: <u>Bandwidth scarcity</u> coupled with high <u>costs for bandwidth</u> drive ISPs towards DPI* → ACCEPTED**

We have used two indicators to test this hypothesis. The first is *international Internet bandwidth per Internet user*. This indicator acts as a proxy for the abundance or scarcity of bandwidth at the country level, as a large part of Internet traffic is destined for (or originates from) locations outside of a user's country via international links. The indicator has a significant negative correlation: countries with higher bandwidth per user on average make less use of DPI.

The second indicator is the *monthly subscription fee for an entry level broadband connection*.[27] This indicator is a proxy for the ability of ISPs to compensate their bandwidth costs by demanding higher prices from customers. It is also negatively correlated with DPI.

Putting the two findings together, we can say that ISPs running low on bandwidth which cannot push the costs for purchasing new capacity to their subscribers, seem more likely to deploy DPI to effectively manage their bandwidth.

Hypothesis: Network <u>security problems</u> drive ISPs towards DPI → REJECTED

To be able to test this hypothesis, we had to take a different approach from the other hypotheses. Cyber-security performance can be both a *driver* for DPI and an *outcome* of its deployment. Consider an ISP that faces severe malware problems among its user base and decides to deploy DPI to boost network security and reduce the spread of malware. Security performance functions as a driver in this situation. Now, after the decision making and deployment phase, and after sufficient time for the implementation to have an actual effect, we expect an improvement in the security performance of the ISP, i.e. a drop in the number of infected machines to levels equal to or better than other. Here, cyber-security performance is the outcome of DPI.

What the empirical evidence shows us is that in fact ISPs using DPI have a worse security performance than their peers. Higher DPI use correlates with a larger *percentage of bot infected machines in the subscriber base*. To explore the possibility that the DPI implementation achieves the desired effect over time, we can look at the correlation per year. Table 10 presents the results. For the first two years, we found no significant relationship. Then, for 2011, we found a positive relationship, in other words, ISPs that have adopted DPI have a higher rate of infected machines in their network. If security had been a driver for DPI adoption, we would expect the relationship to move towards a negative correlation, because the ISPs that adopted the technology, we moving towards lower infection rates. The exact opposite occurs in the evidence. This suggests that that either DPI is not being used for solving security problems, or if it is, it is very ineffective, which seems unlikely.

Table 10 - Correlation between network security performance and DPI use across different years

Driver	Correlation 2009	Correlation 2010	Correlation 2011
Infected machines per subscriber	*Insig.*	*Insig.*	*Sig., positive*

[27] The ITU defines the entry level connection as having a speed of at least 256kbps and a minimum cap of 1GB. The prices are expressed in USD PPP, making the variable comparable across countries.

Hypothesis: <u>Competition</u> in the broadband market lowers the probability of DPI adoption
→ ACCEPTED

The indicator used here is the *Herfindahl–Hirschman Index*, a measure of market concentration. We calculate it using the market share of the largest broadband ISPs in each country (subscriber counts as a percentage of total broadband subscribers). More concentrated markets are thought to have players with more market power and therefore less competition. A positive correlation is found, indicating that DPI use is higher in more concentrated, less competitive markets.

A shortcoming of using HHI to measure competition in in broadband markets, is that it relies on market shares and they can be a bit misleading in some situations. For example, in some large countries, there may be a substantial number of providers, but some of them may actually be regional monopolies. In other words, they do not really compete with each other. In markets with regional monopolies, national market shares do not adequately express the level of competition. In the majority of countries we have mapped, this does not seem to be an issue.

Political drivers

Hypothesis: States that enact high levels of <u>surveillance</u> are more likely to push ISPs to deploy DPI.
Hypothesis: Strong <u>privacy protection regulations</u> lower the probability of DPI adoption by ISPs.
→ ACCEPTED

We present these two hypotheses jointly, as the issue of privacy regulations and surveillance are highly connected. In fact, the *Privacy Index*, developed by the NGO Privacy International, combines both phenomenon in a single index. It includes constitutional protection, statutory protection and privacy enforcement on the one hand, and visual surveillance, communication interception and data retention on the other[28].

The correlation between DPI and privacy index is negative, in accordance with our hypotheses. Countries with stronger privacy safeguards – in law and in practice, show lower rates of DPI use. A caveat regarding the privacy index is that it has been constructed in 2007, while we use it for 2009-2011. We do not expect this to have a significant impact, as privacy regulation and surveillance practices are institutionalized phenomena and therefore not subject to rapid change.

The other indicator, *governmental requests to Google to disclose user data* (normalized by dividing by the number of Internet users in each country) generates an insignificant correlation. In hindsight, this indicator might be a rather weak proxy for government surveillance. It could be the case that some governments that request Google to disclose information on users do so due to an absence of other surveillance mechanisms. It could also be that some governments simply do not have the necessary legal arrangements in place to make such requests to Google and may have reasons not to seek such arrangement. To illustrate, China has made zero requests. This is most likely why only 28 countries have data.

In summary, using the Privacy Index as the proxy variable, we accept these hypotheses.

[28] Please see Privacy International's Surveillance Report 2007 for the complete list of components.

Hypothesis: States that pursue high levels of <u>censorship</u> are more likely to push ISPs to deploy DPI → ACCEPTED

Using three different indicators that quantify the level of censorship in a country, we obtain the same result: a significant positive correlation exists between higher levels of censorship and increased use of DPI.

Two of these indicators have been constructed by the OpenNet Initiative (ONI). One measures the extent to which political websites – sites that publish views opposing government views – are censored. The other measures the extent of censorship for websites touching socially sensitive topics (e.g. sex, gambling and illegal drugs). Both have a positive correlation, as hypothesized. An interesting outcome is that the correlation coefficient for social censorship is much stronger than that of political censorship (0.30 versus 0.20). Multivariate analysis will be needed to explore the meaning of this difference.

The ONI data is only available for forty of the countries in our sample. It is also not available in a time series format. Just one data point is provided for each country, based on the last time that the measurement was performed, ranging from 2007 to 2011. We therefore also used a third indicator which doesn't suffer from these shortcomings.

The third indicator is the *Freedom of the Press Index* from Freedom House. The index is based on an annual survey of media independence and assesses the degree of print, broadcast, and Internet freedom in each country. As mentioned in footnote 26, the indicator is encoded somewhat counter-intuitively, such that higher scores indicate *more* restrictions. The correlation is positive: countries with higher censorship also have higher DPI use among their ISPs, once again in line with the hypothesis.

Meta-indicator: Polity index

The Polity conceptual scheme examines "concomitant qualities of democratic and autocratic authority in governing institutions". It envisions a spectrum ranging from fully institutionalized autocracies, through "anocracies" and all the way fully institutionalized democracies. In effect, this index captures an underlying property that drives government surveillance, privacy legislation and censorship, and hence acts as a meta-indicator in our study. The correlation between the *Polity index* and DPI use is negative, consistent with the other tests of this hypothesis and with the original prediction.

Hypothesis: Stronger <u>copyright industries</u> make it more likely that ISPs will deploy DPI → REJECTED

The rejection of this hypothesis was somewhat of a surprise. Two proxies are used to capture the strength of the copyright industry in countries: *creative services exports as a percentage of total services trade*, and the *software piracy rate*.

High exports of creative services (defined as personal, cultural and recreational services including audio-visual services) points to the presence of strong creative industries. This can determine the lobbying capability of these industries and their power in pushing for regulations, or launching lawsuits that mandate ISPs to deploy DPI (to curb the sharing of copyrighted material online). The observed correlation is negative, rejecting the hypothesis.

High rates of software piracy are also an indication of weak copyright enforcement in a country. Following a similar logic, a weaker copyright industry should result in lower DPI use. The observed correlation is however positive, once more rejecting the hypothesis.

Our interpretation is that in the political stand-off between the copyright industries and the telecommunication companies, the former do not have the upper hand—at least for now. Although this finding is somewhat surprising, it in fact fits well with prior research by Mueller, Kuehn, and Santoso (2012).

The correlation coefficients

As the closing subject, we turn our attention to the Spearman rank correlation coefficients. The coefficients of the significant relationships all lie in the *weak* range (between 0.1 and 0.3), i.e. they show a weak relationship. This should not come as a surprise. It would be more strange if the multi-faceted phenomenon of DPI use by ISPs would be explained to a higher degree via mono-causal relationships. The more important questions, here, are whether the relationships are significant or not, and whether the signs of the coefficients are in the expected directions or not.

5. Discussion and limitations

In this paper we have used a quantitative method to answer our questions about the drivers of DPI. As far as we know, it is the first attempt known to empirically test a number of much-hypothesized relationships. Some of those widely-held ideas found support in the evidence, others didn't.

We also need to acknowledge the limitations of this analysis. First of all, we have to remind readers of the *ecological fallacy,* a familiar problem in the environmental and health sciences: relations that hold at the level of the population need not necessarily hold at the level of individuals in that population. For example, although we observe that lower market concentration is significantly correlated with lower DPI use, this does not in any way hold for Canada or the UK. These countries hold the second and fourth position in terms of lowest market concentration score in our dataset, and both make pervasive use of DPI.

Another important limitation is the reliance on bivariate analysis. To elaborate with an example, we observe that lower competition is correlated with increased use of DPI; so is a high level of censorship. Could it be that the countries with lower competition levels are also the same ones that enact more censorship? In other words, could it simply be that they share a common cause or that one is the by-product of the other? Multivariate analysis and other statistical techniques are be needed to address this matter. We hope to present this in future work.

A second important limitation rises from the fact that while our DPI metric captures the use of DPI only for P2P traffic management, we are using it as a proxy for the wider use of DPI. As discussed in the methodology section we believe that this is a valid proxy. Prior research has revealed that traffic management is the number #1 application of DPI. Furthermore, when DPI is deployed for other use cases, it is also typically used for bandwidth monitoring. This would explain why we observe correlations between different political forces and the DPI score.

A third limitation stems from the crowd-sourced nature of the Glasnost test used to build the DPI-scores. The number of tests performed in 2011 was lower than the previous years, resulting in forty fewer cases for that year in the dataset. This could have an effect on the results. It also means that expanding the research to more countries is not possible. We see no remedy for this limitation until another source for testing the presence of DPI is found.

6. Related work

The use of Glasnost data to look at DPI deployment patterns has been covered in multiple works. These include research papers by Dischinger and his colleagues, websites such as the *Network is Aware Project*[29], and works of other scholars. The media has covered them in many cases, e.g. the New York Times (O'Brein 2011) and Ars Technica (Lasar 2011).

This paper differs in its descriptive findings from the mentioned works by developing the metrics more robustly (for instance by using weighted averages to calculate the country scores). Our main contribution has however been in attaching a battery of independent variables to the Glasnost data to discern the economic and political drivers that push for or inhibit the use of DPI. In doing so, this work builds on qualitative research conducted in the area of DPI and Internet Governance by Bendrath, Kuehn, Wagner and other scholars mentioned in the text.

7. Conclusion

This paper aimed to contribute to the discussions surrounding the use of deep packet inspection technology by answering the following research questions: (1) the extent to which DPI is in active use by ISPs worldwide; (2) identifying which of the economic and political drivers of DPI have a significant correlation with its use. To answer these questions, a DPI score was calculated for 288 Internet service providers in 75 countries.

Regarding the first question, we find that DPI use is widespread. Less than a third of the studied countries showed no or negligent use of DPI in 2011. Of the other two-thirds, half were found to have noticeable use of DPI and the other half showed pervasive use. DPI use has increased in some countries and dropped in others during 2009-2011. On average, however, use of DPI peaked in 2010 and then dropped. This suggests that some ISPs are adopting the technology, while others have stopped using it due to market or political push backs. And finally, current implementations of DPI are less “aggressive” in terms of number of the number users or times of day that throttling takes place than previous years.

Regarding the second research question, we investigated three market drivers and four political ones. Among these, we find that bandwidth scarcity and costs of bandwidth correlates with higher DPI use; as do lower levels of competition. Among the political factors, we find that high levels of governmental censorship and weak privacy protections correlate with higher DPI use. These findings are in agreement with our conceptual model.

Two hypotheses were rejected by the evidence. First, the strength of the copyright industry in a country does not correlate with the amount of DPI use by ISPs. Second, the security performance of ISPs correlates negatively with DPI use, suggesting that network security is not a driver of DPI.

[29] http://deeppacket.info

Acknowledgements

The authors wish to thank Shahab Zehtabchi for his great help and several weeks of effort in mapping more than two thousand Autonomous System Numbers to the actual operating entities across the different countries, making the research possible. This research was partially funded by the Next Generation Infrastructures Foundation and we gratefully acknowledge their support.

References

Asghari, Hadi, Milton Mueller, Michel Van Eeten, and Xiang Wang. 2012. Making Internet Measurements Accessible for Multi-Disciplinary Research: An in depth look at using MLab's Glasnost data for net-neutrality research.

Bendrath, Ralf. 2009. Global technology trends and national regulation: Explaining Variation in the Governance of Deep Packet Inspection.

BEREC. 2012. A view of traffic management and other practices resulting in restrictions to the open Internet in Europe. Body of European Regulators for Electronic Communications.

Dischinger, M., M. Marcon, S. Guha, K.P. Gummadi, R. Mahajan, and S. Saroiu. 2010. Glasnost: Enabling end users to detect traffic differentiation. Paper read at Proceedings of the 7th USENIX conference on Networked systems design and implementation.

Kuehn, Andreas, and Milton Mueller. 2012. "Profiling the Profilers: Deep Packet Inspection and Behavioral Advertising in Europe and the United States." *Available at SSRN 2014181*.

Lasar, Matthew. 2011. Global ISP tracker shows traffic shaping down, but not out. *Ars Technica*, *http://arstechnica.com/tech-policy/2011/10/global-isp-tracker-shows-dpi-use-down-but-not-out/*.

MPI. 2011. Glasnost: Results from Tests for BitTorrent Traffic Shaping. *http://broadband.mpi-sws.org/transparency/results/*.

Mueller, Milton. 2011. DPI Technology from the standpoint of Internet governance studies: An introduction.

Mueller, Milton, and Hadi Asghari. 2012. "Deep packet inspection and bandwidth management: Battles over BitTorrent in Canada and the United States." *Telecommunications Policy* no. 36:462-475. doi: 10.1016/j.telpol.2012.04.003.

Mueller, Milton, Andreas Kuehn, and Stephanie M. Santoso. 2012. "Policing the Network: Using DPI for Copyright Enforcement." *Surveillance & Society* no. 9 (4):348-364.

O'Brein, Kevin J. 2011. Putting the Brakes on Web-Surfing Speeds. *New York Times*, *http://www.nytimes.com/2011/11/14/technology/putting-the-brakes-on-web-surfing-speeds.html*.

Preuschat, Archibald. 2011. KPN Admits To Using Deep Packet Inspection. *Wall Street Journal Blogs*, *http://blogs.wsj.com/tech-europe/2011/05/12/kpn-admits-to-using-deep-packet-inspection/*.

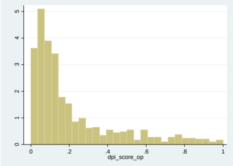
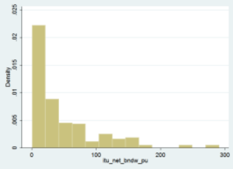
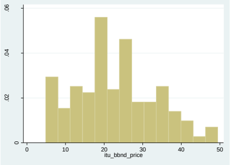
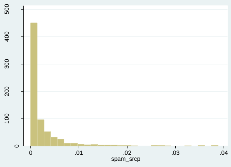
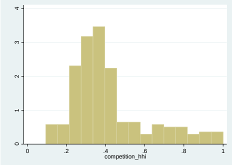
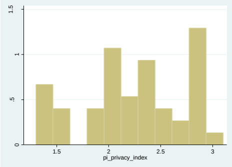
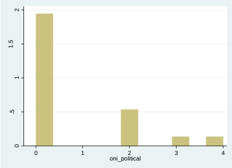
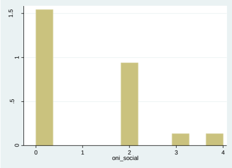
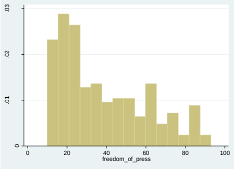
Van Eeten, Michel JG, Johannes M Bauer, Hadi Asghari, Shirin Tabatabaie, and Dave Rand. 2010. "The Role of Internet Service Providers in Botnet Mitigation An Empirical Analysis Based on Spam Data." *Development*:1-31.

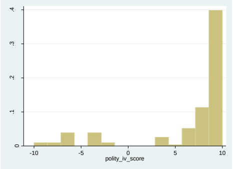
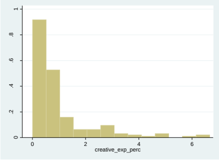
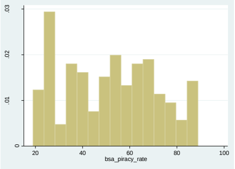
Wagner, B. 2012. "Push-button-autocracy in Tunisia: Analysing the role of Internet infrastructure, institutions and international markets in creating a Tunisian censorship regime." *Telecommunications Policy*.

Appendix

Table of summary statistics and histograms of all variables in the final dataset

For definitions and sources of the variables see the methodology chapter.

Variable	Data points	Range (Mean, SD)	Distribution	
Operator DPI scores *(Own construct, Glasnost data)*	N = 787 75 countries 288 operators 2009-2011	0 – 1 m: 0.21 sd: 0.23		
International Internet bandwidth per Internet user *(ITU, WDI)*	75 countries 2009-2011	0.7 – 965 m: 61 sd: 101		*outliers removed*
Monthly Internet subscription fees *(ITU)*	74* countries 2009-2010 ('10 used for '11) *missing: TW	4.8 – 49.5 m: 24.3 sd: 10.1		
Infected machines per subscriber *(Own construct, spam data)*	N=783 Available for almost all operators 2009-2011	0 - 0.08 m: 0.003 sd: 0.006		*outliers removed*
HHI *(Own construct, TG data)*	74* countries 2009-2011 *missing: MY	0.09 – 1 m: 0.35 sd: 0.16		
Privacy Index *(Privacy International)*	46 countries 2007 (scores copied for all three years)	1.3 – 3.1 (scale of 1-5) m: 2.2 sd: 0.51		
Political censorship on the web *(ONI)*	40 countries one data-point per country in '07-'11 (copied to all years)	0 – 4 m: 0.6 sd: 1.1		
Social censorship on the web *(ONI)*	same as web-political-censorship	0-4 m: 0.8 sd: 1.2		
Freedom of the Press Index *(Freedom House)*	75* countries 2009-2011 *missing: PR,MO	10 – 93 m:33 sd: 20		

Polity Index *(Polity-IV Project)*	70* countries 2009-2010 (‘10 used for ‘11) *missing: HK, IS, LU, PR, TT	-10 to +10 m: 8.0 sd: 4.2	
Creative services exports as % of services trade *(UNCTAD)*	63 countries 2009-2010 (‘10 used for ‘11)	0.0 – 6.7 m: 1.2 sd: 1.3	
Software piracy rate *(BSA)*	73* countries 2009-2011 *missing: MO, TT	19 – 89 m: 46 sd: 19	 .03 .02 .01 0 20 40 60 80 100 bsa_piracy_rate