Mapping Local Internet Control

Hal Roberts and David Larochelle
Berkman Center for Internet & Society
Harvard University
hroberts@cyber.law.harvard.edu
dlarochelle@cyber.law.harvard.edu

Introduction

The Internet is a battleground of control by national governments. That contested control takes many forms, including surveillance, filtering, and prevention (and commission) of cybercrime. To better understand that battleground, it is important to understand how each nation structures the Internet with its borders. One helpful way to understand the structure of national Internets is by mapping autonomous system relationships within each country. Those autonomous systems are the ISPs and other large organizations that are responsible for routing traffic both within the larger Internet and within their own networks and as such act as centers of technical and political control of the Internet.

This paper will describe a method for mapping national networks of autonomous systems, for identifying a small set of autonomous systems that act as points of control for the international routes of each national network, and for measuring the complexity of the networks of autonomous systems within each country. Using these methods, we make several specific findings about the structure of national autonomous system networks. Our primary finding is that across all countries, only a few autonomous systems act as points of control for international traffic. But there are significant differences between autonomous system networks among both countries and regions. China and Russia specifically and Eastern Asian and Eastern European countries generally have dramatically different network structures. China and other Eastern Asian countries are very centralized and very simple – with tens of millions of users per point of control and with Internet users concentrated in only a few of the biggest autonomous systems. Russia and other Easter European countries are much less centralized and much more complex – with only hundreds of thousands of Internet users per point of control and with Internet users scattered through many autonomous systems connected to each other through a much more complex web of relationships.

We have found no evidence that these metrics for national Internet control directly predict whether a country exerts specific kinds of control over its Internet – for instance by filtering content. But they may provide a map for understanding better both how the countries exert control over their networks and how philosophies of political control in the countries have shaped the technical details of their local portions of the Internet. We do not attempt a deep exploration of that relationship here but instead merely propose this as a fertile field for future work that combines computer and social science to understand how countries attempt to exert control over their portions of the Internet.

Background

The Internet is often described as a network of networks. The primary defining characteristic of the Internet protocols is that they connect distinct individual networks with distinct modes of both technical and political control. In one origin story, the Internet was born through an attempt to allow communication between separate networks with separate modes and zones of political control. The first production use of the TCP/IP protocols that underlie the Internet was in 1983 in ARPANET, a research network funded and run by the U.S. defense department. The network had initially been used solely by defense department funded researchers to share computing resources. But since its inception in 1969, the defense department had increasingly grown to use the network for operational military uses in addition to the existing research uses.

By 1983, the defense department had taken over direct management of the network and had become frustrated with the difficulty of enforcing military levels of security on the existing researcher users. So the defense department split ARPANET into two separate networks -- keeping the research users on the existing ARPANET and moving the military users to the new MILNET. (Abbate 1999) To allow users of the two networks to continue to talk to one another, the defense department moved both of the networks from the old operating protocol that ARPANET had been using for fifteen years to the new TCP/IP protocols. The key feature of those new protocols was that they allowed disparate networks to talk to one another – the "Internet" in "Internet Protocol" reflects this support for operating between separate networks. This split of ARPANET into two networks and the resulting switchover to TCP/IP marked the birth of the Internet in the sense that it was the first use of the protocol in an operational network and that ARPANET/MILNET grew – by incorporating other networks – into the larger Internet that we use today. In this origin story, the decision to create the infant Internet was political. It was motivated by the need for different policies of control over two separate but connected networks.

Today's autonomous systems are the descendants of the split ARPANET and MILNET networks.[1] Autonomous systems are the networks that make up the Internet as a network of networks. For an ARPANET machine to send data to a MILNET machine, the ARPANET machine only had to know to deliver the data to the ARPANET gateway. The ARPANET gateway only had to know to deliver the data to the MILNET gateway. And the MILNET gateway was responsible for knowing how to deliver the data to the particular MILNET machine. New networks added to the ARPANET / MILNET core operated in the same way – exchanging data with one another through these gateways. The defining characteristic of this arrangement was that none of the individual networks needed to know anything about how the traffic needed to be delivered in the specific local network. All any network connected to the core ARPANET / MILNET network needed to know was to pass the data to a gateway on the core network, which would deliver it on to the destination network. As the number of connected networks grew, it became cumbersome and inefficient to route all traffic through a set of core networks, so the networks switched to Border Gateway Protocol (BGP). BGP allowed networks to exchange data directly by giving each network the ability and responsibility to announce to its peers for which other networks it was willing to carry traffic.

Many autonomous systems today are Internet Service Providers (ISPs), but many are large companies, universities, and other organizations that essentially act as their own ISPs. An autonomous system is responsible both for determining how traffic flows between machines within its own local network and for passing along to other autonomous systems along BGP advertised paths. BGP is the protocol

[1] Autonomous Systems actually predated the ARPANET / MILNET split, and for specific technical and historical reasons, ARPANET and MILNET actually remained in the same autonomous system after the split. But the combination of the move to TCP / IP during the split and the breakup of the network that remained the core of the Internet marked a key turning point into the specific form easily expanded, heterogeneous network of networks that define autonomous systems today.

through which each autonomous system announces to other autonomous systems which autonomous systems it will carry traffic for. In the figure below, AS #2 announces to AS #3 that it carries traffic for AS #1, then AS #3 announces to AS #4 that it will carry traffic for AS #1 via AS #2:

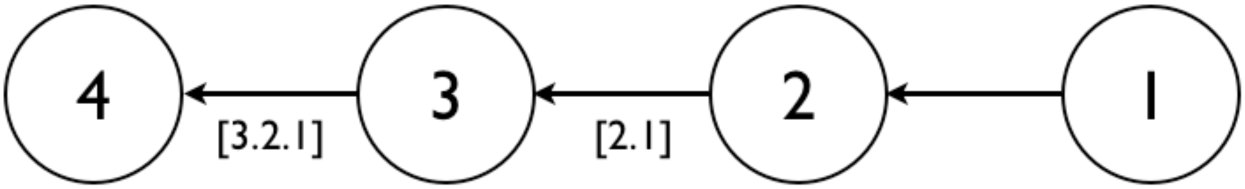


figure 1 – autonomous systems path announcements

In this example, AS #4 knows that it can use the path [3.2.1] to send traffic to AS #1. There are usually several available paths to get from one autonomous system to another, but it is wholly the responsibility of each autonomous system to decide which path to send the data along. So if AS #4 needs to forward data ultimately to AS #1, it may know about both the [3.2.1] and a separate [5.6.7.8.1] path. The most common method for deciding which path to use is simply to forward the traffic to the first autonomous system on the shortest available path, AS #3 in the example above, but it is the responsibility of each autonomous system along the path to make these routing decisions itself. When each autonomous system along the path receives the data, it merely repeats this exercise, forwarding the data on to the first autonomous system along the shortest available path.

This loosely federated architecture allows local networks of vastly differing types – dialup, broadband, wireless, fiber – and different policies – military, commercial, academic, community – to connect to one another easily. The only requirement is that any autonomous system be able to route data within its own network and that it be able to play along in the game of hot potato, passing data for machines in other autonomous systems along these BGP advertised paths. Together, those BGP paths constitute the map of both technical and political control of the Internet. If you want to know how traffic gets from machine A in AS #4 to machine B in AS #3, the technical answer is [3.2.1]. But if you also want to know how to block, surveil, or infect traffic from machine A to machine B, the simplest answer is [3.2.1]. This is the sense in which autonomous systems are key to understanding the Internet as a technical / political system. This distribution of political control is not a mere byproduct of the technical architecture of the network-of-networks Internet. The birth of the Internet as the split of ARPANET into two politically distinct networks was an explicitly political decision – intended to allow distinct modes of political control over the distinct networks.

Just as the U.S. government largely defined the distinct policies of ARPANET and MILNET in the infant Internet, national governments largely define the policies of autonomous system, which are particularly easy targets of regulation as typically large organizations (ISPs, large businesses, universities, etc.). There are strong arguments over what level of control national governments exert over the Internet. Some argue that national governments maintain the same mechanisms to exert control over the Internet as they have over other media (Goldsmith and Wu 2006). Some argue that the Internet fundamentally changes the calculus of control by allowing more people to publish in more ways that are difficult for governments to understand (Kraidy 2006). And others argue that the ultimate role of the Internet in fostering or weakening government control is complex and still not understood

(Chowdhury 2008). But it is clear that the Internet has now become a central site for the battle over the control of information between governments and users. The OpenNet Initiative has tracked extensive filtering of Internet connections in dozens of countries for several years (Deibert et al. 2008). Whether that filtering is having the intended effect of controlling political discourse is open to question, but the extensive efforts by countries to filter the Internet makes clear that the Internet is a key location of the battle for control of social and political discourse.

A few examples of this battle are Iran's geopolitical diversification of its international Internet connections (Cowie 2010); the revelation by AT&T engineer Mark Klein that the U.S. National Security Agency was surveilling Internet traffic at a major U.S. Internet backbone (Klein 2006); the dismantling of the cybercriminal Russian Business Network (Bizeul 2007); and a Pennsylvania law requiring consumer ISPs to block access to illegal pornography (Zittrain 2003). All of these examples center on autonomous systems. Iran recently added a new connection to the Internet through Russia to add to its existing international connections through U.A.E. and Turkey. In this case, the answer to the question “How can Iran gain more control over its connection to the wider Internet?” was a set of paths that add greater geographic and political diversity to the handful of autonomous systems that connect Iran to the wider Internet, making it more difficult for any one other country to control its Internet connection to the rest of the world. Similarly for vast amounts of not only domestic U.S. but also international traffic, the answer to the question “How can I monitor all Internet traffic” is a path in which an AT&T autonomous system sits in the middle, so installing a black box in the closet of AT&T allows monitoring of those vast amounts of Internet traffic. For the dismantling of the Russian Business Network (RBN), the answer to the question, “How do we stop this criminal organization from running malware ISPs?” was a long series of BGP paths that established how the RBN malware ISPs were routing traffic to the rest of the Internet.

And for the Pennsylvania legislature, the answer to the question “How can we stop Pennsylvanians from accessing illegal pornography on the Internet?” was consumer ISPs, typically the last autonomous systems in the paths that Pennsylvanians use to access the Internet – the .1 in [3.2.1]. Jonathan Zittrain uses this Pennsylvania law as an example of how law can operate on specific “points of control” in the Internet – in this case he argues that the Pennsylvania law represented a new effort to exert control at the point closest to the consumer (Zittrain 2003). In Zittrain's version of points of control, the relevant points are the ISPs on each side of a given route, and everything in the middle is a “cloud” that implicitly contains no clear points of control. But that cloud includes only about thirty thousand active autonomous systems worldwide. Those thirty thousand autonomous systems represent a relatively small set of points of control over the billions of individually connected computers (and people) on the Internet. Even assuming control is evenly distributed among those thirty thousand autonomous systems, those autonomous systems funnel of control from the nearly infinite end points of the Internet to the much smaller number of autonomous systems.

But even though the Internet is theoretically structured as a random game of hot potato with no particular autonomous system serving as the center of the network, in practice a very small portion of those autonomous systems carry the traffic for a disproportionate number of routes on the Internet (Huffaker and claffy 2009). Data flowing from a computer in China to a computer in the U.S. will likely travel through one of a handful of Chinese autonomous systems connecting China to the rest of the world and one of a few U.S. autonomous systems connecting the U.S. to the rest of the world. This concentration of traffic on only a few autonomous systems per country further amplifies the technical / political role of those autonomous systems and their resulting role as the loci of national control over the Internet. The key finding of this paper, described in detail below, is that this concentration of autonomous systems holds within individual countries as well as for the Internet as a whole. In any country, a much smaller subset of all of the country's autonomous systems act as a chokepoint for control of the larger set of autonomous systems and for the much larger set of people using the network.

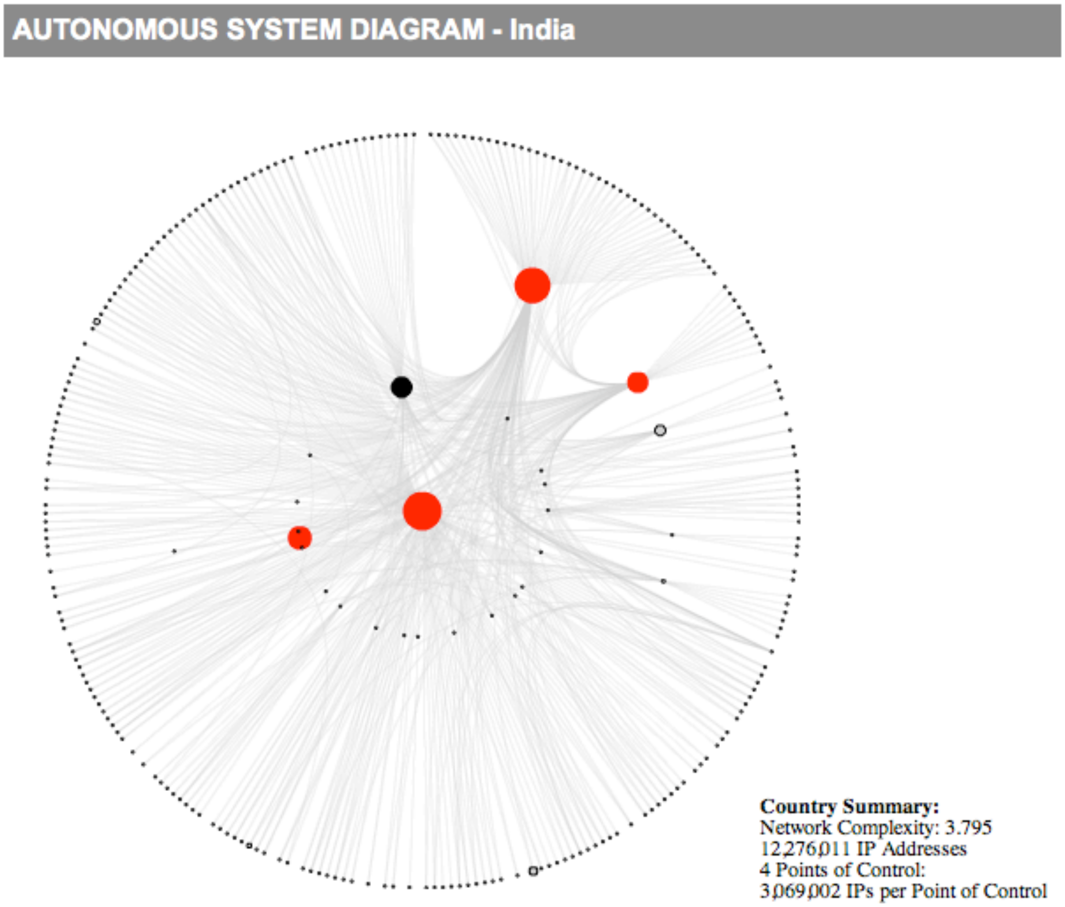
There is existing work that examines the policy implications of the geographic properties of Internet topology. The most current and comprehensive effort to map autonomous system topology globally has been the Cooperative Association for Internet Data Analysis (CAIDA). They have used collections of trace routes to generate global maps of autonomous systems by geography and by number of direct connections to other autonomous systems (Huffaker and claffy 2009). Their maps show that a small number of mostly U.S. autonomous systems have a disproportionate share of direct connections to other autonomous systems. Josh Karlin et al. have analyzed autonomous system topology between countries to determine which countries have the most influence over the international traffic (Karlin 2009). They determined that the United States, Great Britain, and Germany have a large amount of influence over international traffic because a large number of international routes flow through autonomous systems in those countries. The IXmaps project maps the political geography of specific routes by geo-locating the position of each router along the route between two computers and annotating the routers with information relevant to data control at each router (for instance, whether the router is a known NSA surveillance location).

Our work in this paper builds primarily on work inferring autonomous system relationships led by Xenofontas Dimitropoulos at the Cooperative Association for Internet Data Analysis (CAIDA) to infer consumer, peer, and sibling relationships between autonomous systems based on BGP announcements (Dimitropoulos et al. 2007). A consumer relationship is one in which an autonomous system is paying another autonomous system to route traffic to it from the rest of the Internet. A peer relationship is one in which an autonomous system agrees to exchange traffic with another autonomous system, but the only traffic exchanged is traffic directly from one of the peering autonomous systems or one of their consumers. A sibling relationship is between autonomous systems owned by the same entity and as such may include a broad range of different routing agreements negotiated privately. The BGP data is gathered by the University of Oregon Route Views project, which coordinates a collection of servers at various places around the Internet that listen to and collect BGP announcements. CAIDA analyzes these BGP announcements to infer relationships between autonomous systems. For instance, if it consistently sees paths like [3.2.1] and [4.3.2.1] and [5.2.1], it infers that AS 2 is a provider of AS 1. There are some important limitations to the Route Views and autonomous system relationships data that we will discuss in detail in the limitations below, but the most important to note here is that the data set undercount peer relationships (missing over 60% of them according to validation performed by CAIDA) especially among small autonomous systems, making it difficult to tell to what degree autonomous systems at the edge of a country's networks are exchanging directly data among themselves. CAIDA has also used the relationships data to rank autonomous systems based on the number of autonomous systems within recursive consumer relationships with the autonomous system (CAIDA 2010). Bradley Huffaker at CAIDA has used the autonomous systems relationships data to show that the U.S. autonomous systems include a hugely disproportionate number of the world's IP addresses compared to its share of either world population or world GDP (Huffaker 2003).

To better understand how and where countries exert control over the Internet, and to test further the questions of how the Internet strengthens and how it weakens government control of information, it is helpful to understand how autonomous systems structure themselves (or are structured) within individual countries and which autonomous systems in particular are at the center of each nation's Internet. We use three methods to answer these questions: we draw a network graph of the consumer/provider relationships between autonomous systems within the country, we calculate the set of points of control for each country as the minimum set of autonomous systems necessary to connect to 90% of the IP addresses in the country, and we calculate the normalized complexity of the network by considering the number of autonomous systems in the country and the number of IP addresses in autonomous systems at the edge of the country's network. According to these metrics, nations differ dramatically in the infrastructures they use to run, and therefore control, their networks. A primary finding is that within every country a small set of autonomous systems act as points of control for the larger network. However, the number of points of control relative to the number of Internet users and the overall complexity of the network differs dramatically between countries. These metrics do not predict whether countries are exerting specific forms of control over their networks; for example, countries that have national networks with fewer autonomous systems acting as points of control are no more likely to filter their Internet traffic. But the specific identification of the core autonomous systems within each country can identify points of focus for studies of Internet control by identifying where the control is most likely to happen within each country's network – in which ISPs the government is most likely to install filtering and surveillance devices, which autonomous systems are most vulnerable to attack during time of international crisis, in which countries malware communities will find it easiest to hide among the multitude of autonomous systems, and so on. And these metrics point to larger stories about how social, political, economic, and historical factors impact the structures of national networks (and how those factors are in turn impacted by the structure of the network).

Network Maps

Our first approach to understanding the structure of the autonomous system networks within each country is to visually map the network of autonomous systems within each country according to the CAIDA autonomous system relationship data set (CAIDA 2009). These maps visualize the complex (or sometimes not complex) structure of all of the various network paths discussed above within a given country.



Country Summary:
Network Complexity: 3.795
12,276,011 IP Addresses
4 Points of Control:
3,069,002 IPs per Point of Control

figure 2 – network map of India

In this map, each circle represents an Indian autonomous system, each line represents a consumer / provider relationship between two autonomous systems, the black dot represents the Internet outside of India, and the red dots represent the autonomous systems that are the points of control for India. To generate these maps, we assign a country to each autonomous system.[2] For each country, we merge all autonomous systems not in that country into a single "Rest of World" node that represents connections to the rest of the Internet. We determine the number of connected IP addresses for each of the country's autonomous systems – the number of IP addresses in the autonomous systems and its consumers. And we determine the points of control for the network – the autonomous systems that connect to at least 90% of the IP addresses in the country (both of these metrics are described in detail below). We convert the relationships between the country's autonomous systems into a directed graph, with consumer to provider relationships acting as child to parent links. Finally, we map the resulting graph using a circular layout.[3] Autonomous systems with more consumers are closer to the middle of the graph, the size of each node is determined by the number of connected IP addresses for the autonomous

[2] Autonomous systems and IP address blocks are registered through one of five regional Internet registries. These registries maintain the authoritative lists of the autonomous system number and the IP address blocks associated with each autonomous system. They also keep the country in which each autonomous system was registered, which we use to determine the country of each autonomous system. This country of registry generally correlates to the physical and political home of the autonomous system, but there are exceptions, for instance some old African autonomous systems were registered in Israel and other countries before the creation of the African registry. We use this country of registration from the appropriate regional Internet registry via the Team Cymru service.

systems.

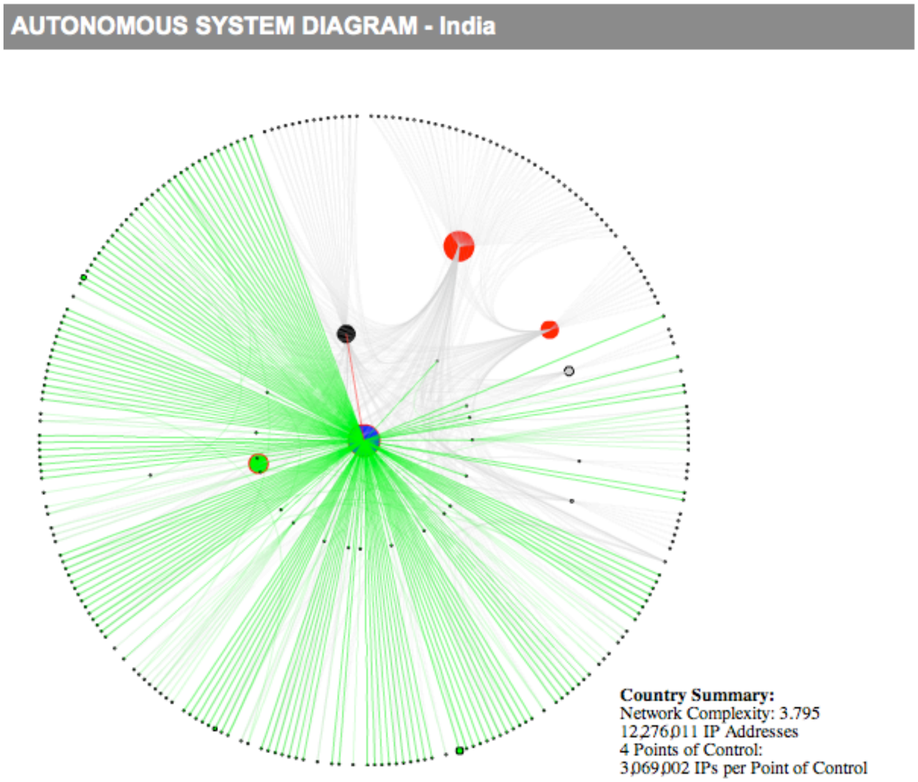

AUTONOMOUS SYSTEM DIAGRAM - India
Country Summary:
Network Complexity: 3.795
12,276,011 IP Addresses
4 Points of Control:
3,069,002 IPs per Point of Control

figure 3 – network map of India with providers and consumers highlighted

In the interactive form above, we also allow the user to click on a given autonomous system and find out either one of it paths to the Internet or all of its providers and consumers. In this same map of India, the provider (red) and consumer (green) links of the center autonomous system are highlighted.

The map of India is mostly representative of our larger results, with the most interesting finding being that a mere 4 autonomous systems act as points of control for nearly all of the 12 million IP addresses in India. The radiating green lines show the wide reach of the single biggest autonomous system.

We have generated network maps of all countries with more than 25,000 total IP addresses, all of which are available at http://cyber.law.harvard.edu/netmaps along with the full code and data needed to reproduce the results in this paper and on the site. Below are the maps for China, Russia, South Korea, and Ukraine that serve as samples of the generated maps and visualize the regional differences in network structure between Eastern Asia and Eastern Europe.

[3] To draw the maps, we use the CircleLayout method of the Flare Toolkit.

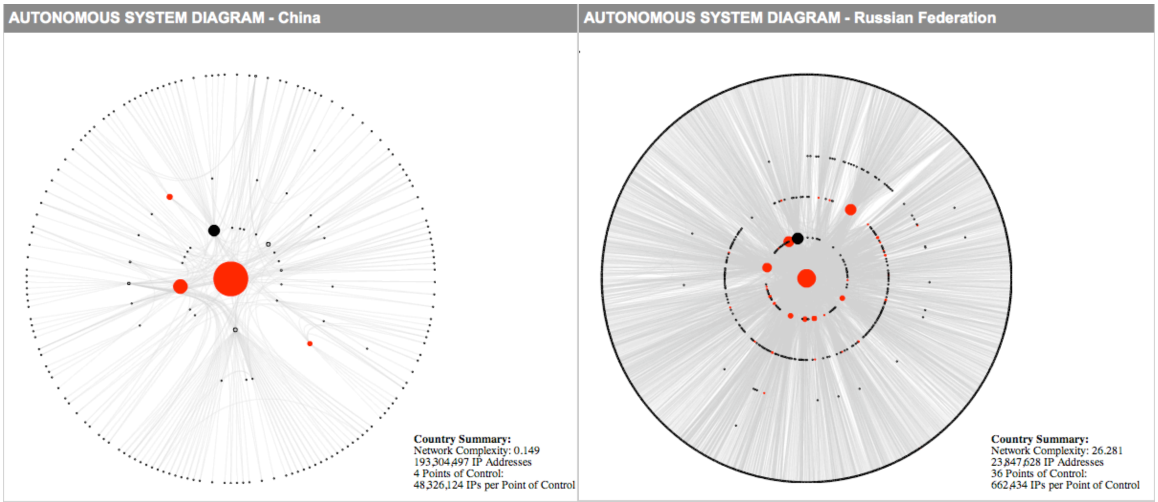

AUTONOMOUS SYSTEM DIAGRAM - China
Country Summary:
Network Complexity: 0.149
193,304,497 IP Addresses
4 Points of Control:
48,326,124 IPs per Point of Control
AUTONOMOUS SYSTEM DIAGRAM - Russian Federation
Country Summary:
Network Complexity: 26.281
23,847,628 IP Addresses
36 Points of Control:
662,434 IPs per Point of Control

Figure 4 – network maps of China and Russia

The above maps visualize the striking difference in the way that China and Russia respectively structure control of the Internet within their borders. China, with 193 million IP addresses, has a dramatically simpler network of autonomous systems than Russia, with only 23 million IP addresses. This difference bears out in the number of points of control as well, with only 4 points of control for China compared to 36 for Russia.

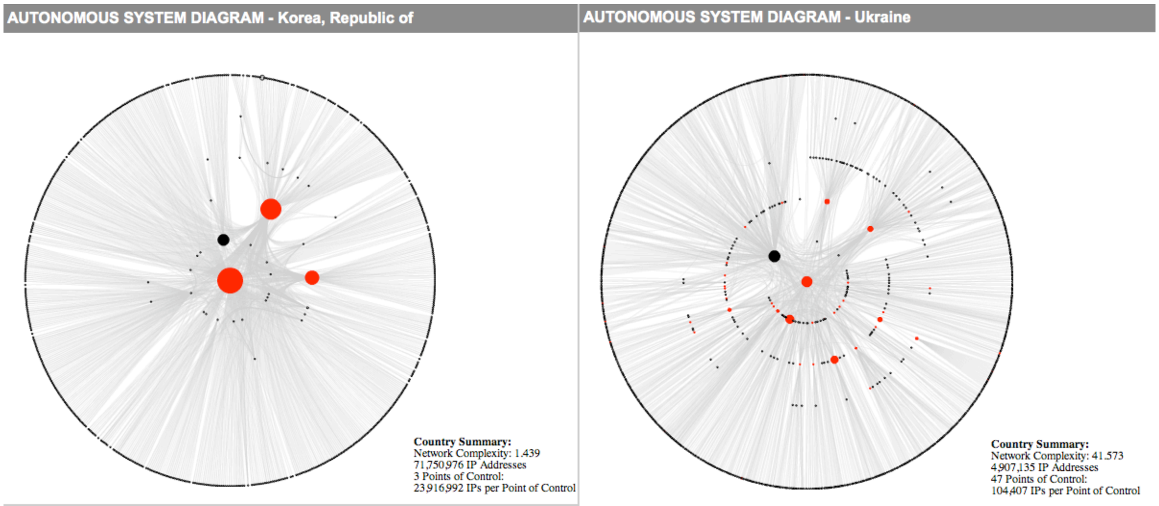

AUTONOMOUS SYSTEM DIAGRAM - Korea, Republic of
Country Summary:
Network Complexity: 1.439
71,750,976 IP Addresses
3 Points of Control:
23,916,992 IPs per Point of Control
AUTONOMOUS SYSTEM DIAGRAM - Ukraine
Country Summary:
Network Complexity: 41.573
4,907,135 IP Addresses
47 Points of Control:
104,407 IPs per Point of Control

figure 5 – network maps of South Korea and Ukraine

These same differences are present but less striking in a comparison of South Korea, with 71 million IP addresses, to Ukraine, with only 5 million IP addresses. South Korea's structure is visibly more dense than China's but still comparable to Ukraine, despite the fact that South Koreas has many times more IP addresses than Ukraine. And despite the comparable density of the networks, South Korea has only 3 points of control to Ukraine's 47, reflecting the fact that South Korea's IP addresses are mostly concentrated in that handful of core autonomous systems.

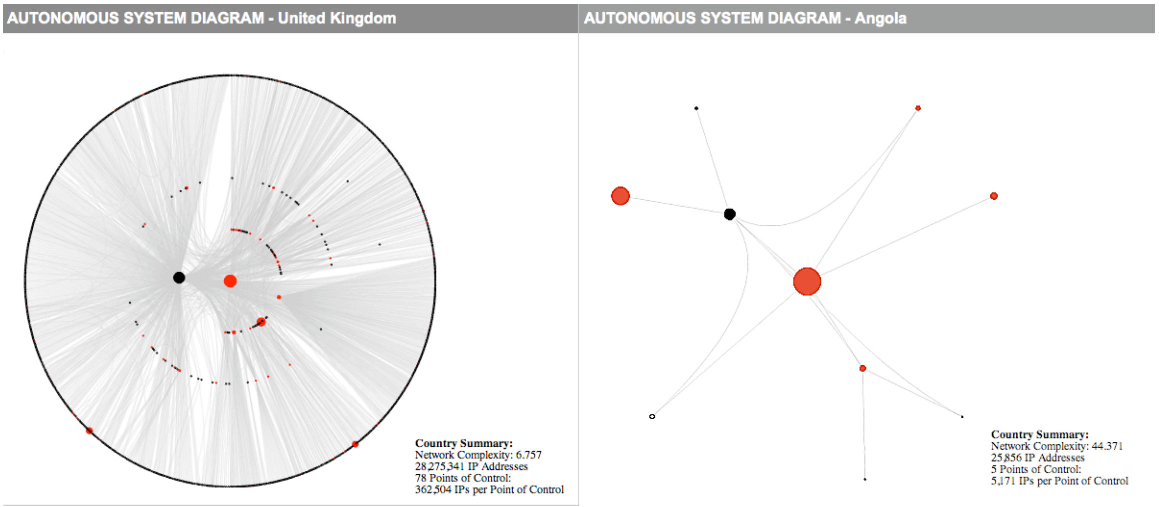

AUTONOMOUS SYSTEM DIAGRAM - United Kingdom
Country Summary:
Network Complexity: 6.757
28,275,341 IP Addresses
78 Points of Control:
362,504 IPs per Point of Control
AUTONOMOUS SYSTEM DIAGRAM - Angola
Country Summary:
Network Complexity: 44.371
25,856 IP Addresses
5 Points of Control:
5,171 IPs per Point of Control

figure 6 – network maps of United Kingdom and Angola

Finally, above are two maps from non-Eastern Asia / Eastern European countries. The U.K. is the network with the most points of control of any country (excluding the U.S. for reasons we discuss below), but it only has 78 points of control to Angola's 5 (for a ratio of 15.6:1) even though it has 28 million IP addresses to Angela's 25 thousand (for a ration of 1120:1). The exponentially smaller number of points of control for the size of the U.K. network makes those point of control autonomous systems key locations for the exertion of technical and political control over the U.K. network.

Points of Control

The purpose of the points of control metric is to determine both the proportion of control potentially executed over Internet routes by any given autonomous system in its country and the smallest set of autonomous systems that have the potential to control virtually all (90%) of the traffic within a given country. Less formally, these points of control are the autonomous systems that appear in most of the networks paths discussed in the Background section above. The points of control are almost always the answers to the kinds of questions mentioned above – Where might a state filter its connection? Where might it surveil its citizens? Where might a malware host hide its connection to the larger the network?

We define the connected IP addresses for a given autonomous system as the set of IP addresses within either the autonomous system itself or within the connected IP addresses of any of its consumers. This definition is recursive, so the connected IP addresses for an autonomous system includes not only its consumers' IP addresses but also its consumers' consumers' IP addresses. This definition of connected IP addresses roughly models the set of IP addresses whose traffic follows routes that are likely to flow through the given autonomous system to the rest of the Internet. We call the set of IP addresses within the autonomous system itself the direct IP addresses to distinguish them from the connected IP addresses.

To generate the direct IP addresses for each autonomous system, we lookup the IP blocks associated with the autonomous system using the CAIDA RouteViews Prefix to AS Mappings data set.[4] To

[4] The CAIDA Routeviews Prefix to AS Mappings data set returns mapping of IP address prefixes, such as 18.0.0.0/8, which represents all IP addresses beginning with 18. We translate those prefixes into the number of possible IP addresses, but in some cases prefixes conflict with each other. For example, one autonomous system might include

generate the connected IP addresses, we recursively traverse up the tree of autonomous systems for the country, adding the number of IP addresses of each autonomous system both to its own connected IP address count and to those of its provider.[5] If an autonomous system has more than one provider, we add to each provider the connected IP addresses of the consumer divided by the number of providers. This is a very rough estimation of multiple provider relationships which we discuss in detail in the limitations section below.

We define the points of control as the smallest set of autonomous system nodes whose connected IP addresses include 90% of a country's total direct IP addresses. We calculate the points of control using a simple greedy algorithm. We start with the autonomous system with the most connected IP addresses as a point of control. We repeatedly find the autonomous system node that will most increase the number of connected IP addresses and add it to the points of control. We continue until the points of control are connected to 90% of the country's IP addresses. When calculating the number of connected IP addresses for each addition to the points of control set, we avoid double counting connected IP addresses – if a provider and consumer are both in the points of control set, the consumer's connected IP addresses are only counted once.

The following table shows the ten countries with the most points of control, meaning that these are the countries in which control over the network is distributed among the largest set of autonomous systems.

country	points of control	autonomous systems	IP addresses	region
United Kingdom	78	1,243	28,275,341	Northern Europe
Ukraine	47	1,165	4,907,135	Eastern Europe
Sweden	44	297	6,590,154	Northern Europe
Netherlands	40	337	14,104,594	Western Europe
Russian Federation	36	2,374	23,847,628	Eastern Europe
Bulgaria	32	345	2,641,920	Eastern Europe
Czech Republic	30	244	4,036,267	Eastern Europe
Switzerland	28	337	9,995,020	Western Europe
Italy	26	459	14,022,848	Southern Europe
Germany	24	950	75,943,183	Western Europe

table 1 – top ten countries by points of control

As mentioned above, the U.K. has the most points of control of any country of the in the set of countries we consider, and only 78 of its 1,243 autonomous systems act as points of control for its

18.0.0.0/8 and another might include 18.100.0.0/16. In those cases, the autonomous system with the most specific prefix is assigned the number of IP addresses in the more specific prefix (18.100.0.0/16) and the autonomous system with the less specific prefix (18.0.0.0/8) is assigned the number of IP addresses in its prefix *minus* the number of IP addresses in the more specific prefix. After correcting for these prefix conflicts, the number of direct IP addresses in each autonomous systems represents the total IP addresses unique to that autonomous system for each prefix listing. So we are able to determine the number of IP addresses in a set of autonomous systems by adding together the direct IP addresses in each autonomous system in the set.

[5] This traversal requires that the country network graph not include cycles. Cycles are rare within country networks – most countries did not have cycles and the few countries with cycles only have a small number. In these rare cases, we modify the network graph to break the cycle by finding the node within the cycle that is furthest away from an international gateway. We remove provider links from this furthest node to other nodes in the cycle – making the node only a customer and not a provider to the other cycle nodes. For example, if node #1 provides service to node #2 which also provides service to node #3 and node #3 provides service to #1, there would be a cycle. If #1 was connected to an international gateway, #3 would be the furthest of the three from the rest of the world so the link between #1 and #3 would be removed. Removing a single link to an edge node is unlikely to alter the points of control or the network complexity.

network. No other country has more than 50 points of control. This low limit on points of control for even large countries confirms our hypothesis that only a few points of control connect the vast majority of the network not only globally but also in each individual country. The full results for the points of control metric for all countries, along with the complexity metric defined below, are available as the appendix of this paper.

However, there are significant differences in the number of points of control between countries. Countries with more IP addresses are likely to have more points of control, but only slightly so (r = 0.17). The most interesting differences are between the Eastern Asia and Eastern Europe regions, which provide statistical support for the visual differences in the maps from the two regions shown above:

region	avg points of control	total ASs	total IP addresses
Western Asia	3.27	714	23,373,952
South Central Asia	3.56	595	17,832,491
Central America	4.66	287	21,649,128
Southern Africa	4.98	83	13,152,896
Eastern Asia	5.67	1,602	410,837,433
South America	7.57	877	51,960,817
South-Eastern Asia	8.61	847	25,783,296
Australia & New Zealand	9.55	813	33,453,528
Southern Europe	15.44	1,149	43,623,766
Northern America	16.97	733	41,426,945
Western Europe	21.96	2,473	143,329,085
Eastern Europe	28.34	6,160	61,428,342
Northern Europe	42.64	2,258	67,010,459

table 2 – average points of control by region

The above table compares the average points of control for all geographic regions with at least 10 million total IP addresses. To generate the average points of control for each region, we weight the points of control of each country in the region by its proportion of the region's total IP addresses. Eastern Asia is the major outlier in this list because of its low number of points of control compared to its huge number of IP addresses. Excluding Eastern Asia countries from the set, the total IP addresses within a country correlate much more strongly to the points of control (r = 0.45). The relatively high number of points of control in Eastern Europe confirms the differences shown in the network maps above but is at least somewhat explainable by its relatively high number of IP addresses compared to all non-Eastern Asia regions. Note also that Northern Europe's high average points of control is explained entirely by the U.K., which has the most points of control of any country and has far more IP addresses than any other country in Northern Europe. If we remove the U.K. and Russia (the country with the most IP addresses and points of control in Easter Europe) from their respective regions, the points of control change to 23.47 for Eastern Europe and 16.82 for Northern Europe.

Network Complexity

The points of control metric applies only to control of data as it flows through the network. It does not apply to questions of who connects to the network. An autonomous system that sits in the middle of network routing traffic for many other autonomous systems has the ability to read and edit the data itself as it passes through, but it cannot directly tell who sent that data or control who gets to connect to the network. That control over connection to the network (which can take the form of either allowing or blocking access or merely watching who is connecting) is held by the autonomous system to which the client directly connects. In some cases, the client (which could be an end user machine or a server) connects directly to one of the core point of control autonomous systems, but in other cases the client connects to an autonomous system at the edge of the network that routes its traffic through one of those point of control autonomous systems. Two political questions in particular that this model of complexity helps to answer are: Where in the network might a malware host be hiding? Where in the network did a user connect to a particular IP address (and what was the offline identity of that user)? For more complex networks, the answers to these questions are a larger set of potential autonomous systems.

The purpose of the network complexity metric is to determine the complexity of controlling who connects to the Internet within a given country, with the assumption that it is more difficult to control who connects to a network that has more autonomous systems in general or whose users connect further away from the core of the network. We use the following equation for complexity:

$$C = (AS / I) * \sum[CI(a) / I]$$

where:

- C = the complexity score for the country
- AS = the total number of autonomous systems for the country
- I = the total number of IP addresses in the country
- $\sum$ = the sum for each autonomous system in the country
- $CI(a)$ = the connected IP addresses for a given autonomous system

We are not proposing this metric as a theoretical measure of network complexity, but rather as a specific way of measuring the complexity of connecting to a national network of autonomous systems. We consider a country's network of autonomous systems more complex if it has more autonomous systems per IP address (and therefore more places through which a given user may connect) or has more of its IP addresses located away from the core of the network (and therefore each user is potentially routed through more providers to get to the Internet). The two halves of the above equation each directly models one of these factors: (AS / I) models the number of autonomous systems per IP address and $\sum(CI(a)) / I)$ models the degree to which direct IP addresses are located at the edge of the network.[6] We include the total number of IP addresses as a divisor in both sides of the equation to normalize the score for the amount of Internet usage in the country, so that we can meaningfully compare the complexity of large and small countries.

This complexity metric does not help answer the question of whether autonomous systems in general make it easier to control who connects to the Internet. It is only meaningful to help compare the complexity of controlling Internet connections between difference countries and regions. The following table lists the average network complexity by region:

[6] For example, consider a simple network in which AS1 is a provider of AS2. If AS1 has 2 direct IP addresses and AS2 has 1 direct IP address, $\sum[CI(a)) / I]$ = (3 / 3) + (1 / 3) = 4 / 3. If AS1 has 1 direct IP address and AS2 has 2 direct addresses, $\sum[CI(a)) / I]$ = (3 / 3) + (2 / 3) = 5 / 3. The second example results in a higher complexity score because it has more addresses at the edge of its network.

region	complexity	total ASs	total IP addresses
Eastern Asia	0.66	1,602	410,837,433
Southern Africa	0.97	83	13,152,896
Central America	1.69	287	21,649,128
South America	2.51	877	51,960,817
Western Europe	2.56	2,473	143,329,085
Northern America	3.35	733	41,426,945
Southern Europe	3.86	1,149	43,623,766
Western Asia	4.19	714	23,373,952
Australia & New Zealand	4.53	813	33,453,528
Northern Europe	4.98	2,258	67,010,459
South Central Asia	5.49	595	17,832,491
South-Eastern Asia	5.66	847	25,783,296
Eastern Europe	20.12	6,160	61,428,342

table 3 – complexity by region

As with points of control, we weight the complexity of each country by its proportion of the total IP addresses within its region. And as with points of control, Eastern Asia and Eastern Europe are at the two opposite ends of the spectrum. In this case, Eastern Europe is the big outlier, with nearly four times the complexity of any other region. This means not only that Eastern European countries have many more points of control than Eastern Asian countries and therefore exert control over the flow information through a much broader range of actors, but also that controlling network access at the end points is also much more complex because clients in Eastern Europe connect through much higher number of autonomous systems.

These findings about the high complexity of Eastern Europe are especially intriguing given the history of cyber crime in the region. In particular, David Bizeul wrote a report in 2007 that detailed how the Russian Business Network, one of the world's largest cyber criminal organizations, had been providing bullet proof malware hosting services in Russia by carefully constructing a hugely complex system of autonomous system relationships to shelter its malware host autonomous systems (Bizeul 2007). Each of those malware autonomous systems connected to the Internet through several different autonomous systems that also acted as local ISPs, essentially providing laundering for the connection of the malware host autonomous systems. Each of those laundering autonomous systems connected to the wider Internet through several different legitimate autonomous systems, making it very difficult to discover (and therefore cut) the connections between the malware autonomous systems and the rest of the Internet. The very high relative complexity of the Russian and Eastern European networks seems likely to make creating these complex, control resistant structures easier. Along these same lines, the fact that Nigeria has the fifth most complex network of any country is intriguing given its well known history as a prominent host of phishing and 419 scams.

Limitations

The large size and decentralized nature of routing on the Internet may make exact measurement of Internet routes impossible, so we have to settle for best efforts at measuring the Internet using the best available data. The analysis in this paper is primarily based on BGP path announcements collected by the Route Views project and analyzed by the CAIDA project. The Route Views project collects BGP announcements on about a dozen routers in various locations around the world that act as core exchange points of the Internet. The distributed nature of BGP announcements means that there is no authoritative source of all of them, so the only way to collect them is simply to setup listeners in as many places as possible to catch as many as possible. The most important limitation of the resulting data is that it misses most peer relationships. The definition of peer relationships is that they are not advertised beyond the two peering autonomous systems, so the only way to discover most peer relationships is to listen directly to announcements in the tens of thousands of autonomous systems at edges of the Internet, rather than just to announcements in the core of the Internet.

In their autonomous systems relationships paper, Dimitropoulos et al. validate their inferred relationships against relationships surveyed from a sample of autonomous systems. They find that they only discovered 38.7% of the surveyed peer relationships. Because of this large underreporting of peer relationships, we only consider consumer-provider relationships in this paper. However, we think that international peer relationships between autonomous systems that are not among the points of control are rare, so at a minimum the findings in this paper apply to international traffic.[7] We suspect that in most countries they will apply to traffic in the country as well because the peer relationships increase the interconnection between the core points of control autonomous systems as much as they do the interconnection between autonomous systems at the edges of each country's network. But more work is necessary to quantify the effect of peer relationships on this work.

The autonomous systems relationships data set only infers relationships between entire autonomous systems, but BGP paths include specific IP address prefixes for the origin autonomous system. For simple consumer / provider relationships, we can infer that all of the IP addresses registered by the consumer autonomous system are routed through the single provider. But for a consumer with multiple providers, we can only guess from the autonomous systems relationships data set which provider is carrying traffic for which IP addresses registered by the consumer. We tested a range of hypothetical routing scenarios for their effect on the relative complexity and points of control of each country and found very little effect on our complexity and points of control metrics.[8] So even though our maps may not hold for a given consumer with multiple providers, the larger metrics hold for each country (and in fact are strengthened in some cases, since the biggest outlier in these correlations was that Russia becomes much more complex in some scenarios).

Another limitation is that we do not include the U.S. in our metrics, mostly because we do not have reliable data on how many IP addresses are being used by each U.S. autonomous systems. For all other countries, we use the IP addresses allocated to the autonomous system as an analogue of IP addresses used. But the U.S. has a much higher number of unallocated IP addresses than any other country, with

[7] We generated the percentage of IP addresses in each country within non-point of control autonomous systems that had peer relationships with foreign autonomous systems. We found only five countries (Netherlands, Austria, Germany, South Africa, United Kingdom) plus the special EU autonomous system region had greater than 5% of their IP addresses peered to foreign autonomous systems and only two of those above 10%. Assuming that the data set is missing 60% of all peer relationships and multiplying the corresponding percentage by 2.5, we still only found seven countries plus the EU greater than 10%. It is still possible that the factor of underreporting is greater than 2.5, so this data about the role international, non point of control peering relationships is only suggestive.

[8] We regenerated the network complexity and IP addresses per points of control numbers for each country under three models for multiple provider consumers: a minimum complexity model in which all IP addresses from a consumer are routed through the provider with the largest number of relationships, a proportional model in which each provider routes an equal share of the IP addresses of the consumer, and a maximum complexity model in which each provider routes all of the IP addresses of the consumer. The correlations between plots of complexity and of IP addresses per points of control for each of the models against the others yielded an R^2 of greater than 0.8 in all cases. We could have generated the specific provider / consumer IP prefix mappings by regenerating the entire autonomous system relationship data ourselves from the Route Views data, but we chose not to given the strength of these correlations.

around 2 billion IP addresses allocated for only about 230 million Internet users.[9] In most other countries, the ratio of IP addresses allocated to Internet users is about 2:3. Because we have no way of even guessing the number of IP addresses for each U.S. autonomous system, we have no way of generating the direct IP addresses for each autonomous system, upon which all of our metrics depend. We also do not include any country with less than 25,000 total IP addresses because the relationships are so sparse in those countries that gross errors are much more likely.

Conclusions and Further Work

Autonomous systems are a key battleground in the fight for control of the Internet by national governments. The terrain of autonomous systems differs widely among countries. Through our analysis of network centrality and complexity, we have taken a first step toward a greater understanding of this terrain and its variation between countries. Only a few autonomous systems act as points of control for the networks of even the biggest countries, but countries and regions differ significantly in the structure, centrality, and complexity of their networks. China and Russia specifically and Eastern Asian and Eastern European countries generally have dramatically different network structures. China and other Eastern Asian countries have many fewer points of control even controlling for their larger size and much less complexity than other regions generally and Eastern Europe specifically. Likewise, Eastern Europe sits at the other end of this spectrum, with more points of control and much more complexity than other regions generally and Eastern Asia especially.

It is tempting to try to use these differences to explain the fact that China strongly filters its network but Russia does not filter or filters its Internet much more weakly (Deibert et al. 2008). If Russia has to work with 36 autonomous systems to keep its network filtered and China has to work with only 4, it seems logical that China is more likely to filter. In practice, we found no correlation between either of the points of control or network complexity metrics and whether or how strongly a country filters the Internet because the motivation to filter and the resulting practice of filtering is a complex combination of technical, political, social, economic, and historical factors that cannot compress to a single measure of network structure.[10]

As with filtering, there is a temptation to use these metrics to predict whether or not a country will exhibit problems with computer security, for instance one might guess that malware hosting or intrusion attempts would be more likely from within countries with more complex networks. Indeed, the high complexity of Russian, Nigeria, and Eastern Europe is particularly noteworthy given their history of cybercrime. However, we could not find a systemic link between cybercrime and more complex networks. As with the practice of filtering, these security issues do not boil down to one simple network metric, however, and so we cannot use the metric to predict, for example, whether a given country will host more malware than another country. Still by providing insight into the environment in which cybercrime occurs, network complexity may be useful for better understanding cybersecurity.

We think that these metrics will provide fruitful starting points for many kinds of research into the technical-political nature of networks. The different network structures of Eastern Asia and Eastern

[9] According to the CAIDA IPv4 BGP Geopolitical Analysis at http://www.caida.org/research/policy/geopolitical/bgp2country/ the U.S. autonomous systems are allocated 62% of all available IP addresses.

[10] We tested our metrics against data from the Open Net Initiative on the filtering behavior of countries.

Europe beg the question of why those particular regions (and the particular countries in the regions) chose to structure their networks in the way that they did. Why did China as a society choose a network in which control over their vast national networks is concentrated in the hands of a very small number of entities? Indeed, was this a conscious choice at all or merely a result of the complex interaction of social and technical forces? Why did Russia as a society choose a network in which no one entity (other than the national government) has control over a significant portion of the network? Are the different structures of the networks merely the locked in result of decisions made at some critical point in the growth of the Internet in each country, or do they represent ongoing decisions by the societies about how to control their networks (and how their networks control them)?

Works Cited

Abbate, Janet. *Inventing the Internet*. MIT Press, 1999.

CAIDA. "CAIDA : research : topology : rank_as." <http://www.caida.org/research/topology/as_core_network/>. accessed 28 Feb 2010.

Chowdhury, Mridul. *The Role of the Internet in Burma's Saffron Revolution.* Berkman Center for Internet & Society, 2008. <http://cyber.law.harvard.edu/publications/2008/Role_of_the_Internet_in_Burmas_Saffron_Revolution> accessed 10 Feb 2010.

Cowie, James. "The Geopolitics of Iranian Connectivity." *renesys blog.* 11 Feb 2010. <http://www.renesys.com/blog/2010/02/irans-internet-the-geopolitics.shtml> accessed 28 Feb 2010.

Deibert, Ronald J. et al. *Access Denied: The Practice and Policy of Global Internet Filtering*. The MIT Press, 2008.

Dimitropoulos, Xenofontas et al. "AS relationships: inference and validation." *SIGCOMM Comput. Commun*. Rev. 37.1 (2007): 29-40.

Goldsmith, Jack L., and Tim Wu. *Who controls the Internet?* Oxford University Press US, 2006.

Huffaker, Bradley. "IP Landscape of the Digital Divide. Cooperative Association for Internet Data Analysis." 2003. <http://www.caida.org/research/policy/geopolitical/bgp2country/> accessed 28 Feb 2010.

Klein, Mark. "AT&T Whistle-Blower's Evidence." *Wired* 17 May 2006.

Kraidy, M. "Hypermedia and governance in Saudi Arabia." *First Monday* (2006).

MacKinnon, R. "China's Censorship 2.0: How companies censor bloggers." *First Monday* 14.2-2 (2009).

The CAIDA AS Relationships Dataset. 2009-09-20. Cooperative Association for Internet Data Analysis. <http://www.caida.org/data/active/as-relationships/>.

Zittrain, Jonathan. "Internet Points of Control." *Boston College Law Review* 44.653 (2003).

Appendix 1 – Full Country Data

Country	Total IP addresses	Total Autonomous Systems	Points of Control	IP addresses Per Point of Control	Complexity
Albania	73,984	10	6	12,330	17.9
Algeria	965,120	9	2	482,560	0.95
Angola	25,856	9	5	5,171	44.37
Argentina	7,382,280	165	8	922,785	3.28
Armenia	167,680	29	3	55,893	31.5
Aruba	30,720	2	1	30,720	6.94
Australia	28,883,032	634	9	3,209,225	4.03
Austria	9,625,856	268	11	875,077	3.42
Azerbaijan	230,656	20	1	230,656	13.99
Bahamas	115,200	2	2	57,600	1.74
Bahrain	112,896	17	4	28,224	22.71
Bangladesh	502,016	75	3	167,338	25.81
Barbados	26,112	5	2	13,056	20.84
Belarus	327,424	44	1	327,424	19.9
Belgium	3,732,866	123	13	287,143	3.86
Belize	49,152	1	1	49,152	2.03
Bermuda	81,152	9	4	20,288	11.48
Bolivia	359,424	9	3	119,808	2.99
Bosnia and Herzegovina	401,408	20	5	80,281	6.2
Brazil	29,487,296	464	9	3,276,366	2.33
Brunei Darussalam	125,440	2	1	125,440	1.59
Bulgaria	2,641,920	345	32	82,560	22.08
Burkina Faso	30,208	2	2	15,104	6.62
Cambodia	132,352	29	11	12,032	30.58
Cameroon	56,320	5	2	28,160	10.81
Canada	41,345,793	724	17	2,432,105	3.33
Cayman Islands	29,952	3	2	14,976	10.02
Chile	4,191,744	85	5	838,348	3.56
China	193,304,497	192	4	48,326,124	0.15
Colombia	4,090,761	53	5	818,152	1.93
Costa Rica	1,456,864	7	2	728,432	0.48
Cote D'Ivoire	91,648	5	4	22,912	6.92
Croatia	678,912	64	7	96,987	12.05
Cuba	104,704	3	1	104,704	5.67
Cyprus	581,632	38	8	72,704	7.17
Czech Republic	4,036,267	244	30	134,542	8.15
Denmark	4,260,096	130	18	236,672	3.59
Dominica	172,416	1	1	172,416	0.58
Dominican Republic	418,816	8	2	209,408	1.95
Ecuador	812,544	31	8	101,568	6.54
Egypt	1,770,240	42	6	295,040	3.63
El Salvador	360,960	13	3	120,320	4.1
Estonia	335,876	27	7	47,982	8.31
Faroe Islands	33,792	4	2	16,896	13.63
Fiji	109,312	2	2	54,656	1.83
Finland	9,007,744	110	10	900,774	1.99
France	29,307,279	430	11	2,664,298	1.92
French Polynesia	29,952	1	1	29,952	3.34
Gabon	151,552	1	1	151,552	0.66
Georgia	507,648	26	3	169,216	6.35
Germany	75,943,183	950	24	3,164,299	2.27
Ghana	125,440	18	7	17,920	20.3
Gibraltar	49,920	7	2	24,960	15.75
Greece	3,094,016	105	13	238,001	5.41
Guam	83,200	4	2	41,600	5.19
Guatemala	460,160	19	5	92,032	6.91
Haiti	40,960	4	4	10,240	11.72
Honduras	114,944	12	6	19,157	11.21
Hong Kong	8,366,344	214	13	643,564	5.18
Hungary	2,477,376	144	20	123,868	7.01
Iceland	667,136	27	3	222,378	5.37
India	12,276,011	270	4	3,069,002	3.8
Indonesia	3,820,032	270	12	318,336	13.19
Iran, Islamic Republic of	1,897,472	85	2	948,736	8.12

Country	Total IP addresses	Total Autonomous Systems	Points of Control	IP addresses Per Point of Control	Complexity
Iraq	31,232	4	2	15,616	12.81
Ireland	4,164,480	83	8	520,560	2.07
Israel	4,857,216	172	3	1,619,072	4.14
Italy	14,022,848	459	26	539,340	5.35
Jamaica	177,408	7	2	88,704	4.12
Japan	136,999,360	508	9	15,222,151	0.67
Jordan	311,808	20	3	103,936	11.24
Kazakhstan	798,976	51	2	399,488	8.39
Kenya	157,866	22	4	39,466	18.81
Korea, Republic of	71,750,976	662	3	23,916,992	1.44
Kuwait	666,496	30	7	95,213	7.68
Kyrgyzstan	128,512	13	3	42,837	15.13
Lao People's Democratic Republic	36,864	4	3	12,288	11.45
Latvia	1,314,368	156	5	262,873	24.38
Lebanon	275,968	31	7	39,424	16.48
Libyan Arab Jamahiriya	295,680	1	1	295,680	0.34
Liechtenstein	53,760	6	4	13,440	13.31
Lithuania	2,061,056	88	8	257,632	6.03
Luxembourg	517,375	21	7	73,910	4.76
Macao	212,224	2	1	212,224	1.23
Macedonia, the Former Yugoslav Republic of	496,384	21	2	248,192	5.52
Malaysia	4,135,936	73	6	689,322	2.1
Maldives	37,376	2	2	18,688	5.35
Malta	396,032	18	2	198,016	5.9
Mauritius	193,792	3	1	193,792	1.55
Mexico	17,895,944	166	5	3,579,188	1.25
Moldova, Republic of	569,344	21	5	113,868	4.7
Monaco	49,152	1	1	49,152	2.03
Mongolia	204,032	24	2	102,016	19.92
Morocco	793,344	5	2	396,672	0.63
Mozambique	43,520	7	2	21,760	18.92
Namibia	132,864	7	3	44,288	11.05
Nepal	145,408	24	4	36,352	27.16
Netherlands	14,104,594	337	40	352,614	3.06
Netherlands Antilles	218,624	11	4	54,656	6.6
New Caledonia	47,872	3	1	47,872	7.47
New Zealand	4,570,496	179	13	351,576	7.67
Nicaragua	185,472	13	5	37,094	8.11
Nigeria	235,776	39	11	21,434	26.67
Norway	10,334,208	97	13	794,939	1.04
Oman	155,648	1	1	155,648	0.64
Pakistan	1,423,360	38	2	711,680	3.76
Palestinian Territory, Occupied	261,120	14	1	261,120	6.93
Panama	1,125,632	56	3	375,210	5.31
Papua New Guinea	29,952	1	1	29,952	3.34
Paraguay	167,424	10	1	167,424	11.08
Peru	1,456,896	13	3	485,632	0.89
Philippines	3,665,152	129	5	733,030	5.26
Poland	14,086,464	876	19	741,392	10.19
Portugal	3,080,704	52	10	308,070	2.74
Puerto Rico	861,696	39	8	107,712	5.68
Qatar	297,728	4	2	148,864	1.35
Romania	7,311,872	880	15	487,458	18.52
Russian Federation	23,847,628	2374	36	662,434	26.28
Rwanda	138,240	4	2	69,120	3.64
Saudi Arabia	2,337,536	64	8	292,192	5.01
Serbia and Montenegro	1,214,464	49	3	404,821	5.72
Singapore	3,468,800	118	12	289,066	5.94
Slovakia	1,222,912	67	13	94,070	7.06
Slovenia	1,056,768	147	5	211,353	17.19
South Africa	13,020,032	76	5	2,604,006	0.87
Spain	20,322,710	253	11	1,847,519	1.63
Sri Lanka	450,048	11	5	90,009	2.85
Sudan	83,968	4	2	41,984	5.34
Suriname	25,600	1	1	25,600	3.91
Sweden	6,590,154	297	44	149,776	5.97

Country	Total IP addresses	Total Autonomous Systems	Points of Control	IP addresses Per Point of Control	Complexity
Switzerland	9,995,020	337	28	356,965	4.41
Syrian Arab Republic	204,800	3	2	102,400	1.85
Taiwan, Province of China	25,185,791	105	7	3,597,970	0.77
Tajikistan	34,816	6	4	8,704	18.25
Tanzania, United Republic of	90,880	17	11	8,261	20.13
Thailand	4,808,960	169	14	343,497	7.63
Trinidad and Tobago	293,120	6	2	146,560	2.17
Turkey	10,288,000	232	2	5,144,000	2.9
Uganda	52,736	8	3	17,578	16.75
Ukraine	4,907,135	1165	47	104,407	41.57
United Arab Emirates	2,053,376	8	3	684,458	0.71
United Kingdom	28,275,341	1243	78	362,504	6.76
Uruguay	614,816	16	3	204,938	2.95
Uzbekistan	138,496	20	9	15,388	19.82
Venezuela	3,372,032	30	4	843,008	0.96
Vietnam	5,589,760	53	4	1,397,440	1
Virgin Islands, U.S.	83,456	5	3	27,818	8.85
Yemen	32,512	1	1	32,512	3.08