Mapping Local Internet Control

Hal Roberts, David Larochelle, Rob Faris, John Palfrey
Berkman Center for Internet & Society at
Harvard University
Cambridge, MA

hroberts@cyber.law.harvard.edu
dlarochelle@cyber.law.harvard.edu
rfaris@cyber.law.harvard.edu
jpalfrey@cyber.law.harvard.edu

Abstract— **The Internet is a battleground of control by national governments, among other actors. That contested control takes the form not only of Internet filtering but also of activities that directly impact cyber security, including surveillance and malware hosting. To better understand that battleground, it is important to understand how each nation structures the Internet within 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 points of technical and political control of the Internet.**

This paper describes a method for mapping national networks of autonomous systems, for identifying a small set of autonomous systems that act as points of control for 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. But there are significant differences between autonomous system networks among both countries and regions. 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.

These findings speak both to how the countries exert control over their networks and to how national philosophies of political control have shaped the technical details of their local portions of the Internet. We propose this map 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.[1]

Keywords-internet; autonomous systems; politics

I. 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 user base of researchers. 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 [1]. 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 to allow 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 that it was the first use of the protocol in an operational network and that ARPANET/MILNET grew–by

[1] All visualizations, data, and code used in this paper are publicly available at http://cyber.law.harvard.edu/netmaps.

incorporating other networks–into the larger Internet that we use today. In this origin story, the decision to create the infant Internet was mostly political, 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.[2] 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 the networks for which it would 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 data to other autonomous systems along BGP advertised paths. BGP is the protocol 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:

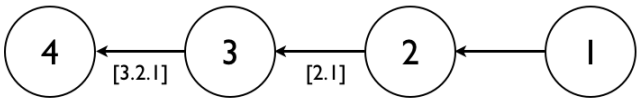

Figure 1. Autonomous systems path announcements

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 along which path to send the data. So if AS #4 needs to forward data ultimately to AS #1, it may know about both the [3.2.1] path 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 #1, 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. It was built into the Internet at its birth. 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 have the power to define the policies of local autonomous systems, 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 [2]. 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 [3]. And others argue that the ultimate role of the Internet in fostering or weakening government control is complex and still not understood [4].

[2] 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.

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 [5]. 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 [6]; 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 [7]; the dismantling of the cybercriminal Russian Business Network [8]; and a Pennsylvania law requiring consumer ISPs to block access to illegal pornography [9]. 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 both domestic U.S. traffic and international traffic, the answer to the question "How can I monitor 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 map of BGP paths that established how the RBN malware ISPs were using complex paths through legitimate-seeming ISPs to launder their 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 [9]. 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 concentrate the control of the billions of end points of the Internet.

But the Internet does not operate as a random game of hot potato between equal autonomous systems. In fact, a very small portion of those autonomous systems carry the traffic for a disproportionate number of routes on the Internet [10]. 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. In 2007, the top 150 autonomous systems carried about 30% of all Internet traffic. By 2009, the top 150 autonomous systems carried about 50% of all Internet traffic [11]. Akamai, a content distribution network, claims to carry fully 20% of all web traffic on its own. The two largest ISPs in the world, Level 3 and Global Crossing, announced a merger in 2011, and the combined entity will carry traffic for over half of the world's IP addresses [12].

This concentration of traffic on only a few autonomous systems per country further amplifies the technical / political role of those autonomous systems. 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.

Existing work 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 [10]. 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 [13]. 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 on autonomous system relationships led by Xenofontas Dimitropoulos at the Cooperative Association for Internet Data Analysis (CAIDA) [14]. Dimitropoulos et al. infer consumer, peer, and sibling relationships between autonomous systems based on BGP announcements. 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. The authors analyze these BGP announcements to infer relationships between autonomous systems. For instance, if there are consistent 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 [15]. 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 [16].

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, we propose that 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 networks. We use three methods to answer these questions: we visually map the consumer/provider relationships between autonomous systems within the country, we calculate the 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.

II. Network Maps

Our first approach to understanding the structure of the autonomous system network within each country is to visually map each national network using data from the CAIDA autonomous system relationship data set [17]. These maps visualize the complex (or sometimes not complex) structure of all of the various autonomous system paths discussed above within a given country.

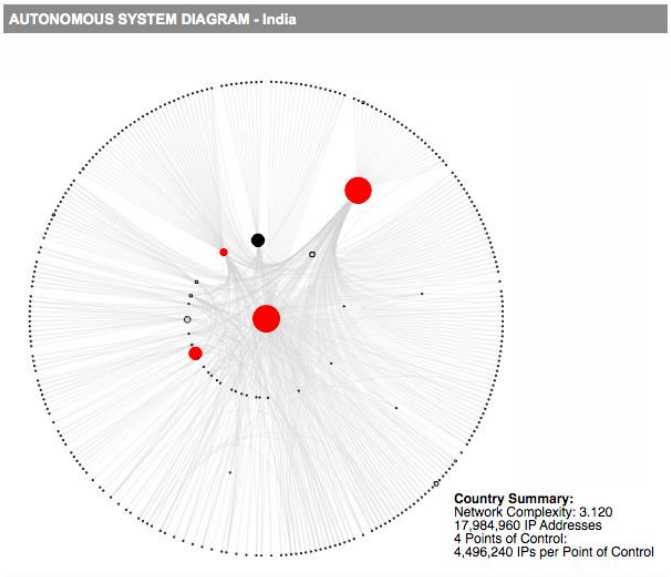

Figure 2. Network map of India

In this map, each circle represents an Indian autonomous system, each line represents a consumer / provider relationship between two autonomous systems, the large 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 first assign each autonomous system to a country using data from the regional Internet registries. [3] 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 system and its consumers. And we determine the *points of control* for the network–the minimum set of autonomous systems that connects 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

[3] 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 at http://www.team-cymru.org/Services/ip-to-asn.html.

graph, with consumer to provider relationships acting as child to parent links. Finally, we map the resulting graph using a circular layout.[4] Autonomous systems with more consumers are closer to the middle of the graph, and the size of each node is determined by the number of connected IP addresses for the autonomous systems.

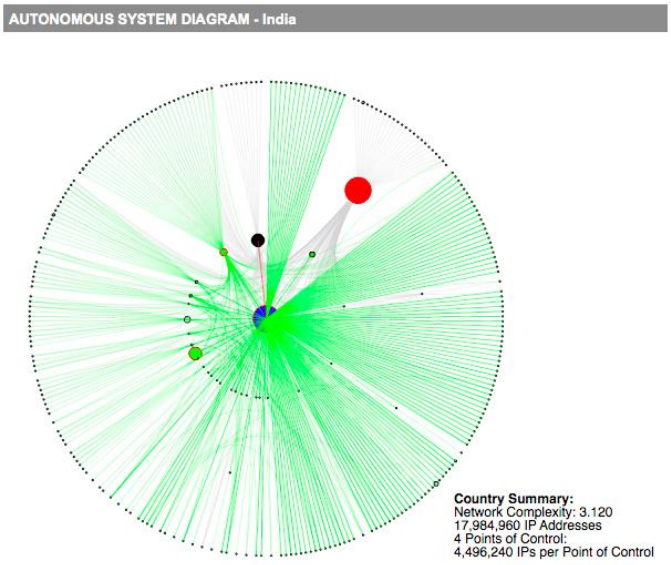

Figure 3. Network map of India with providers and consumers highlighted

In the interactive form, we also allow the user to click on a given autonomous system and find out either its paths to the Internet or its providers and consumers. In this same map of India, the provider (red) and consumer (green) links of the center autonomous system are highlighted. This particular highlight shows that India's most connected autonomous system, Bharti Airtel, has a very high number of local consumers but no local providers (since its only provider link is to the "Rest of World" node). The map of India is representative of most countries, visualizing that a mere 4 autonomous systems act as points of control for nearly all of the 18 million IP addresses in India.

Maps for China, Russia, South Korea, and Ukraine show the regional differences in network structure between Eastern Asia and Eastern Europe.

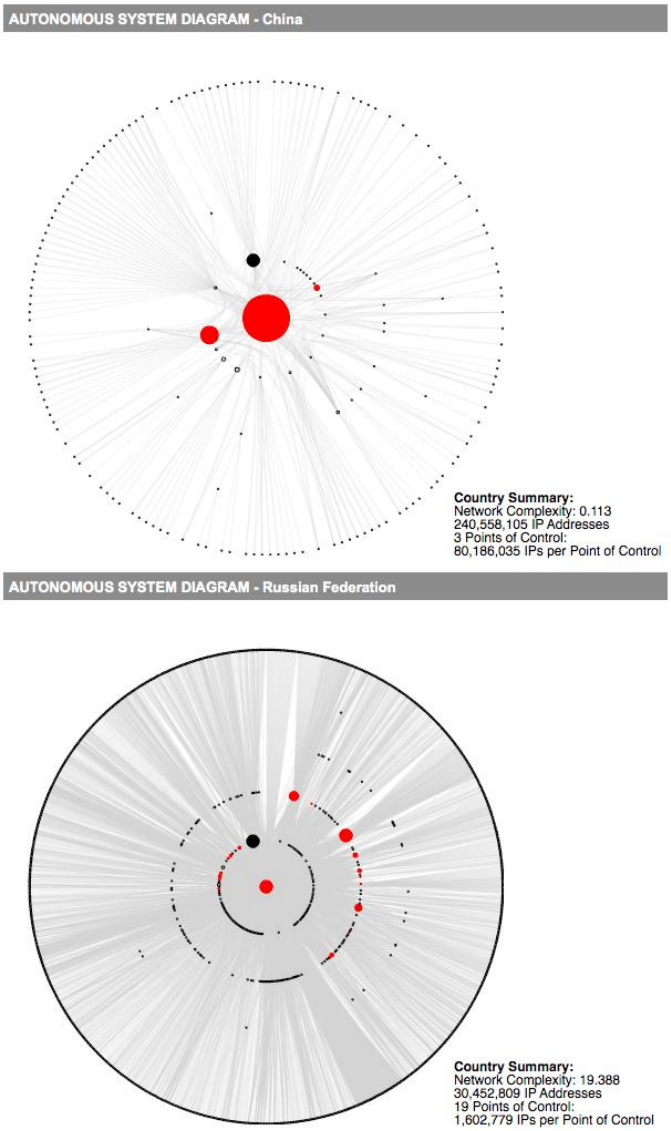

Figure 4. Network maps of China and Russia

The above maps may reflect the way in which China and Russia respectively structure control of the Internet within their borders. China, with 241 million IP addresses, has a dramatically simpler network of autonomous systems than Russia, with only 30 million IP addresses. This difference bears out in the number of points of control as well, with only 3 points of control for China compared to 19 for Russia.

[4] To draw the maps, we use the CircleLayout method of the Flare Toolkit described at http://flare.prefuse.org/api/flare/vis/operator/layout/CircleLayout.html.

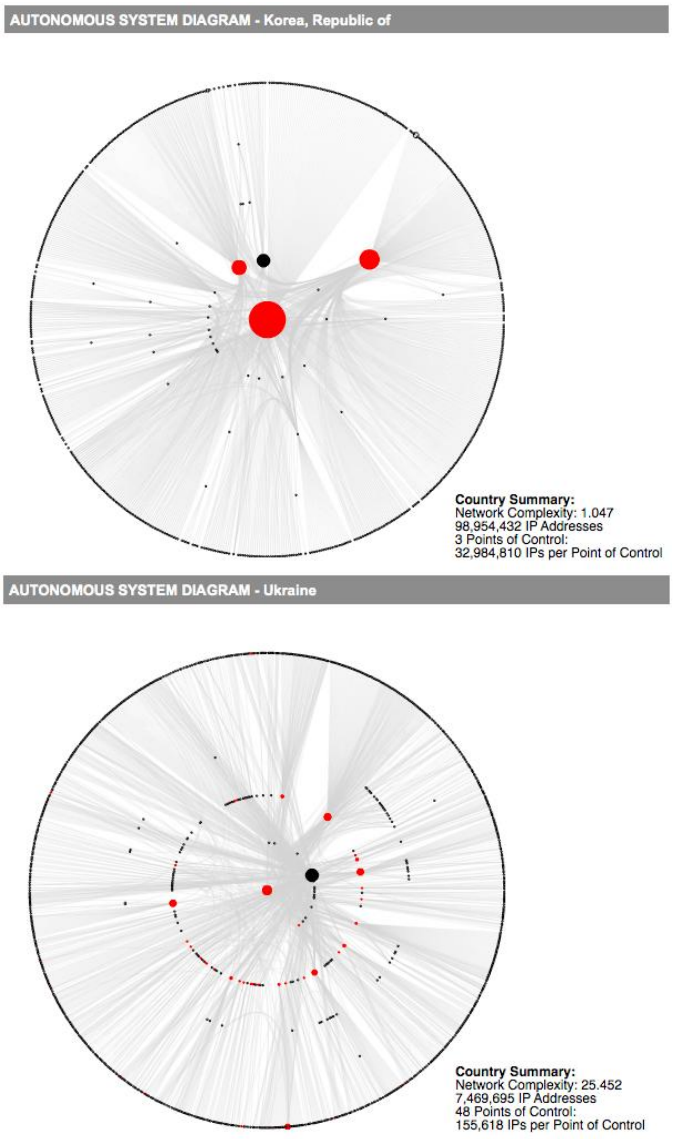

Figure 5. Network maps of South Korea and Ukraine

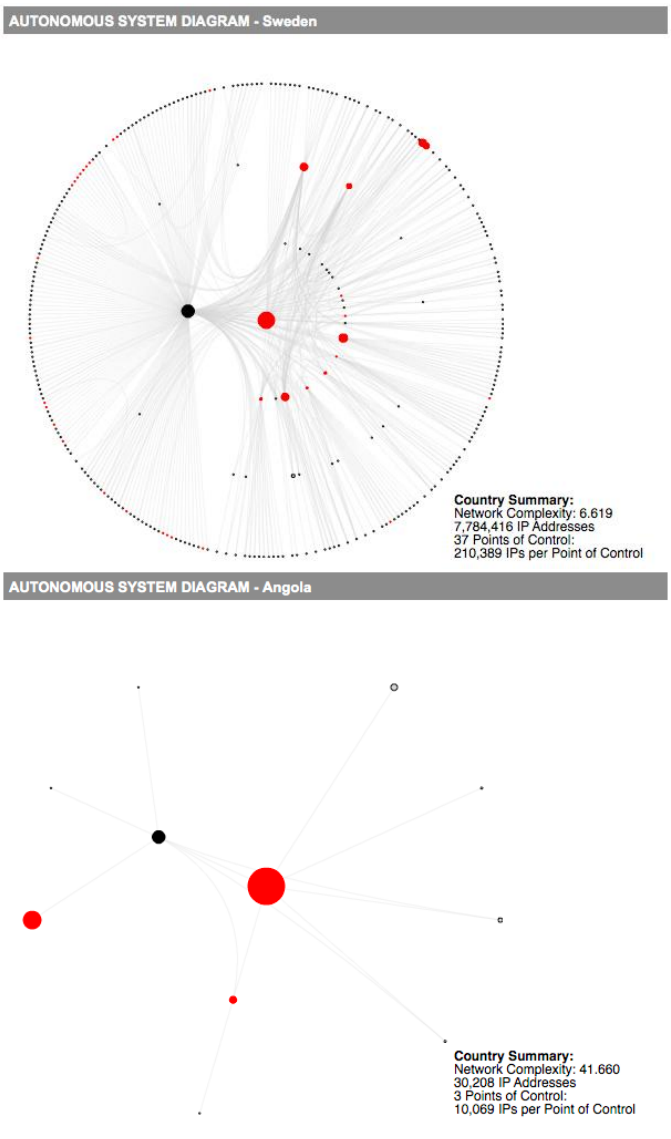

Figure 6. Network maps of Sweden and Angola

These same differences are present but less striking in a comparison of South Korea, with 99 million IP addresses, to Ukraine, with only 7.5 million IP addresses. South Korea's structure is visibly more dense than China's but still comparable to Ukraine, despite the fact that South Korea has about 13 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 48, reflecting the fact that South Korea's IP addresses are mostly concentrated in that handful of core autonomous systems.

Finally, above are two maps from non-Eastern Asia / Eastern European countries, which show the extremes along the spectrum of size and complexity. Sweden is the network with the most points of control of any country other than Ukraine (excluding the U.S. for reasons we discuss below), but it only has 37 points of control to Angola's 3 (for a ratio of 12.33:1) even though it has 7.8 million IP addresses to Angola's 30 thousand (for a ratio of 259:1).

Network maps of all countries with more than 25,000 total IP addresses 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.

III. POINTS OF CONTROL

The purpose of the points of control metric is to determine both the proportion of control potentially executed over local Internet routes by any given autonomous system 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 country filter its connection? Where might it surveil its citizens? Where might a malware host hide its connection to the larger the network? Which are the points that must be disabled to disconnect the country from the Internet entirely?

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 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.[5] To 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.[6] 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 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.

TABLE I. TOP TEN COUNTRIES BY POINTS OF CONTROL

Country	PoC	IPs	Region
Ukraine	48	7,469,695	Eastern Europe
Sweden	37	7,784,416	Northern Europe
Bulgaria	32	3,049,856	Eastern Europe
Czech Republic	31	5,256,576	Eastern Europe
Netherlands	24	15,709,195	Western Europe
Switzerland	22	9,395,844	Western Europe
Germany	19	80,719,913	Western Europe
Russia	19	30,452,809	Eastern Europe
Poland	19	16,588,576	Eastern Europe
Hungary	17	2,647,104	Eastern Europe

As mentioned above, Ukraine has the most points of control of any country in the set we consider, with only 48. And only five other countries have more than 20. 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.

[5] 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 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.

[6] 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.

However, there are significant differences in the number of points of control between countries. The most interesting differences are between the numbers of points of control for the Eastern Asia and Eastern Europe regions:

TABLE II. AVERAGE POINTS OF CONTROL BY REGION

Region	Average PoC	Total IPs
South Central Asia	2.85	26,635,456
Central America	3.00	21,342,440
Western Asia	3.21	28,887,731
South America	4.40	64,518,477
Eastern Asia	4.80	510,641,820
Southern Africa	6.00	13,807,378
South-Eastern Asia	6.50	38,019,138
Southern Europe	6.55	69,446,080
Australia & New Zealand	7.50	43,407,381
Northern Europe	11.00	99,405,852
Western Europe	13.29	155,311,961
Northern America	14.00	44,085,103
Eastern Europe	19.10	74,574,504

The above table compares the average points of control for all geographic regions with at least 10 million 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 ($r = 0.58$). The relatively high number of points of control in Eastern Europe confirms the differences shown in the network maps above.

The full results for the points of control metric for all countries, along with the complexity metric defined below, are available in the appendix to this paper and at http://cyber.law.harvard.edu/netmaps.

IV. 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 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 are impacted by the complexity of a network 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.[7] 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 different countries and regions. The following table lists the average network complexity by region:

[7] 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.

TABLE III. COMPLEXITY BY REGION

Region	Average Complexity	Total IPs
Southern Africa	0.83	13,807,378
Eastern Asia	1.54	510,641,820
Central America	2.63	21,342,440
South America	2.64	64,518,477
Western Europe	3.31	155,311,961
Northern America	3.35	44,085,103
South-Eastern Asia	3.83	38,019,138
Australia & New Zealand	4.71	43,407,381
Western Asia	5.14	28,887,731
Southern Europe	5.24	69,446,080
Northern Europe	5.34	99,405,852
South Central Asia	6.78	26,635,456
Eastern Europe	11.35	74,574,504

As with points of control, Eastern Asia and Eastern Europe are at opposite ends of the spectrum. In this case, Eastern Europe is the big outlier, with nearly twice 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 [8]. 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 facilitates these complex, control resistant structures.

V. 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 announcements 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.[8] 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.[9] So even though our maps may not hold for a given

[8] 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 not authoritative.

[9] We regenerated the network complexity and IP addresses per points of control numbers for each country under three

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 system. 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 around 2 billion IP addresses allocated for only about 230 million Internet users.[10] 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.

VI. CONCLUSIONS

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.

These core findings are robust enough to stand up even in the face of the limitations of our methods and the available data described above. Further evidence for our findings can be found in the recent use of politically motivated national network shutdowns in China, Iran, Egypt, and Libya. Starting in July 2009, China cut off almost all Internet access for the Xinjiang region for ten months in response to local protests [18]. During the June 2009 election protests, the Iranian Internet experienced widespread outages and severe throttling that many attributed to government manipulation at a critical political moment [19]. For six days during the Egyptian protests at the beginning of 2011, the Egyptian government shut down the country's connection to the rest of the Internet almost entirely, with only one internationally connected autonomous system remaining up [20]. And shortly thereafter, the Libya government responded to its own protests by cutting the connection to the rest of the Internet for much of February and March [21].

Our methods do not support mapping a single region within a country, so for the China example we can only point to the overwhelming simplicity of the Chinese network nationally for the ease of shutting down a region. The following maps show the network structure of Iran, Egypt, and Libya, respectively:

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. Th plots of complexity and of IP addresses per points of control for each of the models against the others yielded an $r > 0.8$ 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.

[10] 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.

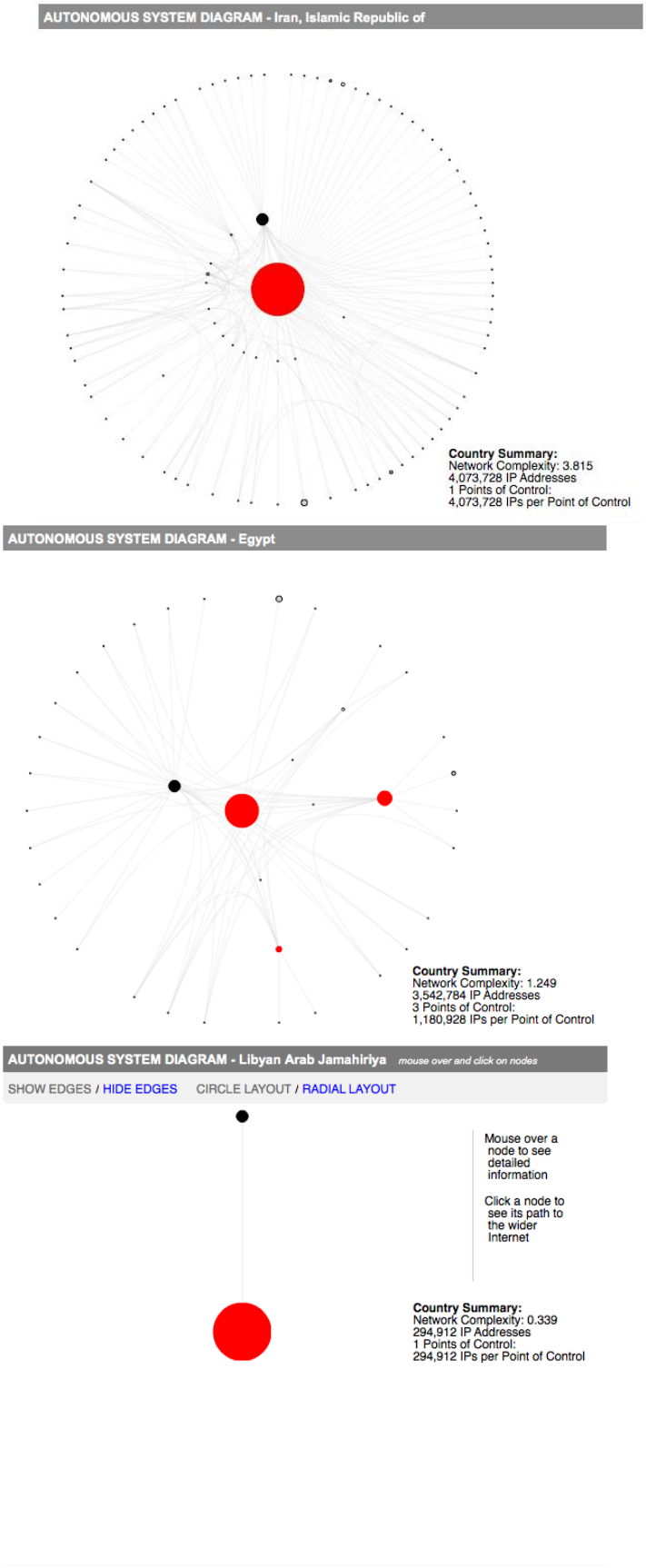

Figure 7. Network maps of Iran, Egypt, and Libya

Both Iran and Libya have only a single point of control (and indeed our data show only a single autonomous system for Libya), and Egypt has only three points of control. The very small number of points of control for each of these countries shows not only the ease of enforcing state control over the Internet through those points but also suggests that the political structure of the countries has influenced the technical structure of the networks. In other words, the lesson from these network structures is not just that shutting down each network takes only a handful of phone calls, but also that countries that structure their networks so simply also have political structures that allow for control of the network.

Interestingly, Iran was reportedly able to shut its network down entirely, at least briefly, as was Libya, but in Egypt a single autonomous system continued to carry international traffic for four days after the initial shutdown. It is not clear why that particular autonomous system remained up, but a likely explanation is that that autonomous system (Noor Group, with 4.9% of the country's connected IP addresses) carries traffic for the Egyptian stock exchange and other important financial sites [20]. Noor Group's resiliency in the face of otherwise total network shutdown is more evidence of the intertwined nature of the political and technical structures. Whether the Noor Group stayed up because it had the political clout to refuse the phone call from the national government or whether the government feared making the phone call because of the potential political fallout of shutting down the country's stock market and other core financial institutions, we can draw the same conclusion that the technical-political structure of the network in Egypt is complicated by its dependency on a liberal economy.

This point is important for understanding the significance of the difference in national network structures. It is significant that Russia has a network that is several orders of magnitude more complicated than China's not because the Russian government would merely have to make more phone calls to shutdown (or filter or surveil) its network; obviously, the logistical difference between a few phone calls and few dozen phone calls is not significant for a national government. The difference is significant because of the differences in political philosophy that the network structure implies; this difference in philosophy is born out through the drastically difference methods of control used by China and Russia. Where China uses brute methods like filtering and network shutdowns as its first line of attack, Russia uses more subtle methods like third party denial of service attacks and youth brigades of pro-government commenters. Because of the much more complex structure of its technical-political network, Russia would be more likely to run into political difficulties like those Egypt ran into with the Noor Group if it tried to make the dozens of calls necessary to shutdown its network. And China has the ability to shutdown a portion of its Internet not just because of the small number of phone calls necessary, but because it has structured is technical-political network to minimize the points of control for such government action.

We think that both the core findings of this paper and the identification of points of control for specific countries will provide fruitful starting points for many kinds of research into the technical-political nature of networks. The network

structures found in Eastern Asia and Eastern Europe beg the question of why the countries in those 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)?

REFERENCES

[1] J. Abbate. *Inventing the Internet*. MIT Press, 1999.

[2] J. Goldsmith and T. Wu. *Who controls the Internet?*. Oxford University Press US, 2006.

[3] M. Kraidy. "Hypermedia and governance in Saudi Arabia." First Monday (2006).

[4] M. Chowdhury. "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.

[5] R. Deibert, J. Palfrey, R. Rohinsky, and J. Zittrain. *Access Denied: The Practice and Policy of Global Internet Filtering*. The MIT Press, 2008.

[6] J. Cowie. "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.

[7] M. Klein. "AT&T Whistle-Blower's Evidence." Wired 17 May 2006.

[8] D. Bizeul. "Russian Business Network study." November 20, 2007. http://www.bizeul.org/files/RBN_study.pdf. accessed 16 January 2008.

[9] J. Zittrain. "Internet Points of Control." Boston College Law Review 44.653 (2003).

[10] B. Huffaker and k. claffy. "IPv4 & IPv6 Internet Topology Map January 2009." Cooperative Association for Internet Data Analysis. 2009. http://www.caida.org/research/topology/as_core_network/pics/ascore-ipv4-ipv6.200903_poster_1250x850.png. accessed 27 February 2010.

[11] C. Labovitz et al. "ATLAS Internet Observatory Annual 2009 Annual Report." North American Network Operator's Group Meeting 47. October 19, 2009, http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf. accessed June 10, 2010.

[12] E. Zmijewski. "Level Crossing." Renesys Blog, April 14, 2011. http://www.renesys.com/blog/2011/04/level-crossing.shtml. accessed May 11, 2011.

[13] J. Karlin, S. Forrest, and J. Rexford. "Nation-State Routing: Censorship, Wiretapping, and BGP." arXiv.org, March 2009.

[14] X. Dimitropoulos et al. "AS relationships: inference and validation." SIGCOMM Comput. Commun. Rev. 37.1 (2007): 29-40.

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

[16] B. Huffaker. "IP Landscape of the Digital Divide. Cooperative Association for Internet Data Analysis." 2003. http://www.caida.org/research/policy/geopolitical/bgp2country/. accessed February 18, 2010.

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

[18] "China restores Xinjiang internet," BBC, May 14, 2010. http://news.bbc.co.uk/2/hi/8682145.stm. accessed May 1, 2011.

[19] J. Cowie. "Strange Changes in Iranian Transit." renesys blog. http://www.renesys.com/blog/2009/06/strange-changes-in-iranian-int.shtml. accessed May 14, 2011.

[20] J. Cowie. "Egypt Leaves the Internet." renesys blog. Janruary 27, 2011. http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml. accessed May 14, 2011.

[21] J. Cowie. "Libyan Disconnect." renesys blog. February 18, 2011. http://www.renesys.com/blog/2011/02/libyan-disconnect-1.shtml. accessed May 14, 2011.

APPENDIX: FULL COUNTRY DATA

The following table includes all metrics generated for all countries with more than 250,000 IP addresses.

Country	IP Addresses	Autonomous Systems	Points of Control	IPs per Point of Control	Complexity
Algeria	1,448,960	11	2	724,480	.77
Argentina	9,073,164	158	7	1,296,166	2.50
Armenia	488,704	28	6	81,450	7.35
Australia	38,026,901	642	7	5,432,414	3.20
Austria	12,390,784	271	4	3,097,696	2.97
Azerbaijan	339,968	17	1	339,968	8.30
Bangladesh	681,600	72	2	340,800	18.80
Belarus	1,162,240	44	1	1,162,240	5.65
Belgium	4,196,352	118	10	419,635	3.04
Bolivia	444,416	9	4	111,104	2.30
Bosnia	572,160	20	6	95,360	4.16
Brazil	35,690,176	483	8	4,461,272	2.95
Bulgaria	3,049,856	342	32	95,308	20.48
Canada	44,085,103	705	14	3,148,935	3.35
Chile	5,262,208	89	5	1,052,441	2.84
China	240,558,105	177	3	80,186,035	.11
Colombia	5,266,817	51	5	1,053,363	1.59
Costa Rica	1,458,688	5	2	729,344	.35
Croatia	815,744	59	6	135,957	9.81
Cyprus	824,192	37	5	164,838	4.73
Czech Republic	5,256,576	319	31	169,566	10.74
Denmark	4,397,056	139	16	274,816	3.73
Dominican Republic	515,840	8	2	257,920	1.58
Ecuador	1,274,112	33	4	318,528	5.08
Egypt	3,542,784	36	3	1,180,928	1.25
El Salvador	463,616	11	3	154,538	2.44
Estonia	754,688	26	3	251,562	3.96
Finland	9,558,656	113	6	1,593,109	1.92
France	31,974,177	434	7	4,567,739	2.09
Georgia	681,344	26	2	340,672	7.25
Germany	80,719,913	932	19	4,248,416	1.68
Greece	3,772,160	88	11	342,923	4.07
Guatemala	570,144	18	4	142,536	4.88
Hong Kong	9,734,556	217	9	1,081,617	4.92
Hungary	2,647,104	143	17	155,712	8.28
Iceland	666,880	27	3	222,293	5.09
India	17,984,960	291	4	4,496,240	3.12
Indonesia	7,946,960	273	7	1,135,280	6.11
Iran	4,073,728	96	1	4,073,728	3.82
Ireland	4,343,723	80	8	542,965	1.94
Israel	5,852,688	165	4	1,463,172	3.24
Italy	36,268,672	454	7	5,181,238	2.08
Japan	161,064,743	495	8	20,133,092	.55
Jordan	422,784	21	3	140,928	8.71
Kazakhstan	1,868,032	51	2	934,016	4.62
Kenya	1,001,216	23	4	250,304	3.31
Korea, Republic of	98,954,432	637	3	32,984,810	1.05
Kuwait	949,120	30	6	158,186	4.70
Latvia	1,369,728	145	6	228,288	19.20
Lebanon	388,864	32	7	55,552	11.99
Libya	294,912	1	1	294,912	.34
Lithuania	2,012,160	80	7	287,451	5.65
Luxembourg	925,696	25	7	132,242	3.02
Macao	329,984	3	1	329,984	1.09
Macedonia	572,928	21	4	143,232	4.21
Malaysia	5,289,728	69	5	1,057,945	1.65
Malta	490,240	14	2	245,120	4.77
Mauritius	461,056	4	1	461,056	.87
Mexico	17,515,592	158	4	4,378,898	1.12
Moldova, Republic of	857,856	23	4	214,464	3.32
Morocco	1,085,952	4	2	542,976	.37
Nepal	461,568	25	5	92,313	8.11
Netherlands	15,709,195	339	24	654,549	3.71

Country	IP Addresses	Autonomous Systems	Points of Control	IPs per Point of Control	Complexity
Netherlands Antilles	290,048	12	4	72,512	4.64
New Zealand	5,380,480	168	8	672,560	6.23
Nigeria	420,224	41	10	42,022	15.44
Norway	11,873,792	101	11	1,079,435	.96
Pakistan	1,033,856	37	2	516,928	6.68
Palestine	315,904	15	1	315,904	6.21
Panama	1,334,400	53	2	667,200	4.39
Paraguay	258,560	10	2	129,280	5.94
Peru	2,206,464	13	3	735,488	.61
Philippines	4,434,624	126	5	886,924	3.68
Poland	16,588,576	876	19	873,082	7.86
Portugal	3,232,000	52	8	404,000	1.91
Puerto Rico	1,005,696	37	7	143,670	4.47
Qatar	509,944	5	2	254,972	1.55
Romania	5,738,752	264	9	637,639	5.79
Russian Federation	30,452,809	2,346	19	1,602,779	19.39
Saudi Arabia	3,294,379	66	3	1,098,126	3.74
Serbia	1,972,053	75	5	394,410	5.54
Singapore	4,328,480	124	10	432,848	5.17
Slovakia	1,351,040	61	11	122,821	6.51
Slovenia	1,490,432	151	5	298,086	14.56
South Africa	13,807,378	75	6	2,301,229	.83
Spain	22,231,744	249	10	2,223,174	1.63
Sri Lanka	531,712	11	4	132,928	2.28
Sweden	7,784,416	301	37	210,389	6.62
Switzerland	9,395,844	356	22	427,083	6.68
Syrian Arab Republic	665,600	3	1	665,600	.85
Taiwan	31,769,916	106	5	6,353,983	.75
Thailand	6,583,442	172	8	822,930	5.61
Trinidad and Tobago	457,984	6	2	228,992	1.41
Turkey	11,632,896	226	2	5,816,448	2.72
Ukraine	7,469,695	1,122	48	155,618	25.45
United Arab Emirates	2,521,344	8	2	1,260,672	.57
United Kingdom	56,644,753	1,177	13	4,357,288	4.34
Uruguay	907,392	15	2	453,696	1.80
Venezuela	4,135,168	30	4	1,033,792	.79
Viet Nam	9,435,904	56	4	2,358,976	.75