Towards Measuring the Geographic and Political Resilience of the Internet

Annika Baumann and Benjamin Fabian
Institute of Information Systems
Humboldt-Universität zu Berlin
Spandauer Str. 1, 10178 Berlin, Germany
annika.baumann@wiwi.hu-berlin.de
bfabian@wiwi.hu-berlin.de

Abstract

Given the importance of the Internet for worldwide communication and services, its resilience against attacks, accidents, or attempts of misusing political control becomes critical for businesses and society. This article focuses on the question how vulnerable specific geographical regions are to an Internet access disruption or to censorship-based impediments due to governmental control. In particular, a new metric is developed that measures the geographical Internet resilience on a country level. For this purpose several indices based on geography, technology as well as control are combined into a single, rank-based score indicating the Internet resilience of a particular country compared to others.

Keywords: Internet, Resilience, Autonomous System, Geography.

1 Introduction

The importance of the Internet as today's communication and information medium is undisputed. To a large extent the current cost efficient and fast worldwide communication has been made possible by it. Moreover, it is the foundation for all online business models and services. The Internet is often defined as a global system of computer networks that uses the standard Internet protocol suite (TCP/IP) to serve users worldwide (Paul et al., 2013). In turn, a computer network, which is a collection of connected machines and IP routing prefixes under control of one or more operators with a common clearly defined routing policy, is referred to as an autonomous system (AS) (Hawkinson and Bates 1996). Today, the Internet consists of about 70,000 ASs (Potaroo, 2013) with a complex structure that complicates any kind of related analysis. Nonetheless, considering the role of the Internet in modern worldwide communication, social interaction, and economics it attracts significant scientific interest.

Internet resilience is a topic of highly practical political relevance which also has become a focus of several research studies from different disciplines. A limited Internet availability could lead to significant financial losses for economies and businesses. Furthermore, the ability to share information and communicate with people around the world increasingly depends on the Internet (ITU, 2011). Over the last decade, several Internet failures have occurred because of different reasons. Some were results of natural disasters, such as earthquakes (Cowie, 2011a), hurricanes (Brown, 2008a), and undersea mud volcano eruptions (Madory, 2013). Some were caused by a human factor such as cable cuts (Popescu, 2008b) or business disputes (Underwood, 2008c); and some originated from on-going battles for control over information streams on the web due to armed conflicts (Zmijewski, 2008d), politics (Cowie, 2011b), censorship (Brown, 2008e), central planning (Madory, 2012a), and terrorism (Madory, 2012b). All these examples highlight how numerous the Internet connectivity threats are, and that geography plays an important role in Internet resilience.

This paper contributes to answering the question how vulnerable specific geographical regions are to Internet access disruptions as well as censorship-based impediments due to local governmental control. For this purpose, we aim to develop a metric that measures the geographical Internet resilience based on a country level. This metric encompasses several geographic-, technology- as well as control-based indices, combining them into a single, rank-based score that estimates the Internet resilience of a particular country compared to others. Such a comparison could serve as an indicator for regions at risk, which could support international politics, Internet businesses or freedom activists in improving Internet resilience and censorship resistance by focusing their activities. The goals and methods proposed in this paper are motivated by investigating if and how an earlier approach by Roberts et al. (2011) could be improved by including additional metrics to account for further dimensions.

The structure of the paper is as follows: Section 2 discusses the background and current state of research on geographical vulnerability of the Internet. Section 3 describes the methodology and the methods used for data collection. Section 4 describes our new geographical resilience metric in detail. Finally, section 5 presents the results and the conclusions as well as limitations and future outlook.

2 Background and Related Work

Dramatic Internet failures already happened in reality. Sterbenz et al. (2010) as well as Wu et al. (2007) refer to various so-called "large-scale disasters". Even if the entire Internet as a whole seems to be quite robust (Sterbenz et al., 2010) these kinds of disruptions still affect the connection to the Internet in certain areas to a various extent. An example for a natural disaster is the hurricane Katrina, which raged in 2005 in the southeast of the USA – some of the most damaged areas were Louisiana (especially New Orleans), Mississippi, Alabama, Florida, and Georgia. A fraction of between 8 percent (in Louisiana) and 38 percent (in Mississippi) of the locally situated networks were

unavailable due to Katrina. It was possible to restore most of the affected networks within a one-day period but especially networks in Louisiana needed a lot more time to recover fully (Cowie et al., 2005).

Another example for a natural disaster is the Taiwan earthquake in the end of 2006. The most severe impact was the damage of seven undersea cables in the affected area, when only two remaining cables still worked. Again, several networks were either unavailable or did not work properly – a maximum of around 4,000 networks were affected, most of them recovering quite fast after the seven earthquakes. The repair of the undersea cables took longer and lasted until the mid of February 2007 (Wilcox, 2007). The global impact of the Taiwan earthquake was stronger than the one of the hurricane Katrina because of the damage of the undersea cables. Locally, both natural disasters had a severe negative effect on Internet accessibility.

Other disasters, which are not necessarily of a natural origin, are for example fires caused by accidents. If such a fire occurs at a place important for the accessibility of the Internet, it might have a severe negative impact on it. In 2001 such an incident happened at the Howard Street Tunnel in Baltimore, Maryland. A fire caused by a train that ran off the rails destroyed fiber cables inside the tunnel. It was possible to reroute most of the traffic, resulting in slow connections for a bulk of Internet users. The repair of the damaged fiber cables required one and a half day (Sterbenz et al., 2010).

Next to these (natural) disasters a variety of further possible impediments exist which could affect the Internet access capability in a negative way. Mistakes – either on the human or system side – might have a severe impact too. A series of aligned faults caused a blackout in Canada and the Northeast USA in 2003. Not only around 50 million people had no access to electricity, but also the Internet access was significantly affected. A maximum of around 3,175 networks was unavailable due to the outage concerning over 50 percent of the ASs in this area and more than 1,700 organizations. For example, around 460 ASs had no working network left for more than four hours during the blackout – belonging to around 1,000 organizations (Renesys, 2003). Furthermore, it seems to be the case that organizations which are connected by many networks are only slightly affected. Organizations which have a small number of networks and are therefore only marginally connected suffer much more from the blackout, e.g., by having all their network connections lost. Therefore, the availability of the Internet depends not only on geography, its own infrastructure and functionality but also on several more factors like the power grid.

Other mistakes are accidental cable cuts such as the one of the Mediterranean Sea Cable in 2008, which seems to happen regularly (Sterbenz et al., 2010), and different forms of misconfigurations like the one by the Pakistan Telecom in 2008, which tried to block videos on YouTube by configuring their own AS as the path to the content. Instead of being valid only for Pakistan, this configuration spread worldwide. This mistake could only be finally undone by disconnecting the Pakistan Telecom. Another example for an accidental misconfiguration happened by a Czech Provider in 2009. A wrong configured BGP-setting of its router led to abnormally long AS paths affecting the worldwide Internet. An until then unobserved bug in Cisco's BGP implementation as well as a lot of misconfigured routers allowing these long AS paths led to this circumstance with global routing instability (Sterbenz et al. 2010).

Intentional attacks are another challenge including for example distributed denial of service attacks (DDoS), worms and acts of terrorism. The terrorist attacks on the World Trade Center on September 11 in 2001 caused severe damage not only on the human side. Important cables for Internet functionality in and next to the building were demolished. Although it affected around 1,000 networks, the effects on the global Internet were rarely noticeably. Most of the traffic could be rerouted or the necessary facilities could be rebuilt within a day (National Research Council, 2003). Furthermore, these kinds of events might even provoke another challenge to maintain the availability of the Internet which is unusually high traffic load caused by people seeking for information.

Worms are another form of intentional attacks targeting the Internet and its users directly. There are many examples of them such as the Nimda worm and Code Red v2, both raging in 2001. Apart from the destruction caused on private computers, these worms result in enormous BGP update rates leading to routing instabilities. In case of the Nimda worm, the update rates increased from around 400 per minute to around 10,000 per minute (Cowie et al., 2002).

In 2011, the European Network and Information Security Agency (ENISA) provided an Internet incident report. Eleven countries reported a total of 51 incidents, whereas nine countries reported no disturbances and another nine countries just sent in no report at all. The most frequent incidents were caused by hardware/software as well as third party failures – accounting for 80 percent of the errors. Human errors, malicious attacks and natural impacts are less frequently the cause of the disturbance. Nevertheless, natural disasters such as storms and earthquakes as well as malicious attacks have the most severe impact causing a much longer duration of the disruption of service than the other three impact categories. This incident report makes obvious that there are many sources of interference causing more than 50 significant Internet service disruptions. However, the estimated number of unreported or undetected cases might be much higher (ENISA, 2012).

All of these examples hint at the importance of considering geography as an important factor of Internet resilience. Several research papers already contributed to the question of how robust certain countries are, focusing on the AS level of the Internet topology. For example, Roberts et al. (2011) used data collected by CAIDA (CAIDA AS Names, 2012) and Team Cymru (Cymru, 2013) to perform an AS-to-country mapping. They used the results of the mapping to construct a metric reflecting the complexity of the Internet infrastructure in a single country, which incorporates the number of IP addresses, the number of ASs, and the number of points of control, which are defined as ASs which provide connection to 90% of IP addresses in a country. Their findings indicate that absolute size does not always determine the complexity and resilience of a national network infrastructure. While there are 177 ASs in China, Roberts et al. identified that there are only 4 points of control. Thus, a threat of a country-wide Internet outage is much higher compared, for instance, to Hungary where the number of ASs and the number of points of control are 143 and 17, respectively.

Reynolds and Tamaddon (2011) extended the results of Roberts et al. (2011) with Internet filtering scores. They used several rankings provided by the OpenNet Initiative to construct two entities – the Political Filtering Score and the Overall Filtering Score. The combined metric is called the Network-Political Resiliency (NPR) and indicates *the internal power of a country to control its own Internet access* (Reynolds and Tamaddon, 2011). In accordance with their calculations, only one country out of 152 analyzed was classified as High NPR – the United Kingdom, while the majority, 101 countries, received a Low NCR rating.

Wählisch et al. (2012) proposed an enhanced IP-block-based approach for identification and classification of ASs related to a specific country. They argue that foreign ASs hosting national organizations should also be considered as a part of national network infrastructure. Thus, to identify such ASs, Wählisch et al. (2012) suggested shifting from the prefix-based level to the IP-block level. The researchers employed the developed technique to spot all of the Germany-related ASs and claim to outperform the prefix-based methodology by 25%.

All of these papers considered solely technology-based metrics to measure the vulnerability of the Internet on the AS-level with respect to certain countries. We will extend this approach by including additional metrics to assess further dimensions and therefore enhance the analysis to a different level.

3 Methodology and Data Collection

In this section, we provide our methodology and describe the data collection. First, we will discuss the concept of geographic proximity in the Internet. Besides the topological vulnerability of ASs, there is another type of vulnerability which is independent of existing connections. This is the vulnerability due to geographic location. One variant of this vulnerability is characterized by a very close

geographic proximity of ASs, e.g., if the power supply to a particular town fails in which several ASs are located. In this case, several ASs might be affected at the same time even if there is no actual topological connection between them. Thus, investigating the close geographic proximity of ASs is challenging.

Therefore, a methodology of geo-locating ASs will be provided. However, depending on the degree of accuracy and detail it might not even be possible to specify for some ASs whether they are close to each other. Thus at first, the geographic location of IP addresses needs to be analyzed. Services such as the IxMapper of Ixia (IxMapper, 2013) or EdgeScape of Akamai (EdgeScape, 2013) mainly use hostname-based mapping methods to find a solution for this (Lakhina et al., 2003). The degree of correctness of GeoIP services is questionable. MaxMind (Maxmind, 2013a), a supplier for GeoIP Services, provides on its website accuracy results for IP geolocations based on an allocation to cities. Correct results within a distance of 40 kilometers range between 45 and 96 percent, while the average value is 73 %. This means that on average 27% of the allocation results are either wrong or the city is unknown (MaxMind, 2013b). In the second step, these IP addresses need to be assigned to the corresponding ASs. This so-called alias resolution process is not yet fully developed and can cause errors during the assignment procedure. This kind of more detailed geographic analysis is conducted, for example, by Lakhina et al. (2003). But due to the immaturity of this method for the close geographical proximity of ASs, it will not be considered in this paper.

The second type of geographic vulnerability is characterized by the fact that ASs are situated in the same country, i.e., their wide geographic proximity within country borders. In such a case it might happen that a whole country is cut off from the Internet. This could happen, for example, if it is easy to control the ASs in a certain area. This might be caused by the fact that there are only few ASs regulated by a small number of responsible companies. Recently such Internet cut-offs happened in Syria and Egypt (Cowie, 2011c; Cowie, 2012c). This analysis is not as strict as the close geographic proximity analysis. The allocation within country borders covers a larger geographical area than a city does, which makes it possible that mapping errors are reduced. This paper will therefore focus on the geographic analysis of ASs situated in the same country.

In addition to the IP geolocation method mentioned above, another approach for retrieving the geographic location of an AS is viable. The Regional Internet Registries (RIRs) contain in their databases a two-letter country code based on ISO 3166 for each AS that indicates the country in which the managerial unit is located or rather where the AS has been registered (APNIC Aut-Num, 2013; Roberts et al., 2011 p.4). This may be the same as the geographic location of the AS but this does not necessarily have to be the case. As mentioned by Roberts et al. (2011, p.4), the country code seems to be accurate enough for such an analysis. There may be some exceptions such as ASs in Africa that have been registered somewhere else in the world due to the earlier nonexistence of the AfriNIC.

The geographic resilience analysis conducted in this paper is therefore based on the data source provided by Team Cymru (Cymru, 2013). This is a service offering listings that obtain daily updated information about country codes directly from the various RIRs including ARIN, RIPE, AfriNIC, APNIC and LACNIC. The data was downloaded on February 18th, 2013. All information on ASNs was selected from these datasets and finally merged. In total there are 54,773 ASNs mentioned in combination with their underlying country code. For the final analysis, entries containing too general location information such as ‘Europe’ or ‘Asia/Pacific region’ were ignored.

To validate the reliability and dynamics of this data, it was cross-checked with the country code data provided by CAIDA’s AS Rank project (CAIDA AS Names, 2012) and the GeoIP data from MaxMind (MaxMind, 2013a). The AS Rank country data (CAIDA AS Names, 2012) was used to investigate how likely it is that country codes (frequently) change. The data is based on the same method, i.e., it provides country codes collected via various RIRs. The difference lies in their date of creation. The AS Rank data originates from the 29th of June 2012 resulting in difference of 235 days between these two datasets. By comparing the intersection of ASs of both data files, it was possible to find out how many country codes have changed. Both data files share 52,306 joint ASNs. Of these,

2,176 ASNs changed their underlying country code, which is around 4.16 percent. In general, the country data remains quite stable over time. A closer look at the data reveals that around 250 entries are specifications of general country information, e.g., general country codes, such as 'EU' (Europe) or 'AP' (Asia/Pacific region), are stated more precisely in the Cymru data and changed, for example, to 'HU', 'AT', or 'JP'. In such a case, this should not be regarded as a real change but as an increase of precision. This means that in summary around 3.7 percent of the entries effectively changed. The dynamics of country codes is therefore present but negligible and thus considered to be quite stable.

In order to analyze the reliability of the Cymru dataset and its country codes, the MaxMind database is used. The MaxMind website provides GeoLite Country (IPv4 and IPv6) as well as a linked GeoLite ASN (IPv4 and IPv6) databases. This is a GeoIP service which might also not be 100 percent reliable but could still be seen as more accurate than country codes of the RIRs. The databases are updated on the first Tuesday of every month. Therefore, they originate from the 5th of February 2013. This results in a difference of 13 days between the Cymru and MaxMind data. All databases for IPv4 and IPv6 were merged together to create a list of all ASNs with their underlying GeoIP location. If more than one country was mentioned for one and the same ASN, all entries related to that ASN were discarded. This was done in order to achieve unambiguous results. Remaining entries that contain a country code such as "satellite provider" or "anonymous proxy" were also dismissed. To ascertain the level of data reliability, the following procedure was applied: at first, the intersection of joint ASNs in both the Cymru and the MaxMind datasets was identified. Overall they share 17,857 common ASNs. In the next step, the intersection of ASNs was checked for changes in terms of the underlying country. The country changes for 814 ASNs, i.e., a fraction of 4.55 percent of the ASs. This difference is not excessively severe. Even if the country code is not highly reliable, the majority of information seems to be correct. In conclusion, the country code information derived from Cymru is reliable enough to conduct a deeper analysis without possible errors caused by GeoIP data.

4 Geographic Resilience Ranking of ASs

The approach proposed in this paper is based on a similar method designed by Roberts et al. (2011). This paper will update some of their results as well as include more metrics that are reflecting different dimensions of Internet robustness. We propose to include the following metrics into the geographic resilience score for specific countries:

1. Absolute number of ASs per country;
2. Number of ASs per a square kilometer per country;
3. The ratio of the number of ASs to the number of inhabitants per country;
4. Number of ASs in relation to the population density of a country;
5. Absolute number of IP addresses per country;
6. Ratio of the number of ASs per number of IP address per country;
7. Number of IP addresses per capita per country;
8. Risk score of becoming a target for cyber-attacks;
9. World Press Freedom index.

In the following sections we will describe the motivation for choosing each metric as well as the data sources used to gather the necessary information. Only those countries and their underlying coding were considered as relevant which are listed in the ISO 3166 coding list on Wikipedia (Wikipedia ISO 3166 Coding List, 2013). Of all 268 countries mentioned in the coding list of Wikipedia, 231 are also mentioned in the data provided by Team Cymru. This leads to 37 missing countries which are either not contained in the dataset or do not have an AS on their own. For the analysis, a rank is assigned to each country for each used metric resulting in nine distinct rankings for each country. This positions the country according to the value of the particular metric. For example, the country with the highest

number of ASs will be ranked on position one while the country with the lowest number of ASs will have the position number 231. In the end, all assigned ranks for each country will be summed up and basing on the sum of ranks a new final rank will be calculated. This will give an indication of the risk for a country to be disconnected from the Internet. The rank-based approach was chosen due to its simplicity and easy practical application.

4.1 Geography-based Metrics

Our proposed resilience score combines several geography-based metrics which are the absolute number of ASs per country, the number of ASs per square kilometer per country, the ratio of the number of ASs to the number of inhabitants per country and the number of ASs in relation to the population density of a country. For this purpose, the geographic area in square kilometers, the population and the population density for every country was determined via Wikipedia.

The first metric is the only absolute one. The total number of ASs indicates how complicated it might be to cut a country off the Internet. A higher absolute number of ASs means that there are more entities which provide Internet infrastructure. It is an important but not sufficient metric in terms of robustness. In general, this metric favors countries that span a large area. They usually have a higher number of ASs present, while small countries and islands possess only few ones due to their small size. Therefore, the absolute number of ASs per country might be misleading. A small country such as the Vatican having an area of 0.44 square kilometers will never have such a large number of ASs as Italy spanning an overall area of 301,338 square kilometers.

Hence, a ratio is applied which takes into account the area a country embraces. It calculates the number of ASs that are present per square kilometer of land – the area ratio. It considers the size of a country and thus the area which needs to be covered by the number of ASs present in that country. It is assumed in this approach that in general the bigger a country is, the more ASs are usually needed to provide reliable Internet access. This time small countries have an advantage. Even if they have just one AS due to their small size, their ratio is usually higher than in case of large countries.

This results in the circumstance that not just the area of a country is important. It is necessary to address countries with a large area but a small population size in a better way. For example, Greenland is a country that has only a small number of present ASs and spans at the same time a large area. Therefore, it is disadvantaged in both metrics mentioned above. Considering its low population density, there even seems to be no need for a better provision of ASs. To address those countries more efficiently, two additional metrics are applied. A ratio related to the population gives an indication on how many people have to share an AS (population ratio). If the available number of ASs cannot serve the number of inhabitants, this could, for example, lead significantly faster to traffic overloads, which in turn might cause a disturbance of the networks. Therefore, this ratio needs to be in balance, too.

Taking into account only the metric based on the population size of a country might be misleading because it also favors small countries. This kind of countries usually have a small number of inhabitants due to space limitations and therefore often a favorable population ratio. Thus, in addition the quotient of the number of ASs and the population density is used (density ratio). Again, this metric refers to the geographical size but takes the existing population size into account at the same time. Therefore, it calculates the area that is covered by an AS per inhabitant. Densely populated countries have a higher demand in terms of AS availability. While a solely population-based metric produces an advantage for small countries, this is not true for the population density. Here, big countries with a low population density have an advantage. However, this is only the case if they simultaneously have a sufficient number of ASs in their geographical area. A combination of these metrics will balance the geographical characteristics in such a way that only those countries will be on the top of the final list which are superior in all areas.

4.2 Technology-based Metrics

Roberts et al. (2011) argue that complexity of a national network infrastructure is defined by two factors – the number of ASs per IP address and the number of IP addresses located away from the core of the network. While the researchers model the first factor as a simple ratio of the number of ASs over the number of IPs in a single country, the second one requires identification of all links between ASs and IPs. Consequently, we suggest extending the geography-based metrics with the following IP-related metrics:

1. The absolute number of IP addresses.
2. The ratio of the number of ASs per IP address.
3. The number of IP addresses per capita.

The first two metrics account for size and complexity of a national network infrastructure. The last one is necessary to better address countries with a low number of IPs as the absolute metric favors countries with high population.

Two online services provide regularly updated statistics about IP addresses on a country level – MaxMind (2013c) and BGPExpert (2013). Both datasets were accessed on July 26th, 2013. This paper uses the statistics collected by MaxMind because of the higher precision of data provided by the service. BGPExpert counts IP addresses per country in millions, significantly decreasing the precision of measurements for small countries.

However, to validate the dataset provided by MaxMind, it was cross-checked with the statistics from BGPExpert for regions with a total number of IP addresses exceeding one million (80 regions in total). Although an average deviation was only 3.7 %, for some countries results differed significantly, for example, Moldova – 23.5 %, France – 20.5 %, Puerto Rico – 16.4 % and Czech Republic – 11.2 %. The average deviation may be explained by a disparity of the dates when the datasets were last updated – MaxMind's on 07/02/2013 and BGPExpert's on 07/24/2013. The reason for high deviations in some specific cases can also originate from a difference in algorithms of a geolocation process. Notably, about 20 million IP addresses belong to the European Union (EU) in the dataset collected by MaxMind, whereas BGPExpert is more successful in revealing particular countries – only 6 million referred to the general EU region and not to some specific member of the union. In the further analysis the records for the EU and other macroregions (Asia, Africa, etc.) as well as other unidentified geographical regions (records for anonymous proxy and satellite providers) were excluded from the analysis.

The sanitized MaxMind's dataset contains 246 rows, while the dataset for ASs only 231. The redundant records represent small administrative territories which were not considered in the ranking due to the absence of any ASs within their borders.

4.3 Control-based Metrics

The geographic resilience score should also be extended by including metrics which account for distribution of control over ASs in a country. This control can be considered as control by organizations administrating ASs, Internet service providers (ISPs) and as a political control of a government over ISPs. In the beginning, we chose to use the absolute number of ISPs and the relative number of ASs per ISP in a country as a proxy for the level of control of organizations administrating ASs. The search for a publicly available list of ISPs across countries yielded two sources of information. The first was the World Factbook issued on a yearly basis by the Central Intelligence Agency (CIA) of the United States of America (2013). The second was the online service Whoisthisip.com (2013).

The statistics for ISPs in the World Factbook was available only for the years 2002 and 2004. Further cross-check of the total number of ISPs per country provided by the CIA and the Whoisthisip.com service generated dramatic deviations. Even though the most obvious explanation for this is that the dataset published in the World Factbook is outdated, figures from the online geolocation service were abnormally high. To perform an additional validity test, two countries were selected for a scrutiny. The choice was made in favor of Australia and New Zealand because governmental statistical services of these countries collect and officially publish information about the national network infrastructures including the number of ISPs (Australian Bureau of Statistics, 2012; Statistics New Zealand Tatauranga Aotearoa, 2012). Results of the analysis are presented in Table 1.

Country	The World Factbook	Whoisthisip.com (accessed July 26th, 2013)	Governmental statistical agencies[1]
Australia	571	9968	76
New Zealand	36	988	106

Table 1. The total number of ISPs in Australia and New Zealand.

The disparity in the figures was dramatic, making an extension of the ranking with valid metrics impossible. Therefore, the suggested ISP-based characteristics were not considered in the further analysis.

In order to model the political control of a government over ISPs, a metric similar to the one developed by Reynolds and Tamaddon (2011) was included into the ranking. Nonetheless, while their developed metric comprised two factors accounting for censorship and Internet filtering, we propose to use the World Press Freedom Index, published by Reporters Without Borders (2013), as a proxy for a degree of overall freedom that journalists, news organizations, and netizens (active Internet users) experience in a country[2]. The index is based on results of a survey conducted in a form of a questionnaire that was distributed among members of the association, journalists, researchers, jurists, human rights activists and non-governmental partner organizations all around the world. The respondents evaluated a state of the following entities in countries: pluralism, media independence, environment and self-censorship, legislative framework, transparency and infrastructure. A cumulative index assesses media freedom in a single country. In general, the lower the index the higher the freedom of press in a particular country. The last available version of this report was published on 01/30/2013 and contained a ranking of 173 countries. This means that there are 58 missing records in comparison to the dataset for ASs. In the absence of a reasonable approach to fill the gaps and in order not to lose a larger portion of information, a purely technical ranking was calculated apart from the final combined ranking. This ranking considered only the network infrastructure metrics and consisted of 231 countries.

The last metric, suggested for inclusion in the geographical AS resilience ranking, is an average attractiveness of ASs in a country to hackers. The idea behind this metric is that certain parts of a national network infrastructure offer attackers a more significant potential payoff than others. Hence, a risk of becoming a target for attacks related to these segments is higher. Let us consider an example of two countries – one with a prevailing financial sector in an economy and one with a more diversified economy. If the total number of ASs in these countries is similar, then in the first country a large portion of backbone networks will be managed by financial organizations which face frequent hacker attacks (IBM, 2013; Symantec, 2013). As a result, the first country is exposed to a higher risk of an

[1] Both statistical governmental agencies consider only ISPs which provide access to the Internet to more than 1000 customers.

[2] The Reporters Without Borders (2013) organization defines the World Press Freedom Index as a measure of "*the degree of freedom that journalists, news organizations and netizens enjoy in each country, and the efforts made by the authorities to respect and ensure respect for this freedom*".

Internet connectivity outage because of simultaneous failures of a bigger number of ASs caused by massive cyber attacks.

Furnell (2002) defines four components which determine a motivation of hackers: financial payoff, curiosity, notoriety and revenge. An ideal approach for an estimation of the probability of an AS becoming a target of a cyber attack would include an evaluation of these four components for all ASs. However, this can hardly be performed due to the number of ASs and lack of publicly available information on the data that is hosted on them. This paper proposes a simplified approach which is based on statistics of cyber attacks collected by major information technology security vendors. The approach includes three distinctive steps:

1. Collection of data on industries which suffered most of the cyber attacks conducted over the last year.
2. Classification of all ASs by industry.
3. Assignment of a score indicating the probability of becoming a target to each AS depending on the industry it belongs to.

The first step consisted of an analysis of reports published by major IT security vendors. Eight latest of the available reports were downloaded on 19/05/2013 and issued by IBM (2013), Symantec (2013), HP (2012), Cisco (2013), CheckPoint (2013), TrustWave (2013), Trend Micro (2012) and Radware (2012). Only two of them contained information about the distribution of cyber attacks across industries, and namely the reports provided by IBM and Symantec (see Table 2 and Table 3). Both firms concluded that companies related to the manufacturing, financial and non-traditional (IT & Telecom) services accounted for about 2/3 of all cyber-attacks in 2012. The categories used by Symantec in their report correspond to the Standard Industrial Classification (SIC) (2011), while IBM did not reveal which particular classification they used. Due to this the further analysis was based on the Symantec distribution of cyber attacks.

Industry	Share, %
Manufacturing	26,5
Finance and insurance	20,9
Information and communication	18,7
Health and social services	7,3
Retail and wholesale	6,6

Table 2. Distribution of cyber-attacks across industries (IBM, 2013).

Industry	Share, %
Manufacturing	24
Finance, Insurance, Real estate	19
Services, non-traditional	17
Government	12
Energy/Utilities	10
Services, professional	8
Aerospace	2
Retail	2
Wholesale	2
Transportation, Communications, Electricity, Gas	1

Table 3. Distribution of cyber-attacks across industries (Symantec, 2013).

To complete the second step we used a keyword-based approach to classify ASs by industries for categories concordant to the Standard Industrial Classification System (SIC) (2011). The effective rate of classified ASs reached 13,961/54,773 = 25%. The best result was achieved for African countries –

more than 40% of ASs were classified by industry, while for the countries located in Europe and South America the classification rate was only 16%. After the reclassification each AS in the dataset got a specific score based on an industry it relates to. ASs classified into the top three industries and having faced 60% of attacks received the score 3. ASs categorized into governmental, energy/utility and professional service groups got 2. ASs of the rest of industries were assigned 1. All ASs which were left unclassified got the average score of 1.41. Lastly, the resulting dataset was used to calculate an average score for a single country.

5 Results and Conclusion

5.1 Results

As already mentioned above, for the analysis a rank was assigned to each country for each of our proposed nine metrics, which positioned the country according to the value of the regarded metric. Finally, all assigned ranks for each country were summed up and based on the sum of ranks a new rank was calculated which constitutes the final rank. Excerpts from the rankings of the ten countries with the most resilient as well as the ten countries with the least resilient networks are displayed in Table 4. The ranking which does not consider the results drawn by the Press Freedom Index due to missing information for some countries is shown in Table 5.

Position	Nation
1	Latvia
2	Switzerland
3	Romania
4	Poland
5	Austria
6	New Zealand
7	Ukraine
8	Slovenia
9	USA
10	Sweden
	…
221	Reunion
222	Turkmenistan
223	Cape Verde
224	Chad
225	Ethiopia
226	South Sudan
227	Guinea-Bissau
228	Eritrea
229	Yemen
230	Senegal
231	Korea, D.P.R of (Nord)

Table 4. Combined ranking for geographical resilience per country without considering the press freedom index.

In summary, mostly European countries are on the top of the ranking. Of all considered European countries, 56 percent can be found at the top quarter of the combined ranking list. No matter what kind of metric is chosen, European countries always dominate the top of the list. North America seems to be somewhere in between. There are resilient countries such as the USA and Canada but also several countries with a low ranking result such as Cuba and Haiti. Considering the size of these well-connected countries, it becomes evident that a large portion of the area of North America is reliably covered. This is caused by the dominant position of the USA and Canada. The same is true for

Oceania since Australia and New Zealand – the two best-connected countries in this continental area – account for around 93 percent of the entire region.

Position	Nation
1	Switzerland
2	Latvia
3	Austria
4	Poland
5	Romania
6	New Zealand
7	Sweden
8	Czech Republic
9	Slovenia
10	USA
…	…
169	Congo, The Dem. Rep. Of (Zaire)
170	Burundi
171	Somalia
172	Chad
173	Turkmenistan
174	Guinea-Bissau
175	Ethiopia
176	Senegal
177	Eritrea
178	Yemen
179	Korea, D.P.R of (Nord)

Table 5. Combined ranking for geographical resilience and freedom from control per country based on all proposed metrics.

In case of Asia, the corresponding countries are almost evenly distributed over the whole combined ranking list (except for the dominant position of the second quarter of the list). It could be the case that there is a trend shifting away from the bottom towards the top, which needs to be checked with historical or future data. Assuming that this is the case, it can be seen as a sign that Asia is already in the advanced state of transition to a more robust Internet connection status. Similarly, South America also seems to be in an early stage of transition. While the top quarter of the combined ranking list contains only three countries from this area, the bottom quarter still accounts for one third of all the South American countries. The shift towards more robustness is observable. African countries fall behind in all areas. More than 65 percent of all African countries can be found at the bottom quarter of the combined ranking list. An exception to this is Gibraltar which has a lot of ASs in relation to its size. This supports the hypothesis that this country is an important connection point for Internet communications. All in all the status quo of each country seems to be closely related to its economic development. Therefore, Africa seems to be still just at the beginning of a period of change.

The inclusion of the degree of political freedom into the ranking led to minor changes in positions of countries on the top and in the bottom of the list. Countries best connected to the Internet mostly represent Central, Northern, and Western Europe, while the most vulnerable networks belong to the African and Asian countries.

The aggregated results per continent show that the largest network infrastructure is located in North America and mostly formed by the U.S. and Canada (see Table 6). On a country level Europe represents the least vulnerable region to Internet access disruption. 56 percent of all European countries are in the first quarter of the list. The second largest network infrastructure belongs to Asia. This fact can be explained by the number of users and very rapid spread of Internet technologies in the region. However, the level of freedom is the lowest compared to other continents.

Continent	# of countries	ASN total	ASN per country	# IP total	# IP per country	Average risk score	Average Freedom Index
Africa	55	968	18	54,129,047	984,165	1.82	35.108
Asia	52	9,835	189	900,677,859	17,320,728	1.82	45.091
Europe	50	17,662	353	677,105,716	13,542,114	1.89	18.921
North America	34	22,338	657	1,697,272,703	49,919,785	1.77	25.267
Oceania	22	1,819	83	55,833,685	2,537,895	1.60	22.649
South America	18	1,352	75	119,957,389	6,664,299	1.90	28.688

Table 6. Aggregated results for geographical resilience analysis.

The network infrastructure of Oceania is very similar to that of North America and is mostly constituted by two countries Australia and New Zealand. Although South America, similarly to Asia, looks like a rapidly developing region, the relatively small number of networks and certain problems with political freedom in the region push countries on the continent to the bottom of the list. More than 55 percent of South American countries are in the third quarter. African countries mostly received positions in the bottom of the list. They took 67 percent of positions in the bottom quarter of the ranking.

Further analysis revealed a high correlation of the metrics representing technical characteristics of network infrastructures on a country level. The results of the analysis are summarized in Table 7. For the metrics based on the total number of ASs per country, total number of IPs per country and ratio of the total number of ASs per country over a population density, the correlation was in the range from 0.686 to 0.869. For metrics based on the number of AS per capita and the number of IPs per capita, the correlation reached the value of 0.748. This can be a basis for a simplification of the ranking by exclusion of some metrics to optimize calculations in future work.

	Rank Freedom	Rank Risk Score	Rank Risk Score	Rank AS/IP	Rank #IPs	Rank AS Dens.	Rank AS Pop.	Rank km^2	Rank ASN #
Rank Freedom	1	0.0471	0.523	-0.052	0.142	0.174	0.567	0.457	0.191
Rank Risk Score		1	-0.092	0.151	-0.232	-0.231	0.004	0.043	-0.228
Rank Risk Score			1	-0.436	0.488	0.173	0.748	0.638	0.353
Rank AS/IP				1	-0.748	-0.322	0.121	0.110	-0.400
Rank #IPs					1	0.686	0.014	-0.002	0.869
Rank AS Dens.						1	-0.016	-0.331	0.778
Rank AS Pop.							1	0.821	0.129
Rank km^2								1	0.088
Rank ASN #									1

Table 7. Correlations between the metrics.

5.2 Limitations and Future Work

This study presented a new approach to the resilience assessment of the national network infrastructures. A limitation is that the metrics included into the final ranking so far do not incorporate the existing links between ASs and therefore do not reflect the real complexity of the network structure. Nevertheless, even in the presence of a yet hypothetically appropriate toolset for the revelation of structure, a major problem still involves the identification of all links between ASs. For instance, Dimitropoulos et al. (2007) argue that they were able to discover only 38.7 % of existing peering connections. Furthermore as part for future work it would be additionally possible to include the physical layer in terms of cable maps next to this AS-layer. Another limitation is the lack of reliable statistics on a country level for ISPs. To the best of our knowledge, currently there are no valid services or data sources that can be used for the analysis of a control distribution over national networks.

Moreover, the inclusion of the World Press Freedom Index into the ranking creates a challenge. The reason for this is that not all administrative entities physically hosting ASs are independent countries. As a solution for this limitation, these territories can be mapped to parent countries and as a result receive the same ranking for this metric. Finally, the applied approach for the evaluation of attractiveness of ASs for hackers was based on the distribution of cyber attacks in the previous year. Thus, the metric based on this information could be either revised on a yearly basis or calculated using some average figures for a chosen period of time.

5.3 Discussion and Implications

For the global information and network society it becomes increasingly critical to assess how vulnerable specific geographical regions are to Internet access disruptions. Closely related are the options and limits of censorship-based impediments implemented by local governments. For this purpose this paper presented a metric assessing the geographical Internet resilience based on a country level. This metric encompasses several geographic-, technology- as well as control-based indices, combining them into a single, rank-based score that estimates the Internet resilience of a particular country compared to others.

Such a comparison can serve as a systematic assessment for regions of the Internet that are at risk and could support international politics, Internet businesses or freedom activists with indicators for improving Internet resilience and censorship resistance by focusing their activities on critical regions. Moreover, the ongoing debate on net neutrality (Deeb, 2009) would benefit from investigating not only the benefits and drawbacks of providing quality of services to end users but also to investigate geographic factors and political conditions that influence the reliability and speed of accessing a particular content, including risks of not being able to access some content at all.

This ranking will also help to understand and possibly counter geographic options for censorship as well as spying techniques and to assess risks for deploying new services in the network economy. The development and deployment of practical privacy-enhancing technologies can also benefit from insights from our index. Mission critical services should be located in regions with robust and censorship-free network access.

5.4 Conclusion

This paper introduces a proposal for a geographic AS resilience ranking that gives information on the resilience of different countries to Internet disruptions. Our approach broadened the range of factors by including aspects such as the World Press Freedom Index. The final result showed a dominance of European countries in terms of the combination of all metrics, while mostly African countries lack behind in all areas. This confirms to some extent the hypothesis that a national network infrastructure is an outcome of several social, political and technical decisions. All the countries in the bottom of the list refer to regions with poorly developed economics and low level of political freedom. An open question remains what factors play the key role in the development of the network structure on a country level, which can also become a topic for future research.

References

Apnic Aut-Num (2013). Object Template. Available from: http://www.apnic.net/apnic-info/whois_search/using-whois/guide/aut-num/. Last accessed 04/22/2014.

Australian Bureau of Statistics (2012). Internet Activity, Australia. December 2012. Available from: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/FA32C4F74EB0F063CA257B4700137847/. Last accessed 04/22/2014.

BGPExpert (2013). IP Addresses by Country. Available from: http://www.bgpexpert.com/addressespercountry.php. Last accessed 04/22/2014.

Brown, M. (2008a). Ike Brings Biggest Multi-State Internet Outage since 2003. September. Available from: http://www.renesys.com/2008/09/ike-brings-biggest-multistate/. Last accessed 05/06/2014.
Brown, M. (2008e). Pakistan hijacks YouTube. February. Available from: http://www.renesys.com/2008/02/pakistan-hijacks-youtube-1/. Last accessed 05/06/2014.
CAIDA AS Names (2012). CAIDA AS Rank Projekt. Available from: http://as-rank.caida.org/?%20mode0=as-dump-info,%202012. Last accessed 04/22/2014.
CheckPoint (2013). Check Point Security Report. Available from: http://www.checkpoint.com/campaigns/security-report. Last accessed 04/22/2014.
CIA. (2013). The World Factbook. Available from: https://www.cia.gov/library/publications/the-world-factbook/. Last accessed 04/22/2014.
Cisco (2013). Annual Security Report. Available from: http://www.cisco.com/en/US/prod/vpndevc/annual_security_report.html. Last accessed 04/22/2014.
Cowie, J. (2011a). Japan Quake. March. Available from: http://www.renesys.com/2011/03/japan-quake/. Last accessed 05/06/2014.
Cowie, J. (2011b). What Libya Learned from Egypt. March. Available from: http://www.renesys.com/2011/03/what-libya-learned-from-egypt/. Last accessed 05/06/2014.
Cowie, J. (2011c). Egypt Leaves the Internet. Available from: http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml. Last accessed 05/06/2014.
Cowie, J. (2012c). Syrian Internet is Off the Air. Available from: http://www.renesys.com/blog/2012/11/syria-off-the-air.shtml. Last accessed 05/06/2014.
Cowie, J., Ogielski, A. and Premore, B. J. (2002). Internet Worms and Global Routing Instabilities. Available from: http://www.renesys.com/wp-content/uploads/2013/05/renesys-spie2002.pdf. Last accessed 05/06/2014.
Cowie, J., Popescu, A. and Underwood, T. (2005). Impact of Hurricane Katrina on Internet Infrastructure. http://www.renesys.com/wp-content/uploads/2013/05/Renesys-Katrina-Report-9sep2005.pdf. Last accessed 05/06/2014.
Deeb, K., O'Brien Sr., S.P., Weiner, M.E. (2009). A Survey on Network Neutrality: A new form of discrimination based on network profiling. International Journal of Networking and Virtual Organisations, 6(4):426-436.
Dimitropoulos, X., Krioukov, D., Fomenkov, M., Huffaker, B., Hyun, Y. and Riley, G. (2007). AS Relationships: Inference and Validation. ACM SIGCOMM Computer Communication Review, 37(1): 29-40.
EdgeScape (2013). http://www.akamai.com. Last accessed 04/22/2014.
Furnell, S. (2002). Cybercrime: Vandalizing the information society. Addison-Wesley, Boston.
Hawkinson, J. and Bates, T. (1996). Guidelines for Creation, Selection, and Registration of an Autonomous System (AS). Technical standards document RFC1930, March. Available from: http://tools.ietf.org/html/rfc1930.
HP (2012). Cyber Risk Report. Available from: http://www.hpenterprisesecurity.com/collateral/whitepaper/HP2012CyberRiskReport_0213.pdf. Last accessed 04/22/2014.
IBM (2013). IBM Security Services Cyber Security Intelligence Index. Available from: http://www-935.ibm.com/services/us/en/security/infographic/cybersecurityindex.html. Last accessed 04/22/2014.
ITU (2012). Measuring the Information Society. International Telecommunication Union Report. Available from: http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2012.aspx. Last accessed 04/20/2014.
IxMapper. http://www.ixiacom.com/products, 2013. Last accessed 04/22/2014.
Lakhina, A., Byers, J.W., Crovella, M. and Matta, I. (2003). On the Geographic Location of Internet Resources. IEEE Journal on Selected Areas in Communications, 21(6): 934-948.
Madory, D. (2012a). Lebanon Loses Lone Link. July. Available from: http://www.renesys.com/2012/07/large-outage-in-lebanon/. Last accessed 05/06/2014.
Madory, D. (2012b). Blast in Turkey Impacts Iran, Iraq. October. Available from: http://www.renesys.com/2012/10/blast-in-turkey-impacts-iran-i/. Last accessed 05/06/2014.
Madory, D. (2013). Faraway Fallout from Black Sea Cut. February. Available from: http://www.renesys.com/2013/02/faraway-fallout-from-black-sea/. Last accessed 05/06/2014.
Maxmind (2013a). GeoLite Free Downloadable Databases. http://dev.maxmind.com/geoip/geolite. Last accessed 04/22/2014.
Maxmind (2013b). GeoIP Ort - Präzision für ausgewählte Länder. http://www.maxmind.com/de/city_accuracy. Last accessed 04/22/2014.
MaxMind (2013c). Allocation of IP Addresses by Country. Available from: http://www.maxmind.com/en/techinfo. Last accessed 04/22/2014.

National Research Council (2003). The Internet Under Crisis Conditions: Learning from September 11, National Academies Press.
Paul, D., Sarkar, S.K. and Mondal, R. (2013). Optimization of Core Network Router for Telecommunication Exchange. American Journal of Networks and Communications, 2(1): 1-8.
Popescu, A. (2008b). Deja Vu All Over Again: Cables Cut in the Mediterranean. December. Available from: http://www.renesys.com/2008/12/deja-vu-all-over-again-cables/. Last accessed 05/06/2014.
Potaroo (2013). 32-bit Autonomous System Number Report. http://www.potaroo.net/tools/asn32/. Last accessed 04/22/2014.
Radware (2012). Global Application and Network Security Report. Available from: http://www.radware.com/Resources/rclp.aspx?campaign=1630844. Last accessed 04/22/2014.
Renesys (2003). Impact of the 2003 Blackouts on Internet Communications. Available from: http://www.renesys.com/wp-content/uploads/2013/05/Renesys_BlackoutReport.pdf. Last accessed 05/06/2014.
Reynolds, C. and Tamaddon, S. (2011). Network-Political Resiliency. Technical Report. Available from: http://www.chloejreynolds.com/portfolio_materials/NPR-Paper.pdf. Last accessed 04/22/2014.
Roberts, H., Larochelle, D., Faris, R. and Palfrey, J. (2011). Mapping Local Internet Control. Computer Communications Workshop, Hyannis, CA.
SIC (2011). Standard Industrial Classification. Available from: http://www.sec.gov/info/edgar/siccodes.htm. Last accessed 04/22/2014.
Statistics New Zealand Tatauranga Aotearoa (2012). Internet Service Provider Survey. Available from: http://www.stats.govt.nz/browse_for_stats/industry_sectors/information_technology_and_communications/ISPSurvey_HOTPJun12/Definitions.aspx. Last accessed 04/22/2014.
Sterbenz, J. P. G., Hutchison, D., Çetinkaya, E. K., Jabbar, A., Rohrer, J. P., Schöller, M., Smith, P. (2010). Resilience and Survivability in Communication Networks: Strategies, Principles, and Survey of Disciplines. Computer Networks, 54(8), pp. 1245-1265.
Symantec (2013). Internet Security Threat Report 2013. Available from: http://www.symantec.com/security_response/publications/threatreport.jsp. Last accessed 04/22/2014.
Team Cymru (2013). Internet Security Research and Insight. Available from: http://www.cymru.com/. Last accessed 04/22/2014.
TrendMicro (2012). TrendLabs Annual Security Roundup. Available from: http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/reports/rpt-evolved-threats-in-a-post-pc-world.pdf. Last accessed 04/22/2014.
TrustWave (2013). Trust Wave Global Security Report. Available from: https://www2.trustwave.com/2013GSR.html?sl=gsr-2012. Last accessed 04/22/2014.
Underwood, T. (2008c). Wrestling With the Zombie: Sprint Depeers Cogent, Internet Partitioned. October 31. Available from: http://www.renesys.com/2008/10/wrestling-with-the-zombie-spri/. Last accessed 05/06/2014.
Wählisch, M., Schmidt, T.C., de Brün, M. and Häberlen, T. (2012). Exposing a Nation-Centric View on the German Internet – A Change in Perspective on AS-Level. Passive and Active Measurement, Springer.
Whoisthisip.com (2013). Available from: http://whoisthisip.com/internt-service-provider-isp-ip-usage.php. Last accessed 04/22/2014.
Wikipedia ISO 3166 Coding List (2013). ISO-3166-1-Kodierliste. http://de.wikipedia.org/wiki/ISO-3166-1-Kodierliste. Last accessed 04/22/2014.
Wilcox, S. (2007). Quaking Tables – The Taiwan Earthquakes and the Internet Routing Table. http://www.renesys.com/wp-content/uploads/2013/05/ripe-quaking-tables.pdf, 2007. Last accessed 05/06/2014.Reporters Without Borders (2013). World Press Freedom Index. Available from: http://en.rsf.org/press-freedom-index-2013,1054.html. Last accessed 04/22/2014.
Zmijewski, E. (2008d). Georgia Clings to the 'Net. August. Available from: http://www.renesys.com/2008/08/georgia-clings-to-the-net/. Last accessed 05/06/2014.