A Framework for Nation-Centric Classification and Observation of the Internet

Matthias Wählisch
Freie Universität Berlin

Sebastian Meiling
HAW Hamburg, Informatik

Thomas C. Schmidt
HAW Hamburg, Informatik

ABSTRACT

The Internet has matured to a mission-critical infrastructure, and recently attracted much attention at political and legal levels in many countries. Civil actions regarding the Internet infrastructure require a thorough understanding of the national components of the global Internet to foresee possible impacts of regulations and operations at a country-level. In this paper we report on a methodology, tool chain and results for identifying and classifying a 'national Internet'. We argue for the importance to consider individual IP-blocks instead of prefixes and quantify the effects of our proposed approach. The methods have been applied to identify a 'German Internet', but are designed general enough to work for most countries, as well.

Categories and Subject Descriptors

C.2.1 [**Computer-Comm. Networks**]: Network Architecture and Design—*Network topology, Internet*

Keywords

AS-level Classification, Methodology, Nation-state Internet Routing, German Internet Graph

1. INTRODUCTION

The Internet was originally shaped to offer transmission services to those who were willing to inter-connect. Its scalable architecture aimed at supporting a global information infrastructure without contributions at a national level, nor governmental support. Meanwhile the Internet has matured to a mission-critical infrastructure for enabling key operations of public administration and business affairs, of research, education, and entertainment within individual countries. National policies and legal aspects apply to such a building block of the central infrastructure, as well as security, safety and reliability concerns.

Addressing any of these aspects requires a significant knowledge and understanding of the characteristic features the relevant Internet infrastructure is comprised of. Rigorous insight into the country-wise nature of the Internet thus carries fundamental importance and it is somewhat surprising that only recently the inter-network structures of nations attracted attention [1].

Linked herewith, the primary contribution of our work is to identify and classify the Internet players of a country by reliable methods, as well as the minimal routing infrastructure that spans these constituents.

In this paper, we analyze the Internet structure on an IP-block basis, and quantify the relevance of our method as compared to a pure prefix-based approach. A classification of networks with national relevance is then obtained from matching blocks to Autonomous Systems (ASes). To gain further insight into the network characteristics of a country, we heuristically derive a sectoral classification for relevant ASes. Finally, by adding routing information, we are able to generate, visualize and analyze [2] the structure of communication flows between relevant public and business sectors.

In the remainder, we summarize the identification and classification steps along the tool chain (§ 2), and exemplify our results discursively (§ 3).

2. FROM IP MEMBERS TO ASES

Our primary objective is to identify all parts of the global Internet that host organisations of a specific country. We argue that the appropriate granularity must be ranges of IP addresses that are assigned to customers from that country. This starting point differs from the prefix-based approach of [1] and will lead to a significant change in results (see § 3). To demonstrate and validate our nation-centric approach, we chose our home country Germany (*DE*). Figure 1 depicts the algorithmic chain to identify national IP-blocks and map them to prefixes
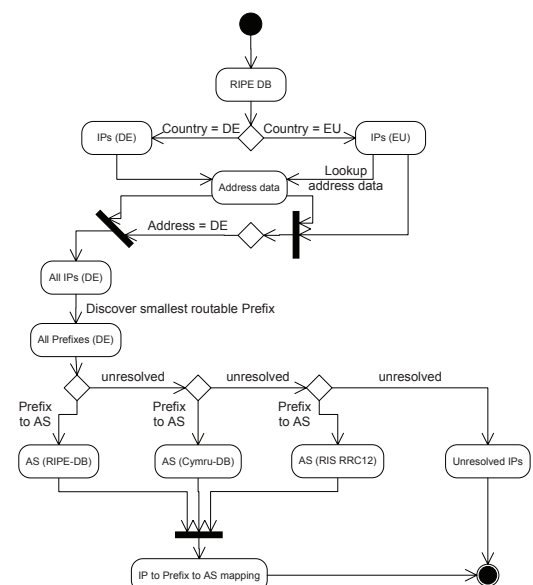

Figure 1: Tool chain for identifying national ASNs.
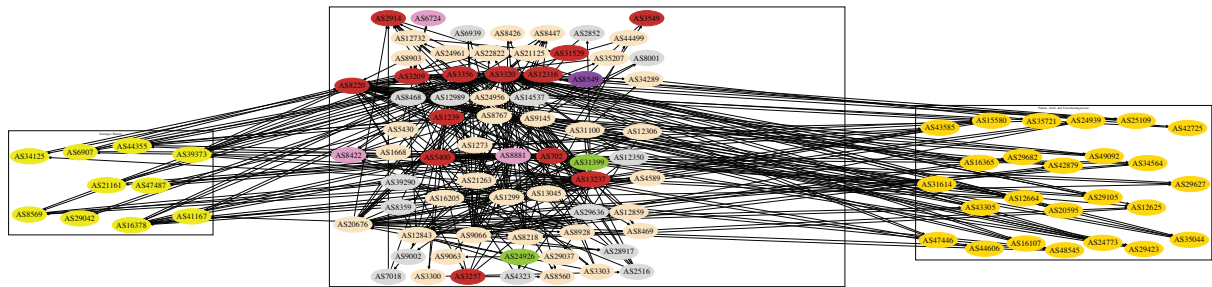
and then AS numbers (ASNs).

Figure 2: [Best viewed in color] Communication flows between traders and financial services

We start by extracting all `inetnum`-objects, i.e. IP-blocks from the database of the regional registry (RIR) that carry the mandatory `country` attribute of either *DE* or *EU*. Additionally, we collect address data for the associated `admin-c`- and `org`-objects. The reason for the latter is twofold: (a) to further resolve *EU* IP-blocks to *DE*, and (b) to verify the *DE* classification of IP-blocks. The result is a list of all IP-blocks allocated by organizations in Germany.

Next, we determine the smallest IP-Prefix for each IP-block and its corresponding AS. Note, that prefix lengths are subject to aggregation and thus depend on the point of observation. Assuming, that RIRs provide the most detailed prefix mapping, it is reasonable to query the RIPE DB[1]. However, using RIPE-DB alone is not sufficient for resolving all IP blocks to prefixes and ASNs. Therefore, we also query the databases of Team Cymru[2] and the route collector RRC12 of RIPE RIS[3] at DE-CIX, Frankfurt/Germany.

These three data sources are applied as follows: For each IP-block, we look up the prefix (`route`-object) in RIPE-DB, retrieve its `origin`-attribute, which refers to an `aut-num`-object – the AS number. For unresolved IPs we try Team Cymru next, and finally RIS RRC12. These consecutive steps of IP to ASN mapping minimize the number of unresolvable IP blocks. The resulting list contains the ASes that compose a nation-centric (German) part of the Internet.

Having categorized the nationality of the stakeholders, we add two further classifications to the selected ASNs. We harvest the topological hierarchy, i.e., *tier1*, *large* and *small ISP*, and *stub*, from [3], and determine a sectoral classification for the most relevant ASes. For the latter, we apply an optimized and manually verified keyword spotting to names, descriptions and address fields of the ASes, using general terms like bank; but also specific company names, as keywords to identify certain company ASes. Thereafter we obtained an additional list of the 'relevant national ASes' including branch tags such as *traders*, or *financial services*.

3. RESULTS & DISCUSSION

Our fully automated tool chain was applied in Oct. 2010 and yielded 245524 German IP-blocks. From these 240237 are embedded in 6278 IP-prefixes that belong to 1472 ASes. 5286 IP-blocks ($\approx 2\%$) could not be resolved to a prefix. 283 ASes have been selected as 'relevant' with a sectoral attribute attached.

Our method of starting from IP-blocks rather than

IP-prefixes identifies significantly more prefixes that carry relevance for the country Germany, as is shown in table 1. When considering prefixes alone, only 5243 out of 6278 ($\approx 84\%$) can be identified as German using RIPE-DB, while Team Cymru yields 4395 *DE* ($\approx 70\%$) and 947 *EU* prefixes respectively. Thus a significant fraction of prefixes that route German traffic do not carry country or address attributes from the same country.

	DE	EU	other
IP-Block Approach	6278		
Prefix Approach (RIPE DB)	5243		1035
Prefix Approach (Team Cymru)	4395	947	936

Table 1: IP-block versus -prefix resolution

Adding AS-level routing relations to our selected and classified AS sets allows us to study and visualize very specific topological set-ups. The network graph in figure 2 displays the information flows between traders (left) and financial services (right) in the Internet. Each colored node represents one German AS, with the color indicating a specific branch (red: large ISPs, pink: access providers etc.). All grey nodes represent ASes that belong to another country. Here it should be noted that the minimal spanning routing system of our AS sets adds transit ASes to the graph.

Despite the relatively small number of ASes for these two sectors, it takes quite many transit ASes (shown in the middle box) to interconnect all selected networks. Moreover, taking a closer look at the graph clearly shows that many of these necessary nodes and links actually remain outside of the 'German Internet'. This may indicate that a nation-state perspective on routing in Germany is not well defined in relevant areas.

These results, as well as a large collection of other routing graphs, analysis' and visualizations that we have prepared, reveal uncommon views and insights about those fractions of the Internet that are under civil debate. We hope that a better understanding of 'national Internet infrastructures' will advance the maturity and long-term vision in debates and decision processes.

4. REFERENCES

[1] J. Karlin, S. Forrest, and J. Rexford, "Nation-State Routing: Censorship, Wiretapping, and BGP," arXiv.org, Tech. Rep. abs/0903.3218, March 2009.

[2] M. Wählisch, T. C. Schmidt, S. Meiling, M. de Brün, and T. Häberlen, "Towards a Nation-Centric Understanding of the Internet," in *ACM CoNEXT. Student Workshop.* Nov. 2010.

[3] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-level Topology," *ACM SIGCOMM CCR*, vol. 35, no. 1, pp. 53–61, 2005.

---

[1] http://www.ripe.net/db/

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

[3] http://www.ripe.net/projects/ris/rawdata.html