Who Runs the Internet?
Classifying Autonomous Systems into Industries

Annika Baumann and Benjamin Fabian
Institute of Information Systems, Humboldt University Berlin, Spandauer Str. 1, 10178 Berlin, Germany

Keywords: Internet, Autonomous System, Industry, Classification.

Abstract: The Internet consists of a network of Autonomous Systems (ASs). To understand which kind of organizations control those ASs can help to better assess the Internet structure in terms of economic interests and reliability. The current paper proposes a novel classification approach by combining AS-specific data with business data from the United States Securities and Exchange Commission. Furthermore, more detailed industry classes than in previous works are considered, inspired by the North American Industry Classification System (NAICS). Using our methodology on a recent data set, we were able to classify 56.69 % of the considered ASs into industries. This lays a foundation for our future work on investigating the important players of the Internet backbone as well as their economic interests and risks.

1 INTRODUCTION

The Internet expanded rapidly during the last decade. From 2001 to 2013, the fraction of the world population using the Internet increased from 8.0 % (International Telecommunication Union, 2011) to an estimated 38.8 % (International Telecommunication Union, 2013), with a simultaneous population increase from 6.1 billion (United Nations Population Fund, 2001) to 7.2 billion people (U.S. Department of State, 2013), resulting in approximately 2.8 billion Internet users today versus 0.49 billion in 2001. This rapid growth in users resulted in a heterogeneous and complex system, making analysis and modelling of the Internet structure difficult.

Our paper is part of an on-going research project that is investigating how the Internet of today is structured in terms of economic interests, control and reliability. Who are the important players of the Internet backbone, what are their economic interests and risks with respect to their business models, and what are the implications for reliability, security and privacy as well as political control?

Our first step towards approaching these goals is to classify the important organisations that control Autonomous Systems (ASs) of the Internet according to business categories, which could support future analyses along all of those dimensions. For example, with respect to reliability and security, common methods assess the robustness of the Internet structure based on graphs and modelling the Internet as an abstract complex network consisting of nodes (each representing an AS) that are connected via edges. However, such approaches solely focus on topology-based robustness and so far ignore the highly economically driven character of the Internet, as well as corresponding heterogeneous risks of attack and control.

At an organizational and global routing level of abstraction, the Internet can be considered as composed of ASs. An AS can be defined as “a group of IP networks run by one or more network operators with a single clearly defined routing policy; when exchanging routing information to the outside, each AS is identified by a unique number (Réseaux IP Européens, 2011). The *Internet Corporation for Assigned Names and Numbers* (ICANN) and, via delegation, the *Regional Internet Registries* (RIR) are responsible for registration of these AS numbers (ASNs). The amount of registered ASs increased from roughly 10,000 in the year 2000 to more than 60,000 in 2013 (Potaroo, 2012), which is also another indicator for the substantial increase of Internet complexity.

Classifying the major players of the Internet backbone is an interesting challenge in itself because publicly available business data is sparse. Our approach presented in this article focuses on

Baumann A. and Fabian B..
Who Runs the Internet? - Classifying Autonomous Systems into Industries.
DOI: 10.5220/0004936803610368
In *Proceedings of the 10th International Conference on Web Information Systems and Technologies* (WEBIST-2014), pages 361-368
ISBN: 978-989-758-023-9

analysing the public registration information for AS numbers. Moreover, we present an approach for the classification of ASs into detailed industry classes in order to better understand the organizational and economic patterns of the Internet.

The rest of the paper is structured as follows: Section 2 discusses related work. Section 3 presents the data sources, followed by Section 4 on our methodology. Section 5 presents our results, and Section 6 concludes the paper.

2 RELATED WORK

Some earlier research articles proposed approaches for classifying ASs into various categories. The classification approach used in our paper was initially inspired by the methods employed by Dimitropoulos et al. (2005). Based on an expert system that uses text classification techniques, the authors used organization names to categorize ASs. Each AS was assigned to one or more of the basic classes Internet service providers (ISP), Internet exchange points (IXP), network information centers (NIC), companies providing no Internet service as well as education- and research-, military-, government- and health-related networks. The authors were able to classify 20,598 out of 32,689 ASs in 2005, which corresponds to 63.01%.

Another work (Dimitropoulos et al., 2006) used even more coarse-grained classification categories, namely only large and small ISPs, customer ASs, universities, IXPs and NICs. The method applied was based on the AdaBoost algorithm (Freund and Schapire, 1997) using several attributes (e.g., organization description; number of inferred providers, customers and peers; number of advertised IP prefixes) to classify the relevant ASs into their respective classes. The authors were able to classify 95.3% of 19,537 ASs with an accuracy of 78.1%.

The main focus of the work by Chang et al. (2005) was to estimate traffic volume between individual ASs. For this, the authors classified ASs regarding their initial utility, which resulted into the three classes web hosting, residential access and business access. The methodology used by the authors is different from other work conducted in this area. Instead of investigating an individual AS and assigning it to a class, they created a class and tried to find relevant ASs on the Internet. The authors were able to identify 56% of all BGP-advertised ASs with their approach.

The primary focus of the paper by Dhamdhere and Dovrolis (2011) was to analyse the evolution of the AS ecosystem over the last 12 years. ASs were classified into the classes enterprise customers, small and large transit providers, access/hosting providers and content providers. A decision tree approach was applied for classification. In order to build the training set, for each class 50 ASs were classified manually. Afterwards, the classification was conducted for 42,000 ASs by using the number of customers and the number of peers as independent variables. Classification accuracy for the classes ranged between 76% and 82%.

All of those articles have in common that the proposed classes are not comprehensive and do not resemble real industries. Thus they contribute not much to a better understanding of the industry structure behind the ASs comprising the Internet. Our work addresses this research gap by proposing a classification approach that adopts fine-grained industry classes.

3 DATA SOURCES

3.1 CAIDA

The Cooperative Association for Internet Data Analysis (CAIDA) “is a collaborative undertaking among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure.” (CAIDA, 2011). One project offered by CAIDA is the *AS Rank* project (CAIDA, 2012). It is based on Border Gateway Protocol (BGP) routing data collected by RouteViews (2013) and the RIPE NCC (2013). The list of ASs that is used in our paper contains the information of 59,576 ASs. An excerpt of the dataset can be seen in Figure 1. For the purpose of classifying ASs into industry classes mainly the *org name* attribute was considered as highly relevant.

```
# format: AS number|source|AS name|country|org name|org_id|date
1|ARIN|LVLT-1|US|Level 3 Communications|LVLT-ARIN|20120224
```

Figure 1: Excerpt of CAIDA *AS Rank*.

3.2 SEC

The U.S. Securities and Exchange Commission (SEC) is a government agency in the USA (United States Securities and Exchange Commission, 2013). Its primary purpose is to regulate securities and

enforce federal securities laws. Every company publicly traded in the United States has to file certain documents with the SEC. The Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system makes those filings available to the public. This can be used to gather the Standard Industrial Classification (SIC) code for the company (Figure 2). An SIC code can be directly mapped to an NAICS code using a mapping table (CareerOneStop U.S., 2013). Thus it is possible to uniquely identify the industry of an AS's organization by use of the EDGAR system. A limitation is that only organizations that are listed on a stock exchange in the USA can be found in the system.

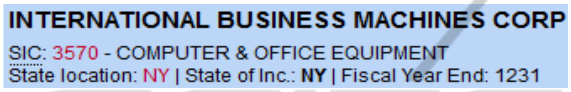

Figure 2: Excerpt of SEC EDGAR result.

3.3 RIR as Information

As an additional information source, data from the RIRs was retrieved. The website *cidr-report.org* contains AS information from all RIRs. It allows searching for individual ASs and returns the information that comes from the WHOIS services of the individual RIRs. In order to simplify the data retrieval process, this website was also used to retrieve AS-specific WHOIS information instead of using the WHOIS services of different RIRs. A sample of such information can be seen in Figure 3.

```
aut-num:        AS6619
as-name:        SAMSUNGSDS-AS-KR
descr:          SamsungSDS Inc.
descr:          Seoul Yeoksam-dong Gangnam-gu   707-19
descr:          135-080
country:        KR
```

Figure 3: Sample of RIR AS information.

4 METHODOLOGY

Figure 4 gives an overview of the process of classifying the ASs presented in this paper. As a first step, the relevant industry classes for the classification approach needed to be defined. Their definition draws from the North American Industry Classification System (NAICS, 2013). Due to the intrinsic online setting of our investigation, special adjustment was necessary, meaning that several of these classes were either merged, dropped or changed. In the case of ASs, some industries are missing at all while some of them are overrepresented. Therefore, the NAICS was only used as a basis for the classification approach in our particular setting. An overview of the classes can be found in the Appendix.

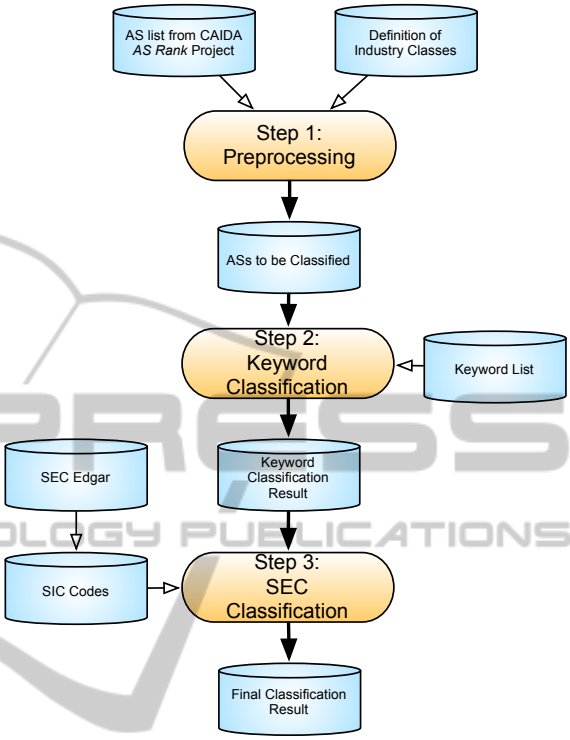

Figure 4: Process of AS industry classification.

Step 1: Preprocessing. The initial AS list included data from the year 2012 and was taken from the CAIDA AS Rank project (CAIDA, 2012); it contained 59,576 ASs. In order to only include reasonable and recent data, the list was preprocessed. At first, the information gathered from the RIRs was used to filter for inactive ASs. This reduced the list by 17,830 ASs, leaving 41,746 ASs to classify. Furthermore, all ASs that did not have an according organization name, i.e., all entries either containing no specification of the underlying organization or being a no registry entry, were removed from the list. Eliminating 1,362 ASs, this step left 40,384 ASs in the list.

Step 2: Keyword Classification. In the next step, a keyword list was created by analysing word and phrase frequencies with the help of an occurrence counting of words, bi-grams and tri-grams. All words and phrases that appeared quite frequently were analysed in more detail. It was assumed that tri-grams needed to occur at least five times, bi-grams ten times and simple words twenty times to be selected for deeper analysis. The rationale behind

this procedure was to include only those words and phrases that are most frequent and therefore important. This makes it possible to classify several ASs at the same time based on a single phrase or keyword. Keywords were mainly defined in such a way that the organization name or a part of it had to comply with the complete keyword. This means that for example in case of the keyword "ship" only the word itself would fit and not "membership" or "ownership". This was done to ensure the reliability of keywords by avoiding undesired mismatches. The selection of keywords itself was randomly cross-checked based on real data to further ensure their reliability and unambiguity. Only those words or phrases were chosen whose unambiguity in relation to industry classification was satisfactory. For example the keyword "Internet service provider" is highly reliable if it comes to sorting into the category ISPs & Networks, while "service provider" might lead to wrong results for the same category. Organizations having a (part of their) name such as "content service provider" would also fit into such a category.

In order to minimize wrong categorizations, an iterative learning process was applied. The procedure was as follows: based on the first selection of keywords, the AS numbers were categorized into the industry classes created so far. Each category was then checked for wrong categorizations. For this purpose, the list of categorized ASs and their underlying organization was reviewed manually. If the categorization of an ASN was wrong, the reason was identified and eliminated with the help of refined or discarded keywords. This procedure helped to ensure that only those keywords remained that are at the same time reliable and general. In order to check for further yet not identified keywords, a list was generated that contained all non-categorized ASNs. This list was then manually checked for further keywords at each iteration. This was particularly important in case of misspelling and language-specific variations. For example, the keyword "university" was represented by many language specific variations such as "universitas", "universidad" or "univ". An example for misspelling is "network *infomation* center" which occurred at least seven times in the list. Such variations were additionally included in the keyword list for each category.

Based on this extended and refined keyword list, the procedure started from the beginning and was repeated again. The complete list of the industry classes created and their respective definition are shown in Figure 7. The keywords used for each industry class are given in the Appendix.

Figure 5: Ambiguous EDGAR search result for "Sprint".

Step 3. SEC Classification. A Java program was written to download information from the SEC EDGAR system. The organization name was used to search for the company. For 40,384 search requests, 2,732 entries could be found in the EDGAR system. However, sometimes the same company has several names, which resulted in more than one outcome for the organization name. An example of such an ambiguity can be found in Figure 5. Because there was no reliable way to uniquely identify the correct entry in such a case automatically, all entries with multiple search results were eliminated which led to 1,706 remaining search results. Furthermore some companies had no SIC code and were eliminated as well. This resulted in 469 ASs that could additionally be classified into industry groups.

5 RESULTS

5.1 Keyword Classification

Applying the method described above and using the keywords shown in the Appendix to classify the 40,384 ASs, resulted in 22,786 or 56.42 % of classified ASs. The industry class distribution based on keyword classification only can be seen in Figure 6. According to this data, most frequently the organizations that own ASs belong to the industry classes *Education & Research*, *Finance & Insurance*, *ISPs & Networks*, and *Telephone & Communications*. This class distribution seems to be intuitive: ISPs, telephone and IT companies as well as universities have more incentives to register an AS than for example a travel agency because ASs classified into these categories are often related to communications, but often also represent major institutions that have a high tendency to own an AS simply because of their size.

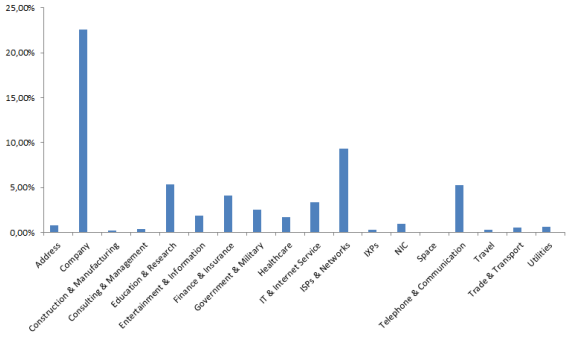

Figure 6: Industry class percentages based on keyword classification.

Of all clusters (apart from the generic *Company* cluster) *ISPs & Networks* is the category that is most frequent and accounts in case of both classification approaches for around ten percent of all the classified entities. This is an expected result since ASs pertain to the communications business, and offering Internet access is a key factor in this particular business area. The categories *Education & Research* and *Telephone & Communication* occupy the second and third positions with 5.32 % and 5.24 % respectively. Therefore, even today, companies in the area of Internet and information technology are still overrepresented because of their particular Internet affinity. A bit more unexpected, however, is that financial institutions seem well represented since the fourth position is taken by the *Finance & Insurance* cluster. All other categories are smaller with percentage values between 2.57 % (*Government & Military*) and 0.07 % (*Space*).

However, a limitation of our results is that the general *Company* cluster still encompasses 22.54 % of the classified ASs. This fact and the remaining number of unclassified ASs indicate that there is still a potential for improvement regarding the classification process. Yet it is questionable whether it is possible to reach much better results with semi-automatic classification approaches because of the presence of non self-explanatory organization names and acronyms such as NGM or EDP. Not only is it difficult to classify those simply based on keywords, it is also challenging to specify what kind of organization they represent without further manual and individual investigations.

5.2 SEC Classification

The industry-class frequency of organizations based on an alternative classification that is solely based on SEC data is shown in Figure 8. The industry classes *Construction & Manufacturing* as well as

Cluster	Definition	Result	Ratio
Address	ASs where no underlying organization is specified but an address, where the AS itself, the underlying organization or its managerial unit is located.	326	0.81 %
Company	Collecting bucket for those ASs which are hard to categorize based on their organization name but at least can be identified as a company.	9,104	22.54 %
Construction & Manufacturing	Mostly building firms and manufacturers are part of this class.	97	0.24 %
Consulting & Management	ASs related to advising and leading of a company.	139	0.34 %
Education & Research	ASs related to learning und gaining of new insights such as schools, universities, research facilities and networks as well as laboratories.	2,150	5.32 %
Entertainment & Information	ASs which are for example related to television, gaming, radio or publishing.	761	1.88 %
Finance & Insurance	This class consists mainly of banks and insurance firms.	1,664	4.12 %
Government & Military	ASs with an authority and military character as well as areal territories such as cities and states are relevant for this class.	1,039	2.57 %
Healthcare	Next to hospital (district) related entities, this class contains pharmaceutical firms.	695	1.72 %
IT & Internet Service	ASs that are affiliated with online as well as offline IT services and computer products. In general, this includes those firms which provide a service or product that is based on the Internet or IT, but which do not offer Internet access.	1,371	3.39 %
ISPs & Networks	Collects ASs of those organizations which offer Internet access or provide the necessary infrastructure.	3,775	9.35 %
IXPs	This class collects all ASs which function as exchange point in the Internet.	134	0.33 %
NIC	Contains those ASs which are "responsible for managing and allocating Internet resources" [6].	390	0.97 %
Space	Contains ASs of the area of astronautics.	30	0.07 %
Telephone & Communication	This class contains (mobile) telephone providers and sellers as well as general communication-based organizations.	2,118	5.24 %
Travel	Contains all ASs that are related to mobility and travel, such as airports, train stations, hotels and travel agencies.	103	0.26 %
Trade & Transport	Collects ASs of the area of wholesale and logistics including apparel and food.	232	0.57 %
Utilities	Organizations which provide electric power, water as well as other basic materials; also services such as waste disposal, coal and mining belong to this class.	256	0.63 %

Figure 7: Industry classes with their definitions and percentages based on the keyword classification.

Consulting & Management have the most organizations with ASs. The classes *Government & Military*, *ISPs & Networks*, *IT & Internet Service*, *IXPs*, *NIC*, as well as *Space* have no ASs at all. However, because not all companies are listed with the SEC and in particular governmental institutions and privately held companies are not registered, the lack of representation of these classes is inherent.

With the help of the SEC data it was possible to classify additional 116 ASs, which could not be classified via keywords only (Figure 9). Furthermore, the industry classes of 206 ASs could be specified more precisely which had previously been assigned to the *Company* class (Figure 10). Most newly specified classifications were assigned to the *Construction & Manufacturing* as well as the *Consulting & Management* industry classes. Their prevalence reflects the results of the SEC classification.

By combining both classification approaches we were therefore able to classify 22,892 of the 40,384 AS of the preprocessed list.

This accounts for 56.69 % of all considered ASs that could be assigned to an industry class. The final result is shown in Figure 11.

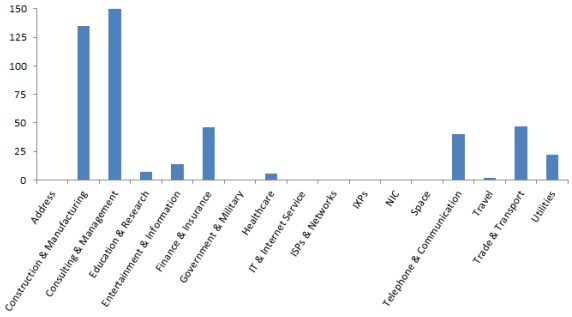

Figure 8: Industry class sizes based on SEC classification.

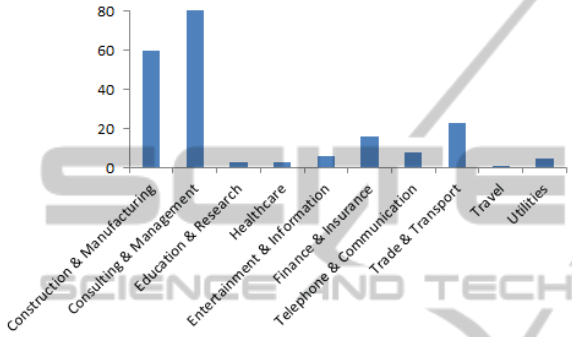

Figure 9: New classifications.

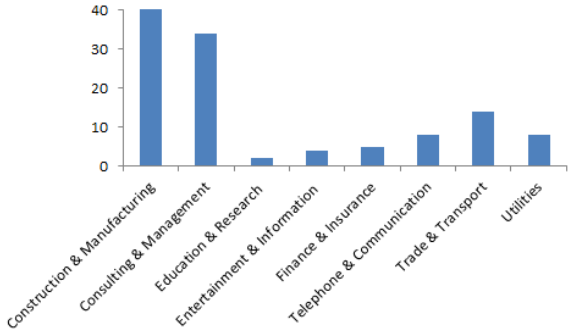

Figure 10: More precise SEC classification.

Cluster	Result	Ratio
Address	326	0.81%
Company	8,898	22.03%
Construction & Manufacturing	198	0.49%
Consulting & Management	254	0.63%
Education & Research	2,155	5.34%
Entertainment & Information	771	1.91%
Finance & Insurance	1,685	4.17%
Government & Military	1,039	2.57%
Healthcare	698	1.73%
IT & Internet Service	1,371	3.39%
ISPs & Networks	3,775	9.35%
IXPs	134	0.33%
NIC	390	0.97%
Space	30	0.07%
Telephone & Communication	2,134	5.28%
Travel	104	0.26%
Trade & Transport	2,546	6.30%
Utilities	269	0.67%

Figure 11: Final classification using both keyword and SEC data (Note: It was possible to categorize an AS into more than one industry class.).

6 CONCLUSIONS

This paper proposed a classification approach for categorizing ASs into detailed industry classes in order to better understand the economic background of the Internet structure. The industry classes are inspired by the NAICS (2013), which had the effect that an unprecedented level of detail regarding the industry classes for classification could be achieved. Data was mainly obtained from the CAIDA AS Rank project as well as from SEC.

The classification of ASs into industry classes based on their underlying organization revealed an on-going strong dominance of telecommunication and IT-related firms in the current Internet as well as of large institutions such as banks and universities.

It was possible to classify 56.69 % of all ASs (after preprocessing). Nevertheless, the amount of unclassified ASs indicates that there is room for improvement regarding the categorization process. A refined and extended keyword selection process could provide better results. Nevertheless, since there is a non-negligible amount of ASs having organizational specifications that are not self-explanatory or acronyms, this would involve a difficult challenge.

Some of our further explorative attempts to find new ways for AS classification with the help of clustering algorithms had limited success so far. However, another possible route could be to apply methods from Natural Language Processing (NLP) to the AS data and also for analysing search results from the Web for acronyms or other challenging organization names.

Moreover, customers of the various ISPs cannot be captured by the current method. It is often the case that large Internet providers also represent smaller customers who are not registered in the organizational information of the ASs. Here, studying the level of IP addresses could provide further insights but will also involve complex challenges.

Various other classification approaches might be feasible. In future work we will try to find other valuable classification systems aiming to take an even closer look at the composition of the Internet. Furthermore, we will use our classification results to further investigate the important players of the Internet backbone as well as to assess their economic interests and risks, at individual as well as global scales. Moreover, we aim to derive implications for Internet reliability and control assessments as well as for security and privacy analyses.

ACKNOWLEDGEMENTS

The authors thank Sebastian Dombrowski for his programming work during parts of this research.

REFERENCES

CAIDA, 2011. About CAIDA. http://www.caida.org/home/about/. (Access Dec, 2013).

CAIDA, 2012. AS Rank Project. http://as-rank.caida.org/?mode0=as-dump-info. (Access Dec, 2013).

CareerOneStop U.S. Department of Labor, Employment and Training Administration, 2013. NAICS-SIC Cross/Reference. http://www.acinet.org/industry/Ind_Sic.aspx?id=&nodeid=1. (Access Dec, 2013).

Chang, H., Jamin, S., Mao, Z., Willinger, W., 2005. An Empirical Approach to Modeling Inter-AS Traffic Matrices. In *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*.

Dhamdhere, A., Dovrolis, C., 2011. Twelve Years in the Evolution of the Internet Ecosystem. *IEEE/ACM Transactions on Networking* 19(5):1420–1433.

Dimitropoulos, X., Krioukov, D., Riley, G., claffy, k, 2005. Classifying the Types of Autonomous Systems in the Internet. *SIGCOMM 2005 Poster*, Philadelphia, Pennsylvania.

Dimitropoulos, X., Krioukov, D., Riley, G., claffy, k. 2006. Revealing the Autonomous System Taxonomy: The Machine Learning Approach. In *Passive and Active Network Measurement Workshop (PAM)*, Adelaide, Australia.

Freund, Y., Schapire, R.E., 1997. A Decision Theoretic Generalization of Online Learning and an Application to Boosting. *Journal of Computer and System Sciences* 55(1):119–139.

International Telecommunication Union, 2011. Internet Users per 100 Inhabitants 2001-2011. http://www.itu.int/ITU-D/ict/statistics/material/excel/2011/Internet_users_01-11.xls. (Access Dec, 2013).

International Telecommunication Union, 2013. Key ICT Indicators for Developed and Developing Countries and the World. http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/ITU_Key_2005-2013_ICT_data.xls. (Access Dec, 2013).

NAICS, 2013. North American Industry Classification System. http://www.census.gov/eos/www/naics/. (Access Dec, 2013).

Potaroo, 2012. 32-bit AS Number Report. http://www.potaroo.net/tools/asn32/. (Access Dec, 2013).

RouteViews, 2013. University of Oregon Route Views Project. http://www.routeviews.org/. (Access Dec, 2013).

Réseaux IP Européens, 2011. AS Number Assignment Policies. http://www.ripe.net/ripe/docs/ripe-525. (Access Dec, 2013).

RIPE NCC, 2013. Réseaux IP Européens Network Coordination Centre. http://www.ripe.net/. (Access Dec, 2013).

U.S. Department of State, 2013. World Population Day 2013. http://www.state.gov/secretary/remarks/2013/07/211828.htm. (Access Dec, 2013).

United Nations Population Fund (UNFPA), 2001. The State of World Population 2001 – Demographic, Social and Economic Indicators. https://www.unfpa.org/swp/2001/english/indicators/indicators2.html. (Access Dec, 2013).

United States Securities and Exchange Commission (SEC), 2013. The Investor's Advocate: How the SEC Protects Investors, Maintains Market Integrity, and Facilitates Capital Formation. http://www.sec.gov/about/whatwedo.shtml. (Access Dec, 2013).

APPENDIX

Cluster	Keywords
Address	avenue, building, flat, floor, strasse, gpo box, handelsweg, mcpo box, no., po box, road, street, suite(s), tower
Company	associates, agency, a\.s\., bv, b\.v\., cjsc, co, co kg, companies, company, coporation, corp, corporation, d.o.o., de c.v., enterprise(s), gmbh, inc, incorporated, l\.l\.c, limited, llc, llp, lp, l\.p\., ltd(a), organization, s.a. de c.v., s.p.a., s.r.l., sp. z o.o., srl, s\.a\., s\.l\., sa, sas, sl, trust, z\.s\.p\.o, zspo
Construction & Manufacturing	architect(s), builders, building company, building society, construcoes, construction, constructora, constructors, electronics, machine, manufacturer(s), manufacturing, producers
Consulting & Management	beratung, business solutions, capgemini, consultance, consultancy, consultant(s), consulting, ernst & young, management company, pricewaterhousecoopers
Education & Research	.*universitaet, academic, academisch, colegio, college(s), desire2learn, ecole, education(al), fachhochschule, forschungsgemeinschaft, forschungsgesellschaft, fraunhofer, institute, instituto, knowledge network, laboratories, laboratory, labs, learning, mitre, physics, polytechnic, recherché, research, school(s), science(s), supercomputer, supercomputing, univ, universidad, universidade, universitaet, universitaria, universitas, universite, universiteit, universitesi, universitet, universiti, universities, university, univerzitet
Entertainment & Information	advertising, bbc, bertelsman, book(s), broadcasting, entertainment, football, fun, game, gaming, library, magazine(s), mcgraw-hill, media, marketing, medien, multimedia, news, newspaper(s), printing, publications, publishing, radio, reuters, television, times, tv, weather, zdf
Finance & Insurance	allianz, american express, asset management, assurance, banca, banco, bank, banka, banque, blue shield, capital, credit, finance, e*trade, financial, goldman, guggenheim, hsbc, insurance, investment, leasing, payment, real estate, reinsurance, rental, societe generale, stock exchange, stonepeak, visa
Government & Military	administration of, agency, air force, army, authority, board of, bureau of, city of(fice), committee, commonwealth of, congress(ional), council, county of(fice), department of, dept, district of(fice), dod, embassy, federated states, gov, government, house of, iles de, military, ministry, nato, navy, northrop grumman, parliament(ary), province of, senate, state of, united nations, united states postal service, US geological survey
Healthcare	bayer, blood, dental, drug(s), drugstore, elektromedizinische, emergency, health(care), hospital(s), johnson & Johnson, klinikum, medical, medicine, medizinische, merck, novartis, pfizer, pfizerswitzerland, pharma(cy), pharmaceuticals, pharmafarm, propharma, social security, transplant
IT & Internet Service	akamai, apple inc, computer hardware, computer products, computer science, computer service(s), computer software, computer solutions, computer systems, content provider, content service provider, content solution(s), data center(s), data corporation, data processing, data service(s), data solution(s), data systems, dell, fujitsu, general electric, google, hewlett-packard, host, hosting, ibm, information systems, information technology, internet service(s), internet systems consortium, it services, microsoft, neterra, network service(s), network systems, oracle, othello, samsung, sap, schuberg philis, siemens, sony, sungard availability, thinktech, verisign, web service(s), yahoo
ISPs & Networks	aol, arcor, at&t, backbone, broadband, bt italia, cable network(s), cogent, comcast, connection(s), esnet, exatel, fibernet, freenet, gts, iletisim hizmetleri, internet access, internet provider, internet service provider(s), internet solution(s), isp, lattelekom-apollo, level 3, linxtelecom, netassist, netcologne, network access, network provider, network service(s), network solution(s), networks, ntt america, optical network, prometey, qwest communication(s), reseau national, reseau regional, retn, road runner, rostelecom, singtel optus, smartcity, sprint, surfnet, swisscom, t-2, telecom, telekom, telia latvija, teo, time warner cable, towerstream, transit, true internet, uzbektelecom, verizon, versatel, vimpelcom, west call, wireless
IXPs	exchange point, internet exchange, internet exchange point, ix, ixp(s), link, open exchange, peering exchange
NIC	afnic, american registry, apnic, arin, east-ukrainian, internic, network information center, network infomation center, network information centre, nic, ripe ncc,
Telephone & Communication	alcatel, bell canada, communication(s), e-plus, elisa, ericsson, lambdarail, mobile, motorola, nokia, o2, phone, radiotelephone, rockefeller group, singtel optus, telecommunication(s), telecomunicaciones, telefonica, telekommunikation, telekomunikacije, telekomunikacja, telekomunikasi, telecomunicazioni, telephone(s), telianet, turkcell, vodafone
Trade & Transport	amazon, apparel, clothing, coca-cola, fedex, food(s), logistic(s), logisticare, retail(ers), shaya magazacilik, shipping, shoe(s), supply, trade, trading, transport(ation), wal-mart, wholsesale
Travel	air canada, airline(s), airport, bahn(hof), boing, flughafen, klm, lufthansa, hotel(s), railway, reisebuero, resort, travel, vacation
Space	aeronautics, aerospace, astronomy, nasa, space administration, space agency, space research, space telescope
Utilities	bp, coal, electric power, electricity, energy, farmer, farms, fiber, gas, mine, mining, offshore, petroleum, utilities, utility, waste, water