Discovery and Mapping of the Dutch National Critical IP Infrastructure

Research Project 2

Fahimeh Alizadeh, Razvan C. Oprea

Supervised by:
Benno Overeinder, NLnet Labs
Marco Davids, SIDN

August 10, 2013

Abstract

The research project entails the mapping and subsequent analysis of the AS-level interconnections between the organisations active as the Dutch critical infrastructure. The discovery of the organisations' AS representation utilises exclusively public sources of information and uses a two-pronged approach. First, a bottom-up process is used—starting from the complete list of Dutch ASNs we select the ones corresponding to Dutch critical organisations. Second, a top-down approach—starting from a list of representative Dutch critical infrastructure organisations we find their AS-level network representation. We use the UCLA's AS topology map files based on BGP routing tables and updates for determining the AS-level interconnections between critical sector organisations. We then implement a visualisation method for constructing the network graphs for each critical sector and analyse their interconnections. We conclude that the Dutch critical infrastructure organisations are well interconnected but rely a lot on foreign entities for IP transit and even for carrying potentially sensitive information via web and email services.

Contents

List of Tables

List of Figures

1 Introduction

Critical infrastructure has always been one of the primary topics of interest for governments, because it plays the crucial role in either national or international wide economy and welfare level in society. The best way for governments to protect and extend the critical infrastructure is first to have a complete understanding of the existing essential infrastructure, and second to analyse its resilience and determine future improvements.

The protection of critical infrastructure, basically aimed at country safeguarding and reducing weaknesses and risks of critical infrastructure, must include the dependencies and connections between the various critical sectors. However, as shown below, the protection of the critical infrastructure has not been formalised until relatively recently.

Starting with the 1996 Executive Order 13010 - Critical Infrastructure Protection issued by President William J. Clinton, creating a president's Commission on Critical Infrastructure Protection and defining eight critical sectors [1], governments formalised the protection plans of critical infrastructure sectors.

In the Netherlands, up to the beginning of this century, the critical infrastructure protection was not a concern at the highest level [21]. At the end of the 90's, an initiative to protect the national critical infrastructure started within the Dutch Ministry of Defence, each infrastructure sector being responsible for its own critical infrastructures, with no central state organisation to coordinate national critical infrastructure protection measures.

In 2002, the Dutch government started the Critical Infrastructure Protection (CIP) project "Bescherming Vitale Infrastructuur" with the objective "The development of an integrated set of measures to protect the infrastructure of government and industry" (which included Information and Communication Technology) [24].

More recently, a policy letter Protecting Critical Infrastructure (2005) and the third progress letter on National Security (2010) includes an analysis of the quality of protection of critical infrastructure [9]. Critical infrastructure is defined as comprising "the business enterprises and public bodies that provide the goods and services essential for the day-to-day lives of most people in the Netherlands". Critical infrastructure is divided into 12 critical sectors, with a total of 31 essential goods and services. Table 1 represents the Dutch critical sectors as they are defined by the Dutch government.

Since the protection of the national critical infrastructure is formalised in government policies, some legitimate questions arise relating to the network interconnections between these organisations: how well interconnected are they and how reliable are these network links? Can we graph and visually analyse these network interconnections? What conclusions can we draw? The answers to these questions are a sound starting point for organising more suitable defence strategies and taking weakness points into account to reduce the risks and threats.

The main goals of this research are thus the discovery, visualisation and analysis of the AS-level network relationships between the Dutch organisations

Category	Critical Sector
energy	electricity, natural gas, oil
telecommunications and ICT	land-line and mobile telephony, radio, broadcasting, Internet
drinking water	the water supply
food	the food supply (including in supermarkets), food safety
health	emergency and hospital care, medicines, vaccines
financial sector	payments and money transfers by public bodies
surface water management	water quality and quantity (control and management)
public order and safety	public order and safety
legal order	the courts and prisons, law enforcement
public administration	diplomacy, public information, the armed forces, decision-making
transport	Amsterdam Schiphol Airport, the port of Rotterdam, highways, waterways, railways
the chemical and nuclear industries	the transport, the storage, the production, the processing of materials

Table 1: The Dutch critical infrastructure sectors

that are part of these critical sectors.

1.1 Research question

The main goal of this research project can be summarised in the following question:

Can we discover and map the Internet entities corresponding to the Dutch national critical infrastructure with a sufficient degree of confidence?

We also set additional goals defined in the following two sub-questions:

- ***What are the authoritative source(s) of information when discovering the Internet presence of an organization?***
- ***What can be said of the resilience of the Dutch critical infrastructure IP-level network?***

1.2 Outline

Section 2 reviews the related research that has been done on this topic and how it differs from this project's approach. Section 3 and 4 introduce the methodology used to discover the relevant AS numbers and to determine the relationships between them. It details on different approaches used in the research to obtain results with a higher degree of confidence. The visualisation techniques used for creating AS-level graphs of different critical sectors is described in section 5. AS-level graphs created from the aggregated collected information are then

analysed in section 6. The conclusions are drawn based on the research findings in section 7.

2 Previous Work

Critical infrastructure protection and Internet interconnection resilience have been studied previously separately:

1. During the past two decades the interest in the protection of the national critical infrastructure started growing in the United States and Europe and was formalised in policy documents, as shown in Section 1.

2. There are also a number of research papers focused on studying the Internet interconnection resilience, such as the April 2011 study "Inter-X: Resilience of the Internet Interconnection Ecosystem" by the European Network and Information Security Agency (ENISA) [23]. The study analysed many potential threats to the Internet interconnections, ranging from the sustainability of the economic model to cascading technical failures, equipment reliability, coordinated BGP attacks, and human infrastructure needed to maintain it during a pandemic flu.

3. Association française pour le nommage Internet en coopération (AFNIC) started publishing since 2011 a report on the resilience of the French Internet. Their latest report [22] is more focused on the technical attack vectors, such as IP prefix hijackings, RPKI, and DNSSEC.

However, there are few previous research papers we found relevant to the subject we focused on — the analysis and visualisation of the critical infrastructure at a network level. Or at least, there are few *public* papers on the subject.

The exception we found is the 2012 paper "Exposing a Nation-Centric View on the German Internet — A Change in Perspective on the AS Level" [25].

In their research they have started from the list of IP prefixes allocated to organisations registered in Germany and then they used RIPE Database, Team Cymru and RIPE RIS to find the originating AS numbers. The AS interconnections were then discovered using BGP dumps. This 2012 paper was the first study we found that analysed the AS-level relationship between entities part of a national critical infrastructure.

In comparison with this paper, our research is, in some ways, more limited in scope by the public nature of our sources of information — we do not know for instance, which IP blocks some organisations use internally. On the other hand, we extend the scope of the research by including an analysis of many of the foreign ASes which act as proxies for web and mail services provided by critical infrastructure Dutch companies.

Additionally, we have detailed in our report a method for building a graph of the ASN relationships and visually analysing the interdependencies.

3 AS Discovery

We have no idea on the organisations' physical connections to the Internet, but since we are interested in the logical IP topology, we decided to work strictly at the AS level.

The ideal resource for our research would be getting a complete and accurate mapping of critical sector infrastructure organisations with their respective AS numbers. While it is expected that all critical infrastructure-related organisations have some Internet presence, one cannot assume that all of them will be using their own AS number. Some might have their online presence (such as website and email servers) hosted by organisations which access the Internet via a certain AS number (we refer to such AS as a"proxy" AS) — and some of these proxy ASes can be located outside of the Netherlands.

There was a need for mapping AS numbers which encompass all these possibilities — and we used two methods for achieving this. Once the AS numbers are determined, the interconnections between them are discovered and the results can be analysed.

Below is a description of the steps related to the methodology we employed.

1. *Bottom-up AS discovery approach*

   The first method for AS discovery is a "bottom-up" approach: starting from the list of all AS numbers allocated to organisations registered in the Netherlands, we select those which are active in one of the 12 critical infrastructure sectors. This approach is intended to identify all critical infrastructure Dutch organisations that have their own AS number ("native" AS).

2. *Top-down AS discovery approach*

   The second approach for AS discovery is a "top-down" approach: we search for representative (prominent) instances of critical sector organisations in the Netherlands and then find out whether they have a native AS (in which case we should see the organisation also listed in the bottom-up approach) or they use a proxy AS.

   For the latter case we look up the IP address of their A, AAAA, and MX DNS records and then identify the larger IP prefix they are part of and the AS number announcing this IP prefix — this is the proxy AS.

3. *Analysis and visualization*

   In this third step we combine the results of the previous two AS discovery approaches and look for all the relationships between the AS numbers. We analyse the resulting data, we visualise it, and draw conclusions.

3.1 Bottom-up AS Discovery Approach

As mentioned in Section 2, in an earlier research [25], AS numbers were determined (using a variety of tools such as the RIPE WHOIS database, Team Cymru, etc.) from IP prefixes.

The ideal scenario is finding an authoritative source of information that would offer us the AS information without going through the complex loop of IP prefix-to-AS mapping. The RIPE NCC publishes a daily updated stats file on their public FTP site [12] which contains, among other information, all the AS numbers allocated to all the organisations in the RIPE service region [13]. Since the RIPE NCC is the sole organisation that hands out AS numbers in the service region the Netherlands is part of, this makes the stats file an authoritative source of information.

Figure 1 shows an excerpt from the RIPE NCC's stats file.

```
ripencc|*|asn|*|28184|summary
ripencc|EU|asn|1196|1|19930901|allocated
ripencc|IE|asn|1197|1|20101118|allocated
ripencc|EU|asn|1198|1|19930901|allocated
ripencc|EU|asn|1199|1|19930901|allocated
ripencc|NL|asn|1200|1|19930901|allocated
ripencc|EU|asn|1203|1|19930901|allocated
ripencc|AT|asn|1205|1|19930901|allocated
ripencc|IE|asn|1213|1|19920617|allocated
ripencc|EU|asn|1234|1|19930901|allocated
ripencc|EU|asn|1235|1|19930901|allocated
ripencc|EU|asn|1241|1|19930901|allocated
ripencc||asn|48198|1||reserved
ripencc||asn|48204|1||reserved
ripencc||asn|48253|1||reserved
ripencc||asn|12410|1||available
ripencc||asn|15449|1||available
ripencc||asn|15907|1||available
ripencc||asn|16078|1||available
```

Figure 1: The RIPE NCC's stats file (excerpt).

It can be observed in the stats file that the AS numbers are allocated to organisations registered in different countries in the RIPE NCC service area. Out of these, our interest is in the organisations registered in the Netherlands (NL country code) and European Union (EU). While the NL allocations are easy to extract from the file, the EU ones could potentially correspond to organisations registered in any of the 28 European Union countries.

To select the Dutch organisations from the EU-labelled ones (around 1,400 entries), we wrote and used a script which queried every entry against the RIPE WHOIS Database [15] and looked for country codes in the "Description" and "Address" fields. Any discrepancies or corner cases are then manually reviewed.

We thus obtained a (complete, as far as we can tell) list of 727 AS numbers allocated to organisations in the Netherlands. The next step is determining which of these organisations are part of the Dutch critical infrastructure. As showed in Section 1, the Dutch government decided which economic sectors and goods and services represent the national critical infrastructure. Unfortunately, there isn't any way of determining the organisations-to-critical sector mapping. First, we needed to determine every organisation's domain name and website

and manually review it to classify it.

We searched each one of the 727 AS numbers using Google, RIPEstat and the Dutch Chamber of Commerce (Kamer van Koophandel [19]), labelled the organisations based on the 12 sectors and 31 goods and services, according to the classification done by the Dutch government, from A (Energy sector) to L (Chemical and Nuclear industries).

We observed that some organisations had several AS numbers registered on their name or their subsidiaries (we found subsidiaries using the KvK data) — for instance, SURFnet has over 60 AS numbers, while Koninklijke KPN N.V. (KPN) has almost 20.

We also noticed that the number of AS numbers in the Internet sector is disproportionately high compared with the total number of "Dutch" ASNs. Unlike non-ICT sectors where relatively big organisations do not have their own AS number, in ICT even the smallest entities sometimes have a native AS. We realised that we cannot claim that every single ICT-related company is part of the national critical infrastructure only because it is classified as an ICT sector company. We therefore decided to exclude companies which do not have their own infrastructure: purely VoIP providers, hosting companies, IT service providers — including Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) providers, IT consultancy, education and research institutions. We included only Internet Service providers (ISPs), Data Centres and Internet Exchange Points (even if their AS number is not related to the core service they provide).

We thus ended up with a list of 335 AS numbers related to critical infrastructure sector organisations, out of which 265 AS numbers relate to the ICT sector (Internet in particular), as shown in Figure 2.

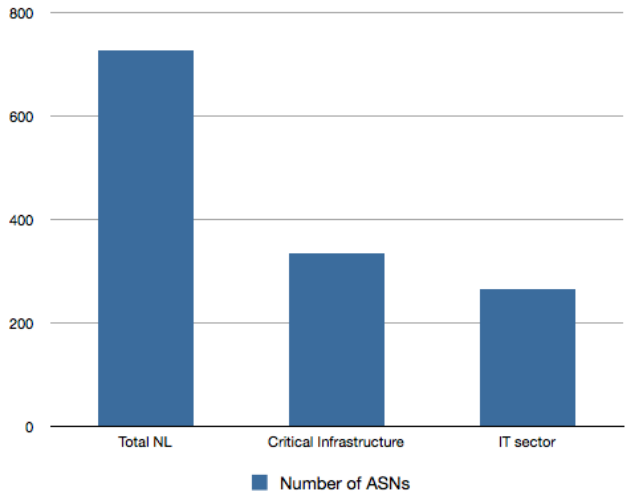


Figure 2: Distribution of Dutch AS numbers

.

3.2 Limitations of the Bottom-up AS Discovery Approach

This approach has some limitations as well:

1. We do not know if all the AS numbers an organisation uses relate to critical infrastructure-related operations. However, if for instance, we classify KPN as a critical infrastructure organisation, we label all of its AS numbers (and of its subsidiaries) as such. Taking into account the fact that KPN can use and reuse its AS numbers as it sees fit and no outside party can control this, there is not a better alternative to marking all of its discovered AS numbers with the same label.

2. We have limited information regarding companies ownership, since we rely on two main public sources of information. On one hand we are are restricted to the information provided in the RIPE WHOIS database, which is updated by the users and where the information regarding mergers, acquisitions and bankruptcies is updated with significant delays. On the other hand, we rely on the KvK data, which is precise, however many times difficult to understand which organisation is the owning which one. We did the best we could to understand the relationships between organisations, an effort which will prove useful when we later gather all the AS relationships related to several AS numbers in one "virtual ASN".

3.3 Top-down AS Discovery Approach

As described in the previous section, the bottom-up approach, while discovering all the Dutch AS numbers, resulted in a very distorted map of the critical infrastructure. For instance, there was notany organisation discovered in the "Surface Water" critical infrastructure sector.

We therefore started the discovery process from the opposite direction — starting with a list of companies we knew they were part of the critical infrastructure. The selection came from multiple sources — through common knowledge, Wikipedia, Google, etc., trying to have samples from each sector. The search yielded a total number of 147 organisations.

After the completion of this step, the KvK website was consulted again to identify the corresponding domain name for each company in turn.

Since we had decided we will only use publicly available information, we are not privy to what IP address space critical sector organisations are using internally. Applying for access to such privileged information, besides the small chances of approval, would have delayed the project too much and it would have restricted the distribution of its results.

An obvious source of information, given a domain name, is represented by its DNS records. Of the available ones, the interesting ones are those that indicate information flow.

For instance, if we consider that a fictive energy company has its A and MX records pointing to an IP address (part of a IP prefix) allocated to a foreign organisation, then we can safely assume that the website and (external) mail

servers (MTA) of this particular energy company are hosted by the foreign entity. Even if the hosted website is just a front-end interface, while the databases and storage stores are not externalised (a very likely scenario), there is data flow from inside the company to the foreign entity (web site updates) and likely from outside to the inside (for example web forms, access statistics, etc.). In case of mail, it is even more obvious — while internal company mail could be restricted to the internal network, every incoming mail from a different domain would have to pass through the mail servers indicated by the MX records (the foreign entity).

This is not the case with other types of DNS records — there is no traffic exchanged between a company's internal network and the machines identified by the NS records, for instance, except for DNS updates. Same could be said of other types of DNS records, such as SRV and TXT.

We concluded therefore that the A, AAAA, and MX records are the ones we should focus on. We queried for the AAAA record to cover for the corner case in which an organisation's web site is reachable over IPv4 on a certain web host and via IPv6 on another. We used the Internet Systems Consortium "dig" utility for gathering the IP addresses corresponding to these DNS records, and then RIPEstat to find the prefix they are part of and the originating AS (the proxy AS).

3.4 Limitations of the Top-down AS Discovery Approach

Besides the already mentioned observation that we only work with publicly accessible data, there are some other limitations of this approach:

1. We do not see any backup and private links the organisations might have. The only visible information, except for the discovery of a native AS (which would have been seen already from the bottom-up approach), is represented by the A, AAAA, and MX DNS records.

2. Complete mapping of critical sector industries requires specialised knowledge. In the critical sector "Food", for instance, one could consider almost all organisations in the food supply chain as being part of the critical infrastructure. And there are many, many different organisations — breeding companies, animal food manufacturers, pig farms, slaughterhouses, food transport companies, butcheries, supermarkets. A complete list of national critical infrastructure organisations could only be obtained using specialised knowledge (potentially via industry associations) or privileged information (via for instance the Chamber of Commerce).

4 ASNs Interconnection Discovery

The combined results of the two approaches for AS discovery, bottom-up and top-down give us a "master" list of AS numbers. This list represents all the AS numbers allocated to organisations registered in the Netherlands (via the bottom-up approach) and a representative selection — almost 150 hand-picked organisations (via the top-down approach).

Our next goal is to see what kind of relationships (interconnections or links) exist between all these AS numbers allocated to Dutch critical infrastructure organisations. At an AS level, this kind of information can be obtained via the BGP tables. But a more accurate view on these links can be obtained by using multiple observation points — and there are few methods for obtaining multi-views on the BGP links:

- *Routing Information Service (RIS)* [14] is a RIPE NCC project that collects and stores Internet routing data from 14 locations around the globe. It is using servers acting as software routers called Remote Route Collectors (RRC) that collects default free BGP routing information from many participating peers.
- *Route Views* is a project of the University of Oregon [20]). In a fashion similar to RIPE RIS, it operates data collectors throughout the world. Collectors serve as real-time operational tools, as well as data sources for the RouteViews central data archive. It was originally created to help ISPs debug and optimise their networks, nowadays it is also used in academic research.
- *Route Servers* — some organisations still install publicly-accessible routers designed to peer with as many parties as possible for the benefit of the networking community. Network operators can access these routers using Telnet or SSH and run a variety of commands (ping, traceroute) or look at the routing tables to help them troubleshoot network problems or get a better understanding on AS topology. Routeserver.org [11] and BGP4.net [8] provide lists of public route servers across the globe.
- *Looking Glasses* — some operators allow public access to some routers via web CGI's which provide a view into the BGP table of that particular router. BG4.net [3] has a rather large list of looking glasses servers worldwide. Similarly to the observation made with the route servers, there are services which aggregate the information provided by several looking glasses into one master BGP routing table.

Among the services that aggregate the data from RIPE RIS, Route Views, route servers and looking glasses three stand out: The University of California, Los Angeles (UCLA) Internet Research Lab, The Cooperative Association for Internet Data Analysis (CAIDA) and University of Washington's iPlane.

We looked the type of data provided by each one of these sources:

- University of Washington's iPlane [7] is more geared towards measuring links performance (latency, bandwidth, capacity, loss rates) than a complete AS-level topology. Nevertheless, an AS map is being constructed and

updated daily by using traceroute from different vantage points (PlanetLab nodes and traceroute servers).

- CAIDA collects several different types of data from various vantage points and makes this data available to the research community by publishing it on their website [18] We were particularly interested in the AS relationships data [2], with the latest version of the file dated June 2012. CAIDA was also able to specify the confidence they have in the type of AS relationship data: "99.1% of the links we infer to be provider-customer are validated to be provider-customer, with the correct direction. 94.7% of the links we infer to be peer-peer are validated to be peer-peer."

- UCLA's Internet Research Lab [6] released daily AS topology map files based on BGP routing tables and updates (RIPE RIS, Route Views), Route Servers and Looking Glasses. UCLA released node files (lists of AS numbers), links files (relationships between AS numbers) but also very importantly, map files containing direct links between any two AS numbers and the type of relationships - peer-to-peer (P2P), provider-to-customer (P2C) and customer-to-provider (C2P). At the time of our research, the newest data file was from 7 April 2013.

  However, upon closer examination, we observed that the fields in the map files showing the type of relationships has not been updated since November 2012. Our supervisor, Benno Overeinder contacted the UCLA Internet AS-level topology maintainer, Yu Zhang and the data started being re-published, in a slightly changed format starting with June 2013 and available at a new URL [5]

Figure 3 shows an excerpt of the map file.

```
1103    21345   p2p
3917    5400    c2p
3917    8167    c2p
3917    17379   c2p
8455    12945   p2c
8455    15426   p2c
8455    15879   p2c
8455    16237   p2c
8455    20562   c2p
```

Figure 3: UCLA IRL map file (excerpt).

Upon considering our options we decided to drop the plan to use University of Washington's iPlane data sets. While their data was the newest, we missed the AS relationship types between AS numbers and their set was limited to traceroute discoveries from PlanetLab servers (so no RIPE RIS or Route Views)

We had to decide then between CAIDA and UCLA data and we chose for the latter, for the simple fact that the data was newer - at the time we decided, we still believed we will work with the April 2013 data, not with the November 2012 we ended up with. There was no sense in using both datasets (UCLA and CAIDA's) because in case of discrepancies we wouldn't have known whether this is due to CAIDA and UCLA seeing different routes or simply due to the

fact that there are routes which were present when CAIDA undertook the measurements and were then dropped.

With a fairly comprehensive list of AS numbers, relationships and relationship types, we needed to find a way to map them.

The first approach, obviously, was to select all the links in the UCLA map file in which both nodes are from the master file of all critical infrastructure AS numbers (Dutch and foreign, discovered via the bottom-up plus the top down approaches). The result would be a list containing exclusively the links between the ASes corresponding to the Dutch critical infrastructure.

Unfortunately, the resulting graph had so many disconnected nodes that we realised quickly that we had to consider adding more intermediary or transit nodes to the map file — this would enable us to see whether there are concentration points, where many links are terminated for instance.

Obviously the goal is to build the minimum graph that connects all the critical infrastructure AS numbers, native and proxy.

Which AS numbers to include to show relevant links? One initial idea we had was to include non-Dutch Tier 1's and Tier 2's - or, at least the transit providers mostly recognised as being such — and we created a list of 21 such organisations. We added these 21 AS numbers to our list and the number of links increased dramatically — actually the number of resulting links was so big (the 21 Tier 1 and Tier 2 ASNs doubled the amount of links for a bigger critical sector, like Energy) that we were now experiencing the opposite scenario than before: because of the excessive amount of links between these major providers, it was hard to extract the useful information from so much non-relevant data in the graphs.

It was clear by now that we needed a middle path: enough links to show the full connection mesh between our critical infrastructure sectors and without any extraneous links which would otherwise "swamp" the graphs and obscure any useful information.

The idea we finally implemented involved using the UCLA map file for gathering AS relationship types: for every unique AS number in our list we added its transit provider to the list and then we built the full mesh of relationships between the new master list: AS numbers plus their providers.

Figure 4 is a schematic representation of this idea applied on a fictive sector.

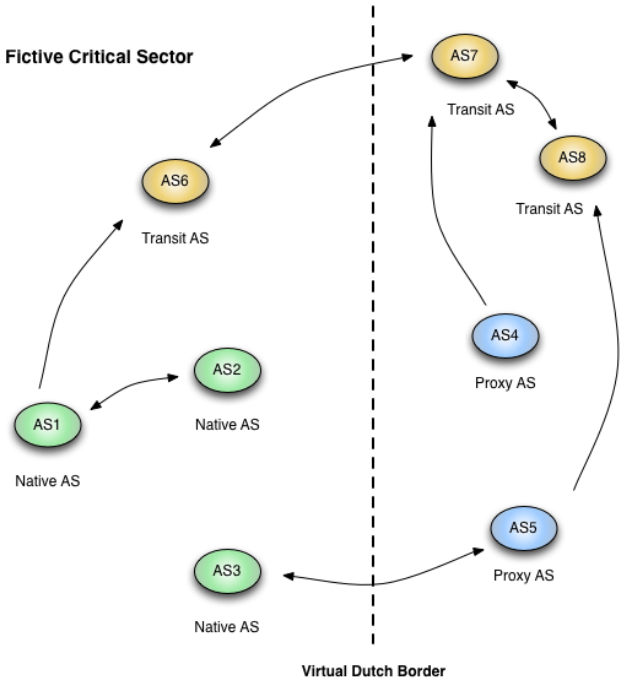


Figure 4: Constructing the ASN relationships graph (fictive critical sector).

5 ASNs Network Visualization

Not only gathering the proper information from various data sources and aggregating them optimally is important, but also the way the collected information will be used and presented. Visualisation helps troubleshooting process operate more suitable in large scale environments. In case of a small number of ASNs, tracing all the relations between them would not be an issue but managing all the AS numbers registered in a country manually is close to impossible. The best approach would be either splitting the input dataset to smaller data sets and visualising each one of them, or, if the input dataset is clear enough, the visualisation process can be done without further division.

In this project we tried to use interactive visualisation methods to keep the level of clarity and informativeness at the highest level. The implemented program provides a set of graphs for each critical sector to maintain simplicity and also one graph for the entire critical infrastructure. These set of graphs can be considered an additional input for further investigations and analysis in terms of resilience and determining the connectivity degree.

5.1 Technical Considerations

There are number of technologies used in each development stage of the visualisation method, such as bash scripting, python, CSS and JavaScript. The following describes the options taken into account to reach the best result graph.

5.1.1 D3.js Library

D3.js [4] stands for *Data-Driven Documents*, a JavaScript library that uses well-structured JSON files as input. The fields and values of input JSON files can be modified based on application requirements, but a specific format has to be kept. D3.js library is used for creating AS-level diagrams, representing each AS number as a node and each inter-AS relation as a link. Nodes will be distinct with the associated labels and colours.

The input dataset has to be formatted in a way that is retrievable for D3.js library. There are two input files used for each AS-level graph: one containing the nodes and the other containing the links. The following code block shows an example of two nodes, with one of them having its own AS number and the other one being the customer of another company (proxy AS).

```
[
 {
"as":"286",
"company":"Brabant Water",
"sector": "C1",
"input":["proxy",{"record":"A","company":"Koninklijke KPN N.V.","
    country":"nl"}]
},
{
"as":"56517",
"company":"Vitens N.V.",
"sector": "C1",
"input":["native"]
}
]
```

We tried to put as much as information in each node. Each node contains the AS number, the company name (corresponding to that AS number), the critical sector code (represented with a unique combination of digits and letters) and some further information on the AS number origin. As we described before (Section 3), there are two possibilities for each company:

- If the company has its own AS number, it is specified in the nodes list as "native".
- If the company does not have its own AS number and it is dependent on another company having its own AS number, it is specified in the nodes list as "proxy". In the node specifications there is also more information related to the origin company and the country that origin company is registered in.

For the links input file, the case is much simpler. For each link only two fields are filled: "source" and "target". Both fields are representative of AS numbers that have a direct link. All the links are defined as bidirectional, so the pair of AS numbers in links file will appear at most once. Also it does not make any difference if the AS numbers in the "source" and "target" fields are swapped. Since AS numbers are unique, they will be treated as pointers to unique companies.

```
[
    {"source":286,"target":56517}
]
```

Positioning AS nodes on the screen in the most efficient way is one of the main goals of the visualisation technique. The first step is to visually differentiate the Dutch AS numbers from the foreign ASNs either with colours, borders or grouping nodes together.

Several options can be taken into account for positioning nodes (AS numbers and companies) on the screen based on the requirements. **Force Layout** is the placement method used in this project. D3.js library uses some positioning functions by default, which are meant to set "top" and "left" CSS fields of each node based on the calculations of best fitted positions. D3.js also provides parameters that can be set to fine-tune the placement of nodes. Force layout is a simulation of force in Physics science. Each link is the representation of a coil and there is a force between each two nodes. There are a number of parameters, which can be set to refine the final position of the nodes on the screen. The graph can be moved easily and the nodes can be dragged, but it is still necessary to add zoom-in or Fisheye effects to see more details of nodes more clearly.

5.1.2 Sigma.js Library

Sigma.js [17] is an open source library, which can be extended easily in terms of additional plug-ins. Comparing it with the D3.js library, we chose to use the Sigma.js mainly because it included many functions which were missing in the in D3.js library, such as zoom-in.

In order to parse JSON input files, jQuery functionality is involved - the same JSON files from D3.js are used in this method as well.

In contrast to D3.js, Sigma.js does not provide positioning layout. Positioning function is needed to display the nodes and links on the screen in the most efficient way. The following list shows the requirements that have to be met in order to make the graph comprehensive enough:

- The isolated nodes have to be found easily in the graph. The difficulty with graphs with high number of nodes is that there must be a dynamic positioning function which selects the best fitted position for each node on the screen. If the nodes are put on top of each other or too close together, then the links can not be differentiated very well and it is not clear whether a node is part of a particular link or not. With zooming into the graph links originating from each node will be shown clearly but it is more preferable if isolated nodes and nodes with higher degree (number of links the node is part of) are distinguishable without further actions and at first sight.

  Nodes are spread on the screen on circles with the same centre and different radiuses. Nodes with a higher degree are set on the inner circles and nodes with lower degrees are set on the outer circles. This ensures that isolated nodes appear separately and away from dense areas.

- The nodes have to be distributed on the screen equally. We first tried putting them on random positions and the resulting graph was not clear.

We then decided it is more preferable to locate nodes with higher degree close together, since this way the majority of links are shown shorter in the graph and results in a more apprehensible graph.

The nodes are managed within an array sorted by descending number of links. Therefore, the higher index in array represents a node with lower degree. The following code block shows the formula used for each node to calculate its position:

```
y = sin(a) * r
x = cos(a) * r
```

As shown in figure 5, **r** is the radius, which is calculated by using the index number of the node in the array. So the node with highest degree will have the radius 0 and will be located on the centre of the circle. All other nodes with lower degrees will be arranged around it with different radiuses. Also **a** is the radian angle, calculated by using the index number of the node in the array. Although the same value could be used for each node's angle, to end up with better graph, it is better to spread nodes with different angles and use empty spaces on the screen. $\mathbf{(x_0,y_0)}$ is the center of all the circles, which in this case is (0,0).

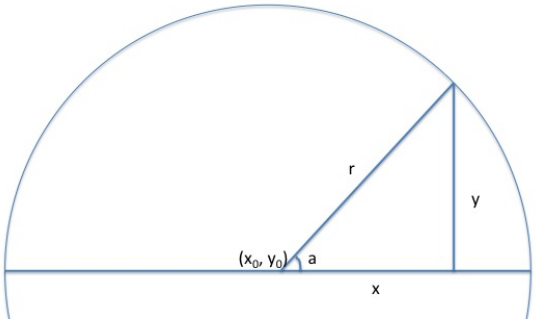
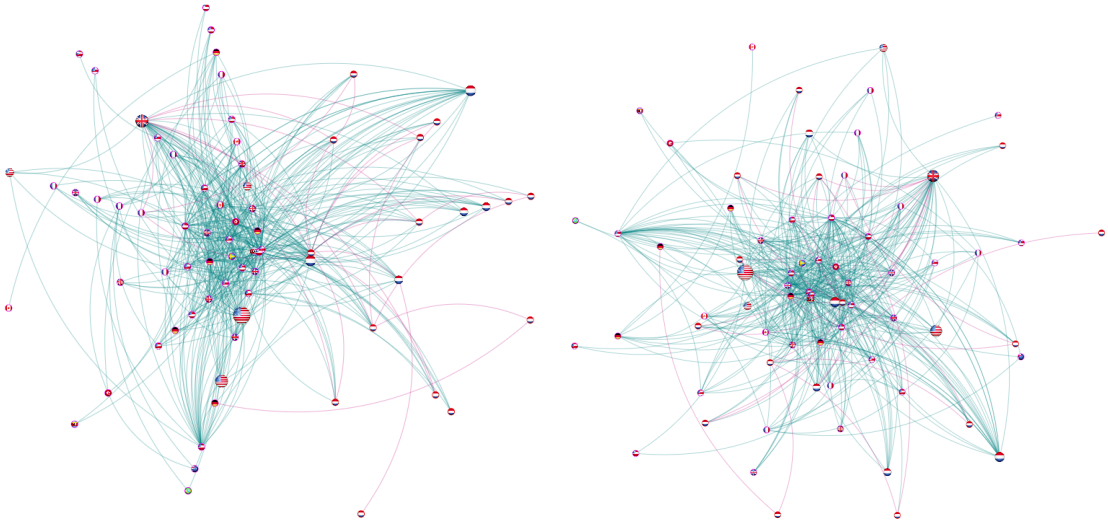


Figure 5: (x,y) selection for each node

- Due to the focus of this research and to make the analysis the network graph easier, there must be a clear visual distinction between the Dutch ASNs and the foreign ones. We tried various solutions and in the end we settled for using different colours for nodes, displaying flags of the country that the corresponding ASN is allocated to and having two reference centres on the screen. In sector 6 it is shown that foreign ASNs are more connected in general than Dutch ones. Having all the links related to foreign ASNs next to the links of Dutch ASNs makes the analysis harder for the viewer. The difference this latter option makes is visible in Figure 6, where the left graph shows a clearer picture of how links are distributed between foreign and Dutch ASNs.

  Separately positioning of ASN numbers has to be implemented with a slightly different method. Figure 7 shows that how two different initial centres of circles are considered. In this case the radian angle used in positioning function can not be in this range: $[0,2\pi]$ because of overlapping



(a) (b)

Figure 6: Two versions of node positioning (left one optimised) **The chemical and nuclear industries** critical sector

of left-located circles and right-located ones. Semicircles can have the proper target with $[\pi/2,3\pi/2]$ for left-located ones and $[3\pi/2,5\pi/2]$ for the right-located ones.

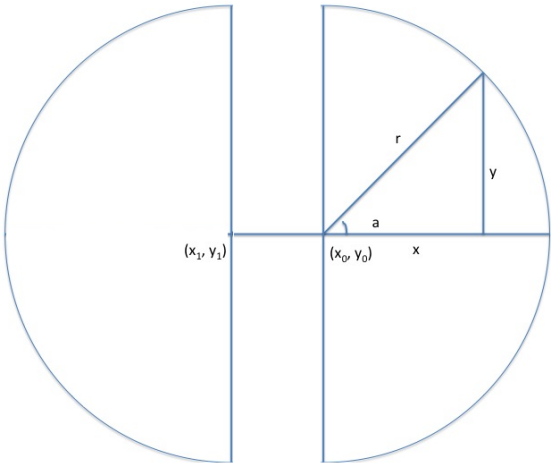


Figure 7: (x,y) selection for each node (with foreign and Dutch ASNs separately positioned)

- We have shown in the graphs the number of companies dependent on each ASN. ASNs and their inter-connections are shown as nodes and links. The colour gradient is another parameter that can be taken into account to show the density of each node in terms of number of companies dependent on it but we use it to distinguish between national ASNs and foreign ones.

In order to add this additional information in the graph, we implemented variable size method for each node; the node with larger size is a proxy (or transit provider) for a higher number of companies.

Some other improvements we applied to the visualisation method to increase the clarity of the graphs are using different colours for links between Dutch nodes and foreign nodes and displaying name of the owner company of ASNs as labels.

6 Data Analysis

In this section we find out how well the companies active in critical sectors are connected and if there are any interesting findings in the ASN mapping of Dutch critical sectors. The network graphs are used as input data source for further research. The network graphs of each critical sector is commented below. In the list below, we used the term "subsector" instead of the official terminology of "goods" and "services", mostly to increase the clarity of the explanations.

1. **Energy critical sector** We investigated the energy critical infrastructure sector which has 3 sub-sectors: electricity, gas and oil. Figure 8 shows the ASN network graph for this sector when only direct links between each two ASNs are taken into account. It is obvious that the graph is too disconnected for drawing any conclusions. Although every two ASNs will be connected in a graph containing all ASNs, the number of isolated nodes in this graph is too high. The links are distributed almost equally in two sides: 44% for the foreign ASNs and 56% for the Dutch ASNs.

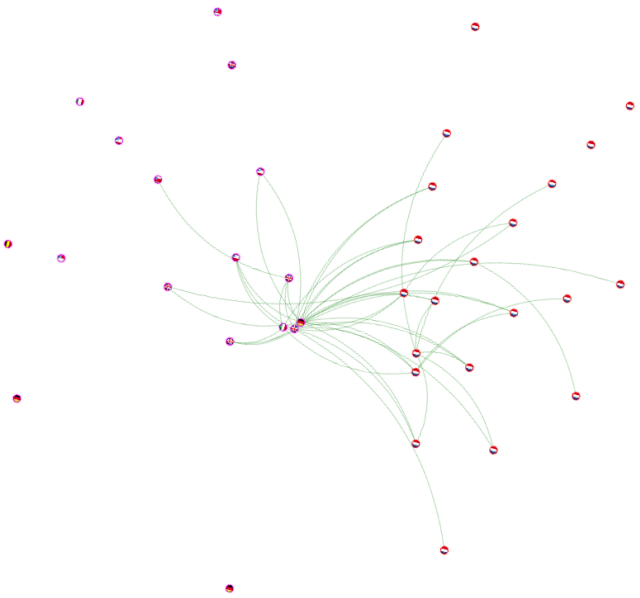


Figure 8: Energy critical sector without providers

As shown in Section 4, with so many isolated nodes in the graph we had to include the immediate transit providers. In most cases a transit provider offers the connectivity for an AS, and the provider usually has the higher degree of connectivity. Using the maps file provided by UCLA (Section 4), we intersect the Energy ASNs from the first column with the rows

containing **customer-to-provider** in their third column and the graph is redrawn. Figure 9 shows new network graph with transit providers included. The graph is more connected than before and the distribution of links is totally changed: 69% for foreign part and 31% for Dutch part.

Although it is expected for each node to have at least one link (it will be the connection of the node to its provider), we still can find one isolated ASN in Dutch part. According to RIPE Stats [16], ASN 61013, which belongs to Alliander N.V. company has never announced any IP prefix. Alliander N.V. is one of the largest companies in maintenance, expansion and adaptation of the gas and electricity network in the Netherlands. The web server and mail server of this company are hosted by British Telecommunications plc (ASN 5400).

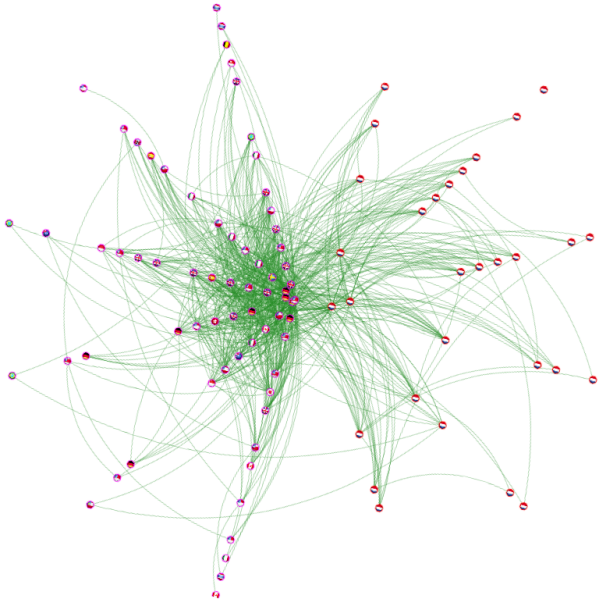


Figure 9: Energy critical sector with providers

For this sector, KPN B.V. is the Dutch ASN with highest degree and also with the highest number of companies depending on it. In the foreign part, Level 3 Communications Inc. and Swisscom appear in the graph with the highest degree, while Microsoft Corp. and British Telecommunications plc have the highest number of companies depending on them.

2. **Telecommunications and ICT** For this critical sector, 4 subsectors are defined by Dutch government: mobile telephony and land-line, broadcast, radio and Internet. As explained in Section 3, "Internet" is quite a general term for all the services that can be provided by companies active in this area. We had to list the services in order and choose only few with highest

priority. Finally, 368 Dutch ASNs were kept and 106 foreign ASNs were discovered as proxy and transit ASNs.

In the centre of the graph, some well-known companies with the highest degree of connectivity appear such as: KPN B.V., SURFnet B.V., BIT BV, Eweka Internet Services B.V. and Open Peering Initiative. SURFnet B.V. also is hosting provider for highest number of companies in this sector.

It is likely that when a company is near to the centre of the graph, its node representation is also larger in size (higher degree and higher number of companies depending on it) both in Dutch and foreign parts. In this sector there are many ASNs with only one or two links in foreign part. This can be interpreted as following:

- It may be possible that in one industry, companies choice for their web server and mail server hosting providers are the same. In this case they will be shown larger in size in the network graph.
- In telecommunications and ICT critical sector, we could hardly see any difference between the proxy and transit providers. Because we did not consider all the links between all the ASNs, we can not argue about connectivity of nodes with one or two links but we can talk about connectivity of one node to other nodes active in this specific sector (native, proxy and transit ASNs). Having a high number of nodes shown as proxy ASNs with only one or two links reveals that the choices of companies in this critical sector for hosting services is quite diverse and they did not limit their options to only the most frequently used by other related companies.

Besides, there are many isolated nodes in Dutch part and with further investigations in RIPE Stats, we discovered that those ASNs do not announce any IP prefixes.

3. **Drinking water** This critical sector is mainly about water supply industry. The network graph for this sector is connected and it does not contain any isolated node. ADIX B.V. and KPN B.V. are the ASNs with highest number of companies depending on them and ASN (56517), allocated to Vitens N.V. is the only native ASN in this sector.

4. **Food** This critical sector contains the food supply and the food safety as its subsectors. Food supply includes a wide range of activities starting from farms and ending to supermarkets. There is no isolated node in its network graph and there are two native ASNs in this sector: The Greenery B.V. and Multi Corporation B.V. which are active in food supply subsector.

   Figure 10 reviews the network graph of food critical sector. As it is clear, the distribution of links between Dutch ASNs and foreign ones is not uniform: 78% for foreign part and 22% for Dutch part.

5. **Health** Emergency and hospital care, medicine and vaccines are subsectors of the Health critical sector. Mediq N.V., E-Zorg B.V. and Medisch Centrum Leeuwarden own the only native ASNs in this sector, which came

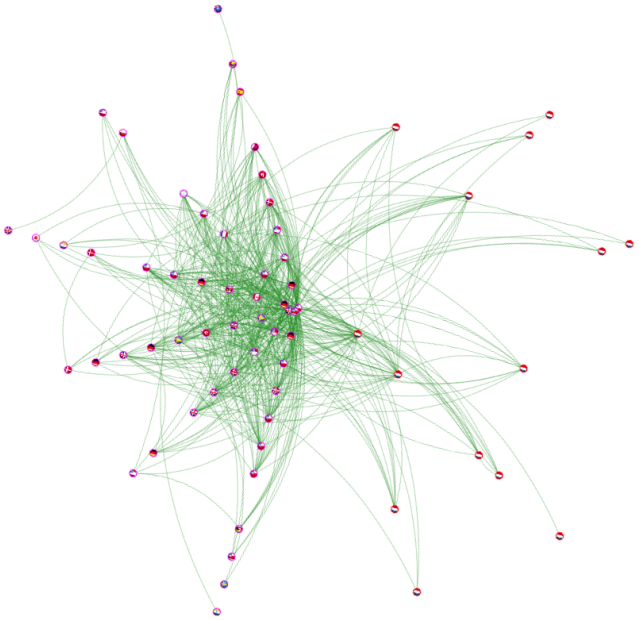


Figure 10: Food critical sector with providers

up through the bottom-up approach. The graph is connected and each node has a degree more than 1 except two native nodes (Mediq N.V. and E-Zorg B.V.), where their connectivity is offered by their providers.

It is more likely that the nodes in centre of the graph (corresponding to a higher degree and a higher number of links) are active in ICT and telecommunications, even for critical infrastructure sectors not related to ICT and telecommunications, For example we have seen KPN B.V. mostly as one of top nodes in terms of degree of the node. Also in this sector companies SURFnet B.V., KPN B.V., and Atrato IP Networks appear as having highest degree.

6. **Financial sector** This sector has a generic subsector of payments and money transfers by public bodies. For this sector, we safely excluded entities such as Nationale Nederlanden (pensions, investments and mortgages) and GSX (hosted services including payment gateways) because they do not make "payments and money transfers by public bodies". Public body is an organisation whose work is part of the process of government but not part of the government itself [10]. Besides we included organisations like banks, which receive the salary of government employees. Due to the vague government definition of what exactly critical infrastructure sectors, goods and services encompass, there is a certain level of freedom in choosing representative entities for our list of critical infrastructure or-

ganisations. Interestingly 64% of nodes in this graph are native although the graph has two isolated native nodes, corresponding to organisations which, even though have ASN numbers assigned to them, have decided to host their services elsewhere.

7. **Surface water management** The subsector defined here by the government are the surface water quality and quantity. In this sector, there is only one native ASN: 42894 (Ministerie van Verkeer en Waterstaat/Rijkswaterstaat). The graph is connected and Microsoft Corp. has the highest number of companies dependent on.

8. **Public order and safety** All city halls can be considered, which are essentially active in this sector. Although about 50% of all nodes have their own ASN, in the graph we observe some native isolated nodes, which according to RIPEStat do not announce any IP prefixes.

9. **Legal order** Legal order critical sector is defined with two subsectors: the courts and prisons and law enforcement. For the latter, Dutch Police and Nationale Politie, which are local divisions of Dutch Police were selected. The only native ASN in this graph belongs to "Voorziening tot samenwerking Politie Nederland". The graph is well connected in Dutch part; every Dutch ASN (proxy or native) has a minimum degree of 3, which means except its direct provider, it has one other BGP neighbour.

   The Courts of Justice do not have their own ASNs but they have outsourced their mail server and web server inside the country (ASP4ALL Hosting for the web server and Tele2 for the mail server).

10. **Public administration** Diplomacy, Public information, the armed forces and decision-making are subsectors introduced for this critical sector. All the ministries may be considered in decision-making subsector. Also most of them can be taken into account as a provider of public information. Ministerie van Algemene Zaken, which applies coordination of government policy and communications on the Royal House and Ministerie van Buitenlandse Zaken are selected for diplomacy subsector. 26% of our nodes in this sector are representative of native ASNs but still there are ministries dependent on foreign hosting companies. Dutch ministries are accessible through two umbrella domains:

    - **government.nl**: Mail services are outsourced to MessageLabs (UK and US) but the web server is hosted in Prolocation B.V. in the Netherlands.
    - **rijksoverheid.nl**: Mail server and web server are outsourced but are hosted inside the country (Prolocation B.V. for web server, KPN B.V. for mail server).

11. **Transport** In this critical sector there are four subsectors defined such as: Amsterdam Schiphol Airport, the port of Rotterdam, waterways, highways and railways. Schiphol airport and the port of Rotterdam are separated because of their importance and vital role in transport industry.

    British Telecommunications plc is the most frequently used foreign hosting company by the organisations active in this sector. One isolated native

Sector	Dutch Provider	Foreign Provider	Top 1 Foreign Provider
energy	56%	44%	Microsoft Corp., US
telecommunications and ICT	96%	4%	Websense Hosted, UK
drinking water	61%	39%	MessageLabs Inc., US
food	63%	37%	No clear leading provider
health	75%	25%	MessageLabs Ltd., UK
financial sector	81%	9%	MessageLabs Ltd., UK
surface water management	57%	43%	Microsoft Corp., US
public order and safety	92%	8%	ClaraNet Ltd., UK
legal order	67%	33%	BT plc, UK
public administration	74%	26%	MessageLabs Ltd., UK
transport	61%	39%	BT plc, UK
the chemical and nuclear industries	36%	64%	MessageLabs Inc., US

Table 2: Distribution of Mail providers in each sector

node from Zeeland Seaports N.V. appears in the graph. Also other native ASNs in the graph apparently do not have high number of links. In the centre of circle we can see KPN B.V. with the highest degree and with the highest number of companies depending on it.

12. **The chemical and nuclear industries** Based on the steps needed for a chemical process, there are 4 subsectors introduced in this sector: the transport, the storage, the production and the processing of materials. For instance **Urenco Netherland B.V.** is working in the uranium enrichment area and providing nuclear fuel cycle. On the other hand **Covra** does the radioactive waste management in the Netherlands. It is unlikely to find a company active in all subsectors.

    The graph is connected and contains only one native ASN. Microsoft Corp., MessageLabs Inc. and MessageLabs Ltd. are the most frequently used for outsourcing purposes by companies active in this sector.

Table 2 reviews the distribution of mail providers as foreign or Dutch. As it is clear in all sectors the number of Dutch providers used is higher than foreign providers but one: The chemical and nuclear industries, which is dependent mostly on MessageLabs Inc.

The biggest mail providers are **MessageLabs (UK and US)**, **KPN B.V.**, **Microsoft Corp.**, **Tele2 Nederland B.V.** and **Ziggo**. In fact, MessageLabs, a division of Symantec Corp., is the single biggest mail provider in our list.

Another interesting observation is that sometimes related companies/industries choose the same providers:

- MessageLabs Ltd.: ABN AMRO and Triodos Bank.

- MessageLabs Ltd.: AkzoNobel and GGD.
- BT plc.: NS and ProRail.
- Microsoft Corp.: Royal Dutch Shell, Gasunie and Argos Energies.

7 Conclusions

In this research we mapped the representative Dutch critical infrastructure organisations using the two discovery methods (bottom-up and top-down). The discovered organisations were verified manually one-by-one so we have a high degree of confidence in the accuracy of the results.

However, we only worked with public sources of information and thus we did not see physical, private and back-up links (as shown in Section 3). We do not know how different organisations physically connect to the internet, neither those that use private AS numbers or private links. But we do know how the organisations listed interconnect to each other at an AS level and we were able to draw the corresponding graph and extract some interesting results.

A more comprehensive list of organisations can only be obtained with specialised information (consider the food-sector example we gave in Section 3), or with privileged access to information. Such information would allow us to know what IP address space is actually being used inside every organisation and thus find the "native" AS number, in addition to the proxy ASNs. It would also allow us to gather the full list of Dutch organisations in all the critical sectors (via for instance the Chamber of Commerce) and the top-down AS discovery approach would then produce comprehensive results, instead of representative samples.

From the research done we observed that many critical infrastructure organisations have reliable connections to the Internet (the native and proxy ASes are well interconnected), but rely a lot on foreign providers for their communication needs. If we would consider the imaginary scenario of an emergency in which critical sector organisations can only communicate using Dutch links, then around half of them (those that use foreign proxy ASes) would be cut-off from the network.

How important is however the IP network for a particular critical sector organisation, such as an energy provider? We simply cannot know this without access to specialised and privileged information. But we do know that with its A and MX records part of an IP prefix announced by a foreign AS, all web traffic and emails received by this organisation pass through foreign entities. Let us consider the hypothetical scenario of such an energy provider providing a web portal for its customers (www.myenergyprovider.nl) in which one can see his invoices, provide payment methods (bank accounts and/or credit cards), set his contact details (address, phone number, etc.). How important is, from a privacy and a confidentiality perspective at least, that such information goes first through a foreign provider whose infrastructure, privacy policies, traffic retention policies and (deep) traffic inspection procedures are unknown to us and which can change at any time, without notice to us, the end user? How valid are then the confidentiality agreements we sign with our energy provider? What happens when instead of an energy provider company we consider the case of a medical unit or a medical insurance company which allows its customers to visualise and edit their records online (medical bills, address, telephone, bank account)?

In this context we find it would be useful to start a discussion regarding the

security and privacy implications of having critical infrastructure organisations' email and websites hosted with foreign entities, especially so with those from outside the European Union (EU) since they do not not necessarily have the same laws regarding data privacy and confidentiality. Which laws apply on the "privacy sensitive" data if many companies have online presence out-sourced to foreign or even non-EU countries? This is becoming more and more important as many critical sector organisations allow their customers to manage their accounts online (Energy providers for instance) or make various requests via email.

A clear conclusion is that we do not see that national critical infrastructure organisations regard their network infrastructure and the data communication flow as being of critical importance, yet.

References

[1] Amendment to Executive Order 13010, the President's Commission on Critical Infrastructure Protection. Website. http://www.archives.gov/federal-register/executive-orders/1996.html.

[2] AS Relationships in CAIDA project. Website. http://www.caida.org/data/active/as-relationships/.

[3] BGP Looking Glasses List. Website. http://www.bgp4.as/looking-glasses.

[4] Data-Driven Documents JavaScript Library. Website. http://d3js.org/.

[5] Internet AS-level Topology Data by UCLA. Website. http://irl.cs.ucla.edu/topology/.

[6] Internet Topology Collection by UCLA. Website. http://irl.cs.ucla.edu/topology/index-old.html.

[7] iPlane: An Information Plane for Distributed Services. Website. http://iplane.cs.washington.edu/.

[8] IPv4 Route Servers List. Website. http://www.bgp4.net/rs.

[9] Protecting critical infrastructure. Website. http://www.government.nl/issues/crisis-national-security-and-terrorism/protecting-critical-infrastructure.

[10] Public body definition by MacMillan dictionary. Website. http://www.macmillandictionary.com/dictionary/british/public-body.

[11] Public Route Servers List. Website. http://routeserver.org.

[12] RIPE NCC ASN information list. Website. ftp://ftp.ripe.net/pub/stats/ripencc.

[13] RIPE NCC service region. Website. http://www.nro.net/about-the-nro/list-of-country-codes-and-rirs/list-of-country-codes-in-the-ripe-ncc-region.

[14] RIPE Routing Information Service (RIS). Website. http://www.ripe.net/data-tools/stats/ris/routing-information-service.

[15] RIPE WHOIS Database. Website. http://whois.ripe.net.

[16] RIPEStat information for AS61013. Website. https://stat.ripe.net/AS61013$#$tabId=at-a-glance.

[17] Sigma.js JavaScript Library. Website. http://sigmajs.org/.

[18] The Cooperative Association for Internet Data Analysis (CAIDA). Website. http://www.caida.org/data/overview/.

[19] The Dutch Chamber of Commerce. Website. http://www.kvk.nl.

[20] University of Oregon Route Views Project. Website. http://www.routeviews.org/.

[21] Critical Infrastructure Protection: Survey of World-Wide Activities. *Bundesamt für Sicherheit in der Informationstechnik (BSI)*, 2004.

[22] Résilience de l'Internet Français (ENISA report). Website, 2012. http://www.afnic.fr/medias/documents/Dossiers_pour_breves_et_CP/Observatoire-sur-la-resilience-Internet-en-France-2012.pdf.

[23] Hall, Chris, Ross Anderson, Richard Clayton, Evangelos Ouzounis, and Panagiotis Trimintzios. Resilience of the Internet Interconnection Ecosystem., 2011. http://www.enisa.europa.eu/activities/Resilience-and-CIIP/critical-infrastructure-and-services/inter-x/interx/report.

[24] Luiijf, Eric AM, Helen H. Burger, and Marieke HA Klaver. Critical (information) Infrastructure Protection in the Netherlands. *GI Jahrestagung (Schwerpunkt "Sicherheit-Schutz und Zuverlässigkeit")*, pages 9–19, 2003.

[25] Matthias Wählisch, Thomas C Schmidt, Markus de Brün and Thomas Häberlen. Exposing a nation-centric view on the German internet - a change in perspective on the AS level. *Proc. of the 13th Passive and Active Measurement Conference (PAM)*, 7192:200–210, 2012.