The state of .nl

A portrait of the Netherlands' national internet domain, based on DNS measurements

2021

Contents

Foreword

Foreword

Our dependence on the internet continues to grow. However, if we are to confidently go on making ever more use of the internet, we must be able to trust in its safety. At SIDN Labs, we therefore investigate ways of enhancing the security of the internet infrastructure on which our society depends. Much of our research is based on 'big data' from the Domain Name System (DNS), such as the billions of DNS queries that SIDN processes every day for .nl, data gathered from the thousands of RIPE ATLAS sensors distributed around the internet, and the results of regular crawls of .nl's 6.2 million domain names.

'The state of .nl' describes the developments we have observed in such data. We present statistical information based on our measurements and explain the rolling statistics published on stats.sidnlabs.nl. As well as discussing the numbers themselves, we place our findings in their broader social context.

Here are a few examples of the observations reported in 'The state of .nl':

- Tension on the housing market is reflected in domain name registration data (subsection 4.2).
- Pandemic-related press conferences held by the Dutch government and large companies have a direct effect on domain name registrations (subsection 4.3).
- An increasingly large proportion of internet services are delivered by an increasingly small number of providers, not all of them US 'big tech' companies (section 6).
- The use of new technologies to improve security and stability is increasing, but the picture is mixed (sections 7 and 8).

We hope that you find this report as interesting to read as we found it to write. If you have any questions or feedback, we'd love to hear from you. Simply drop a line to sidnlabs@sidn.nl.

The SIDN Labs team.

01

Introduction

The internet is now an integral part of everyday life for individuals and organisations alike. We use it for staying in touch with family and friends, for work, for interacting with government agencies, for our hobbies and for countless other things. Since the coronavirus pandemic began, our dependence on the internet has only increased. We've been 'Zooming' with friends and 'Teamsing' with colleagues, while online shopping has become the norm for many.

01

Introduction

Base of the iceberg

Although we are so dependent on the internet, few people stop to consider how the **internet's infrastructure** works. That infrastructure is the foundation beneath all the technical systems (routers, switches, DNS servers, etc) that enable internet-connected devices around the world to communicate with each other. It is the invisible base of the technological iceberg whose tip most people think of as being the internet. Our reliance on it is rarely appreciated until something goes wrong.

Large-scale measurement campaigns

SIDN Labs is the research arm of SIDN, operator of the .nl domain. Labs' mission is to maximise the reliability of the internet infrastructure on which our society depends. To that end, our activities include a lot of technical 'data-driven' research into the **security and stability** of the internet infrastructure, particularly in relation to the Netherlands and the .nl domain. Much of our research is based on **bulk measurements** involving, for example, the billions of DNS queries that SIDN processes every day, data gathered from the thousands of RIPE ATLAS sensors distributed around the internet, and the results of daily crawls of .nl's 6.2 million domain names. Some of the measurements are published on a fully automated, rolling basis on our interactive statistics site, stats.sidnlabs.nl.

Social developments

In producing 'The state of .nl', our aim was to describe the developments we have been able to observe in our data. We were motivated by the **awareness** that our measurements sometimes reflect the effects of **social developments**. For instance, we were able to discern the effects of the coronavirus pandemic in our data as early as March 2020. We can also see how the internet infrastructure is evolving to meet the ever-changing demands of society. One example of that is the increasing use of security protocols, such as DNSSEC and routing security extensions, which increase the security of the internet infrastructure, so that people and organisations have the confidence to accept still greater reliance on it.

Underlying datasets

For 'The state of .nl', we have drawn upon a number of specific measurements performed over the last few years, and upon our analyses of the resulting datasets. Our primary data source is the **Domain Name System (DNS)**, the worldwide mechanism that translates domain names into IP addresses. The datasets we used are: **ENTRADA** (passive DNS measurements, grows by 2.7 billion queries per day), **DMAP** (website crawls, 50 million measurements per month) and **OpenINTEL** (active DNS measurements, 4 billion new datapoints per day). ENTRADA and DMAP are our own datasets, while OpenINTEL is a shared dataset with the University of Twente, SURF and NLnet Labs.

Domain Name System

Before getting down to the numbers and charts themselves, we present an overview of the DNS in section 2. We felt it was important to begin with a look at the DNS, because it is one of the pillars of the internet infrastructure and the source of the measurements underpinning this report. In addition, section 2 briefly examines a small number of other technologies that are pertinent to the DNS and are the subject of data presented later in the report. Additional information about the datasets we have used is provided in section 2 as well.

What you can expect

In the main body of the report, you will find measured data and analyses regarding:

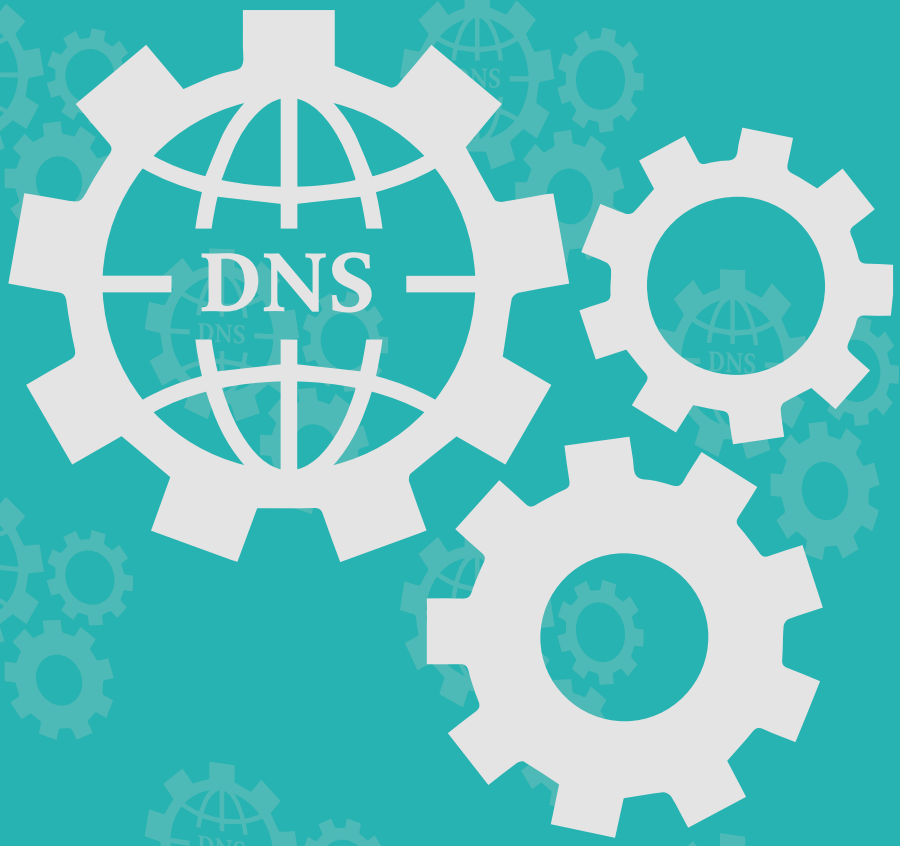
- .nl domain names, e.g. the number of active .nl domain names per registration year and the length of domain names (section 3)
- Trending words used in domain names in 2020 and 2021, e.g. coronavirus-related terms that quickly started appearing in newly registered domain names (section 4)
- DNS resolvers (the servers that retrieve DNS information for users), e.g. the international distribution of resolvers (section 5)
- Centralisation of the internet, e.g. the volume of DNS traffic received from big public DNS resolvers such as Google Public DNS (section 6)
- DNSSEC, e.g. implementation levels in different sectors and the adoption of new DNSSEC algorithms in the .nl domain (section 7)
- IPv6, e.g. the number of DNS resolvers asking for the IPv6 addresses of .nl domain names (section 8)

The methodologies used for our measurements and analyses are described in a separate subsection at the end of each section of the report. Some of the charts presented have previously appeared on stats.sidnlabs.nl, but have been included in this report because we felt further explanation would be useful.

02

Our data source: the Domain Name System

'The state of .nl' is based on measurements and analyses of various components of the Domain Name System (DNS), the global system that translates domain names into IP addresses within milliseconds and forms part of the internet's core infrastructure. The DNS works 'under the hood', meaning that it's effectively invisible to most users. Nevertheless, it's involved in almost every internet transaction, such as visiting a website or sending an e-mail.



02

Our data source: the Domain Name System

A brief explanation of how the DNS works is presented below to facilitate understanding of the charts in the rest of the report. Additional information about the datasets used for 'The state of .nl' is also provided.

A short SIDN video explaining how the DNS works is available at www.supersokken.nl.

2.1 Domain names

A domain name is a memorable 'shorthand' for an IP address (see subsection 2.2). It's divided into a number of 'labels'. For example, 'www.sidnlabs.nl' is made up of three labels: 'www', 'sidnlabs' and 'nl'. In the context of the DNS, the term 'domain name' covers all such names, regardless of how many labels they consist of. So 'login.sso.ict.east.example.nl' is a domain name, and so is 'nl'.

Second-level domain names

In everyday use, 'domain name' usually means the name of a second-level domain – one with two labels, like example.nl. In other words, the kind of domain name you register with SIDN or another registry. That's how we use the term 'domain name' in this report: to mean a second-level domain name. That isn't explicitly clarified every time the term appears in the body of the report.

Domain names under .nl contain between two and sixty-three characters (not counting '.nl'). They can be made up of letters, numbers and/or hyphens in any combination, as long as they don't start or end with a hyphen or include two consecutive hyphens.

2.2 IP addresses

Every computer connected to the internet has a unique IP address for communicating with other computers. Internet users aren't generally aware of their IP addresses, because they normally use domain names instead.

There are two 'flavours' of IP address: IPv4 addresses and IPv6 addresses. An IPv4 address will look something like `203.0.113.80`. IPv4 is widely used, but is the old version of the Internet Protocol. IPv6 addresses are formatted differently. A typical example is `2001:db8::7974:80`.

Almost all IPv4 addresses are now in use, or reserved for future use. However, IPv6 provides scope for creating many, many more IP addresses. Because IPv4 and IPv6 use totally different address formats, they're mutually incompatible. So someone who only has an IPv4 address can't communicate with someone who only has an IPv6 address. Because not everyone can get an IPv4 address anymore, it's very important that everyone is able to use IPv6.

2.3 From domain names to IP addresses

The job of the DNS is to translate a domain name, such as example.nl, into an IP address. That saves you the trouble of trying to remember a long, complicated number and lets you use a memorable name instead.

In other words, the DNS is like a 'phone book' for internet addresses. However, that's not all the DNS does. It can also provide information about certain security protocols. For example, it can tell you which e-mail servers are allowed to send mail from a given domain. That makes mail traffic more secure by enabling the detection of spoofed messages.

Using the DNS involves the interaction of three components: applications (e.g. web browsers and apps) that need information from the domain name system, DNS 'resolvers' that look up the information for the applications, and authoritative DNS servers that provide the information. To understand how that interaction works, it's helpful to consider an example.

Example: getting the IP address for www.example.nl

Suppose that you're using a browser such as Chrome, Edge or Safari, and you want to visit www.example.nl for the first time. Your browser begins by asking your operating system (Windows, MacOS, etc) for the website's IP address, so that it can fetch the site's content. To answer the browser's question, your operating system contacts a DNS resolver.

The resolver then sets about obtaining the IP address of www.example.nl. It starts at the end of the domain name. In our example, that's 'nl'. So the resolver contacts the DNS root servers to ask whether they know where to find 'nl'. A root server responds by directing the resolver to SIDN. Next, the resolver asks SIDN's DNS server where to look for 'example.nl'. Armed with SIDN's response, the resolver approaches the DNS name server for example.nl. The resolver asks the name server for the IP address of 'www.example.nl'. And the answer comes back: `2a00:d78:0:712:94:198:159:35` for IPv6 or `94.198.159.35` for IPv4. The resolver then gives the address to your browser, so that it can fetch the website content and display it.

The DNS resolvers involved in that procedure may be operated by your internet access provider, or may be public resolvers such as Quad9 or Google Public DNS. Many internet access providers set things up so that your computer automatically uses their resolver.
E-mail programs (e.g. Outlook) use DNS resolvers as well, but they use them to look up the IP addresses of mail servers, not web servers. Once the e-mail program has the address information of its mail server(s), it can use that to send mail to the recipient's domain.

Name servers for .nl

SIDN operates the authoritative name servers for the .nl zone. The zone contains second-level domain names (e.g. example.nl), directions to the level below (e.g. www.example.nl) for use by resolvers, plus cryptographic material for DNSSEC (see subsection 2.4). The servers for .nl are sited at dozens of locations around the world and use a technology called 'anycast' to maximise their availability. With anycast, different resolvers reach different name servers, depending on where the resolvers are located on the network.

In our role as operator of the .nl domain, we have access to the messages that resolvers send to the .nl name servers asking for the IP addresses of .nl domain names. According to data collected on 6 September 2021, SIDN's authoritative servers process roughly 2.8 billion DNS queries a day.

However, we don't get to see every single query that's made regarding a .nl domain name. That's because DNS resolvers 'cache' the responses they get from name servers. In other words, they save them for a while, to give to other users who want the same address. That enables them to process the huge volumes of queries they get as quickly as possible, because they don't have to contact an authoritative server about each and every query. As a result, the DNS traffic to our servers is just a sample of the total, but it does come from all around internet.

The traffic between computers and websites never comes our way: the information we provide enables them to talk to each other directly without our involvement.

2.4 DNSSEC and DANE

The DNS has been around for thirty-three years and wasn't designed with security in mind. Sadly, that means it's possible to maliciously interfere with the DNS. For example, responses can be manipulated by 'injecting' false information into a resolver's cache. We call that a 'DNS cache poisoning' attack because the attacker

'poisons' the resolver's memory ('cache') by planting a forged address. The resolver then directs internet users to the 'planted' IP address, which might be, say, a malicious website mocked up to look like the one the user wants to visit. What's more, new ways of maliciously interfering with the DNS are found on a regular basis.

DNSSEC

Fortunately, we can detect attacks like that by using the DNS Security Extensions (DNSSEC). With DNSSEC, a domain name's registrant can attach a cryptographic signature to the DNS information linked to the name, such as the web server's IP address. A resolver that wants that information can then check ('validate') the signature when they get a response from an authoritative server. If the signature is correct, the resolver knows that the response can be trusted and the IP address can safely be passed on to the user. Of course, the process is fully automated and everything happens in the blink of an eye.

DANE

As well as information about the IP addresses of web servers and the names of mail servers, almost any other kind of information can in principle be recorded in the DNS. DANE (DNS-based Authentication of Named Entities) uses that capability to enhance the protection of communication between e-mail servers. DANE's extended protection would not be possible without DNSSEC.
Ordinarily, the communication between e-mail servers is not encrypted, making e-mail traffic vulnerable to interception and manipulation. Although one mail server can ask another for their mail to go via an encrypted connection, an attacker can interfere with that process too, so that the communication ends up being unencrypted anyway. DANE ensures that the connection is encrypted, partly by using the security provided by DNSSEC.

Visit sidn.nl for more information about DNSSEC and DANE.

2.5 Datasets

Various datasets were used for the analyses presented in sections 3 to 8. The .nl zone itself is, of course, an important source of data. As well as listing all the registered domain names, it tells us whether each name is DNSSEC-enabled, for example. The other datasets we use are listed in table 2.1.

Dataset	Number of measurement points	Number of datapoints	Frequency
ENTRADA	24 .nl name servers, distributed globally	2.7 billion queries per day	Continuous
DMAP	6.2 million .nl domain names	50 million measurements per month	Monthly
OpenIntel	236 million domain names in .nl, .se, .com, .us and other TLDs	4 billion per day	Daily

Table 2.1 | Data sources used for this report.

ENTRADA

Data on resolvers comes from our ENTRADA DNS data platform. DNS queries and responses processed by the .nl name servers are saved in ENTRADA for use in research designed to support the security and stability of the internet. In order to establish whether a resolver validates DNSSEC signatures, we observe whether it actually requests DNSSEC records (e.g. signatures and public keys). The method is not 100 per cent reliable, but gives us a good general picture of resolver behaviour.

DMAP

We classify websites on the basis of content or type using data collected by DMAP: the crawler we created to automatically scan all .nl domain names once a month. Using DMAP, we can investigate various things, such as what kind of website is linked to a domain name.

OpenINTEL

Finally, we also work with data from OpenINTEL, a platform that gathers DNS information on more dan 236 million domain names, including all .nl domain names, every day. Our information about which DNSSEC algorithms are being used and how many domains are using DANE comes from OpenINTEL. OpenINTEL is a collaborative initiative by SIDN Labs, NLnet Labs, SURF and the University of Twente.

03

.nl domain names

Let us begin with an analysis of the information in the .nl zone itself: the .nl domain names that we publish via .nl's authoritative name servers on behalf of the names' registrants (see section 2). We'll look at the number of .nl domain names, their length, their age and the regions where they are registered.

03

.nl domain names

3.1 Number of registrations

On 1 January 2022, there were 6,229,639 registered .nl domain names. For a relatively small country like the Netherlands, that's a very large number: roughly one .nl domain name for every three Dutch people. In Europe, only Germany (.de) and the United Kingdom (.uk) have more domain names in their top-level domains. According to CENTR, those countries had, respectively, 17,110,294 and 11,107,255 domain names at the end of 2021. Globally, .nl is the fifth biggest country-code domain.

In the early years, .nl's administrative processes were manual. It wasn't long, however, before burgeoning demand for domain names led to comprehensive automation, as described in The History of SIDN and the podcast 'The Story of .nl'. From the late 1990s to 2014, .nl grew substantially, as shown in figure 3.1. The domain continued to grow beyond 2014, but at a much slower rate, leading to registration of the 6 millionth .nl domain name on 18 June 2020.

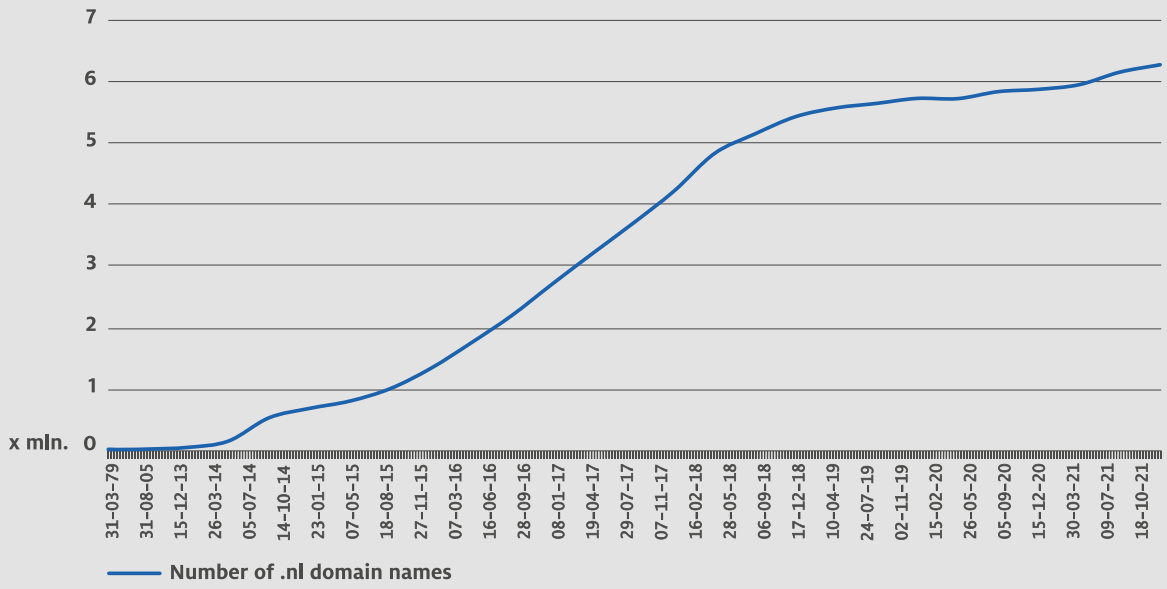


Figure 3.1 | Number of .nl domain names down the years.

3.2 Length of .nl domain names

Domain names made up of just a few characters are relatively popular in .nl. All of the 1,296 possible two-character domain names (not counting '.nl' itself) were registered long ago. At the other end of the spectrum, there are twenty-seven .nl domain names with the maximum number of characters registered (sixty-three), including inhetverledenbehaalderesultatenbiedengeengarantievoordetoekomst.nl ('results obtained in the past are no guarantee for the future dot-nl'. That's up from the twenty-three full-length names we had a few years ago. (See also What's the longest domain name?)

The most common length for a .nl domain name is 12 characters: there are nearly half a million of that length. Across the domain, the average length is thirteen characters. Figure 3.2 shows how many .nl domain names there are of each possible length.

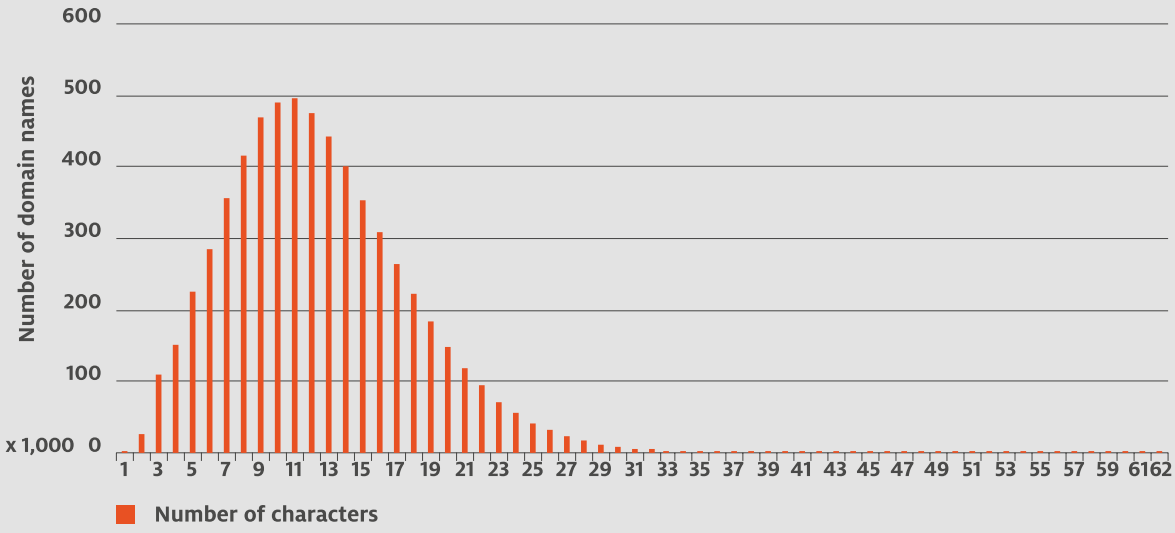


Figure 3.2 | Number of registered .nl domain names of each possible length.

3.3 Regional distribution of .nl domain names

A look at the cities where domain names are registered reveals marked geographical differences. The postcodes recorded for registrants indicate that the highest concentration of .nl domain names is in Amsterdam. Other hotspots include big (university) cities, such as Rotterdam, Utrecht, and Wageningen. Figure 3.3 is a map of the Netherlands, with each municipality colour-coded to show how many .nl domain names are registered there. An interactive version of the map is available on stats.sidnlabs.nl.

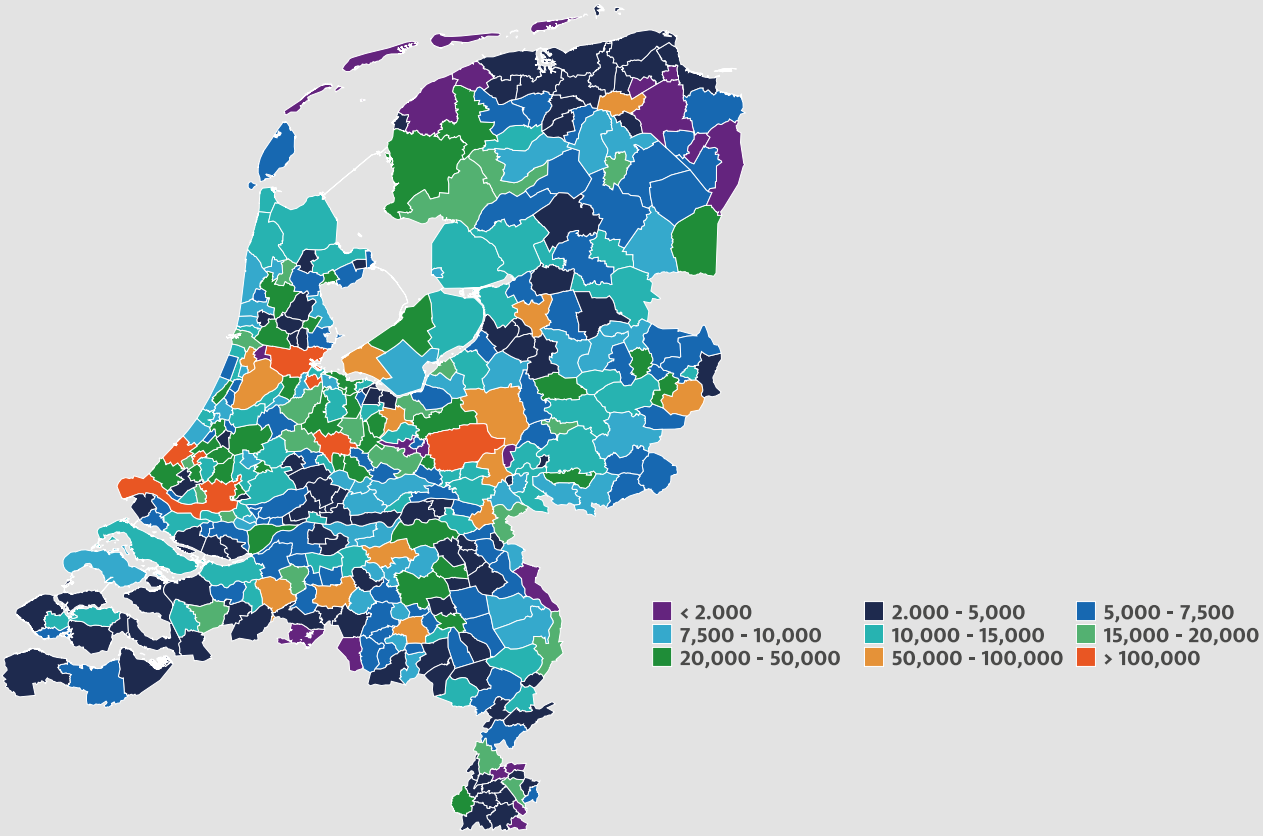


Figure 3.3 | Number of registered .nl domain names per Dutch municipality, based on registrant postcode.

3.4 Age of .nl domain names

Every year, thousands of new .nl domain names are registered. Many are used for a while and then dropped. However, there is a special group of names that have remained in use for a long time. Figure 3.4 shows the percentage of domain names registered per year that were still active at the end of 2021.

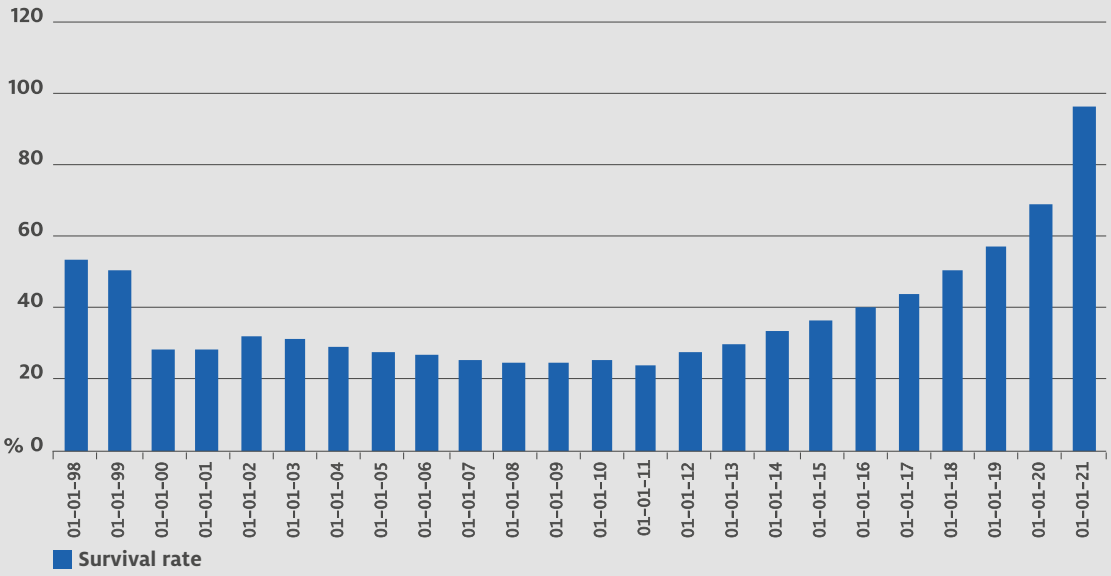


Figure 3.4 | Percentage of .nl domain names registered per year still active at the end of 2021.

Our historical registration data goes back to 1998. (Earlier data has not been retained.) It's notable that a higher proportion of the domain names registered in the first two years of our sequence have remained active than those registered in the years around 2018. That's probably because there were far fewer .nl domain names back in 1998 and 99, making it much easier to claim a name with lasting value.

Because the chart focuses on names still active at the start of this year, the percentages remaining from the last five years are relatively high. Some of those names are ultimately likely to prove short-lived, but have yet to reach the end of their working life. Five years from now, we expect the percentages remaining from those years to be much lower, as domain names referring to outdated entities and initiatives are cancelled.

3.5 Measurement methods

Subsections 3.1 and 3.4 are based on historical data on the .nl zone retained by SIDN.

Figures for the number of registered domain names and their length were obtained by simple counts of the names in the .nl zone.

The geographical distribution data is derived from registrants' postcodes, as provided by registrars.

04

Words in new domain names

This section of the report provides a picture of the words that are most popular for use in new domain names. Do we see surges of names associated with particular topics, for example? Do word choice patterns reflect what's happening in Dutch society? Can we detect any other developments in the registration of domain names? Our analysis focuses mainly on 2020 and 2021.

04

Words in new domain names

4.1 New registrations in 2021

Figure 4.1 is a word cloud made up of the words used in domain names first registered in 2021. The bigger a word appears in the cloud, the more often it was used in a new domain name that year. The cloud therefore reflects the popularity of the various words.

Figure 4.1 | The most popular words in .nl registrations in 2021.

The words 'online', 'shop', 'straat' ('street'), 'test', 'zorg' ('care'/'healthcare'), 'studio' and 'huis' ('house'/'home') clearly stand out as the most popular terms for use in new domain names. In the following subsections, we consider how the popularity of certain words changed from month to month and the reasons for the observed trends.

4.2 Trending words in 2021

While a word cloud can show us how often words are used, they don't necessarily highlight the year's 'trending words'. Can we identify such words, and relate them to what's happening in society?

Table 4.1 lists the ten biggest trending words in 2021. The trending words were identified by comparing the frequency of each word's use in domain names newly registered in 2021 with its prevalence all .nl domain names active that year. That enables us to distinguish words that became much more popular in 2021 from words with enduring popularity. See subsection 4.7 for more information about the identification of trending words.

Ranking	Word	JLH score	New in 2021	Number in all .nl domain names	Percentage in 2021
1	straat	0.0182	4,606	14,189	32%
2	meta	0.0085	1,085	1,962	55%
3	test	0.0082	3,396	14,453	23%
4	crypto	0.0077	1,633	4,419	37%
5	the	0.0067	8,299	58,688	14%
6	laan	0.0056	1,673	5,758	29%
7	padel	0.0055	880	1,906	46%
8	pcr	0.0055	478	626	76%
9	verse	0.0041	604	1,234	49%
10	happy	0.0039	1.820	8,351	22%

Table 4.1 | Trending words in 2021.

The trending words listed in table 5.1 clearly reflect what was happening in society in 2021. At the top comes 'straat' ('street'), with 'laan' ('lane') in fifth place. We investigated the domain names featuring those words more closely and found that nearly all related to property addresses. Their popularity reflects the trending practice of creating a dedicated website to sell a house. Particularly where higher-value properties are concerned, it's now common for a vendor to make extensive information about their property available on a specially created website with a domain name based on the property's address.

Other events and developments show up in the trending name data too. For example, the popularity of 'meta' has shot up since Facebook announced that its holding company was changing its name to Meta. And, with cryptocurrencies and the sport of padel attracting interest, 'crypto' and 'padel' have appeared in more domain names. Of all the domain names containing those words, 37 and 56 per cent, respectively, were first registered in 2021.

Finally, COVID-19 has of course had a big impact on domain name registrations. The number of domain names including 'test' soared, although not all can be linked to COVID-19. The initials 'PCR', with their obvious pandemic link, were also in frequent use for domain names.

4.3 Trend response speeds in .nl registrations

Having considered how social themes are reflected in the words used in domain names, it's interesting to see how quickly word popularity trends change in response to developments.

We've investigated response speeds by focusing on registrations that feature pandemic-related words. Data from both 2021 and 2020 has been examined, in order to pick up trends since the start of the crisis

Figure 4.2 shows the effect of coronavirus press conferences on the number of domain names containing certain terms. The x-axis shows the period from February 2020 to February 2021, with the black vertical lines representing key events, such as the press conference at which the first COVID-19 case in the Netherlands was announced. The number of registrations containing each of three pandemic-related terms – 'corona', 'anderhalve' ('one and a half') and 'mondkap' ('facemask') – are plotted on the y-axis.

From the graph, it's clear that the press conferences had an almost immediate effect on domain name registrations. The number of domain names containing 'anderhalve' (orange line) was fairly stable until early April. That's when the government began advising people to keep 1.5 metres apart, giving rise to the phrase 'anderhalvemetersamenleving' ('one-and-a-half-metre society'). A similar effect is visible when we look at domain names that include 'mondkap' ('facemask'; blue line). Registrations began rising steadily after the first case was reported, then exploded in early May 2020, when masks became mandatory on public transport. Domain names containing 'test' or 'pcr' (red line) increased steadily from the start of the pandemic, but accelerated markedly when the second infection wave hit in October 2020.

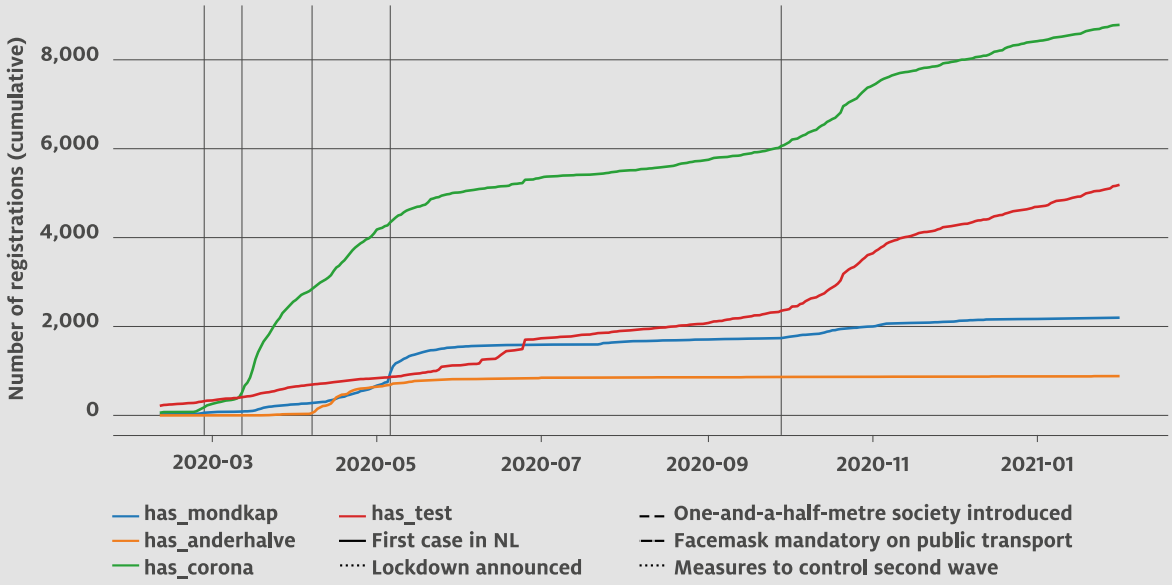


Figure 4.2 | Speed of registration trend response following press conferences.

4.4 Context of words in .nl domain names

Moving on from the prevalence of certain words in .nl domain names, it's instructive to consider the words used on web pages linked to domain names. Web pages contain much more text, enabling the analysis of other variables.

For example, we've investigated the context in which words are used. If two words are used in the same context, that indicates that they are often used in combination. For example, the words 'stoel' ('chair') and 'tafel' ('table') frequently occur in the same sentence or paragraph. Because they share the same context, you can assume that the words have closely related meanings.

However, the context of a word can of course change over time. Table 4.2 lists a number of keywords in the first column and the words associated with them in the second and third columns. Column 2 lists the associations in summer 2018, while column 3 lists those in autumn 2021.

Keyword	Associated words in 2018	Associated words in 2021
Corona	desperados, birra, anejo, drambuie, cerveceria, cerveza, paulaner, campari, nastro, maho	covid, covid19, corana, coronavirus, coronamaatregelen, coronacrisis, carona, afgekondigde, lockdown, coranavirus
Crisis	crises, recessie, kredietcrisis, dreiging, teruggang, malaise, bezuinigingen, hervormingen, protesten, werkeloosheid	pandemie, coronacrisis, crises, recessie, epidemie, lockdowns coronapandemie, gezondheidscrisis, kredietcrisis, lockdown
Anderhalve	tweeënhalve, drieënhalve, welgeteld, vierhonderd, dertiende, halverwege, 13de, pakweg, dertig, hooguit	afstandsregel, afstandregel, anderhalvemeter, veiligheids-afstand, avondklok, afstandhouden, afstandsregels, 5meter, 4voorfier, teruglever

Table 4.2 | Contexts of words.

As you can see, the context in which 'corona' was used changed enormously. In 2018, the word was associated mainly with alcoholic drinks, reflecting its use as a beer brand. In 2021, it was used mainly in the context of COVID-19. A similar shift took place in the use of 'crisis'. In 2018, it was being used predominantly in an economic context, but in 2021 the crisis on everyone's minds was COVID-19.

4.5 Popular words in each month of 2021

Presented below is a word cloud for each month of 2021, showing the most popular words in domain names newly registered that month.

The most notable feature of the word cloud series above is the consistent popularity of 'online' and 'shop'. That isn't linked to a specific event, but is easily explained by use of those words as standard domain name components. The word 'online' obviously goes with any internet service, while the English word 'shop' is widely used in Dutch for the names of webshops. The prevalence of 'shop' tells us, of course, that numerous new webshops were created in 2021, but that is only to be expected.
As explained in subsection 4.2, the word 'straat' was prominent in 2021, and the monthly word clouds confirm that it was consistently popular throughout the year. To a lesser extent, the same is true of 'crypto'. Another word mentioned in subsection 4.2 was 'meta'. What the monthly clouds tell us there is that, unlike 'straat' and 'crypto', 'meta' shot to popularity only in October, when Facebook announced the name change. The social media platform Clubhouse was associated with a similar but less pronounced effect earlier in the year: 'club' enjoyed a brief surge in popularity in February, March, and April.

4.6 Popular words in each month of 2020

Since we didn't produce a 'State of .nl' report on 2020, we thought it would be interesting to retrospectively apply the same technique used for the analysis of 2021 to examine the first year of the pandemic as well. In March 2020, the Dutch government introduced the first measures to control the spread of coronavirus. A sense of the seriousness of the emerging pandemic is clearly reflected in the registration patterns: 'corona' suddenly shoots to the top of the word list. The word 'thuis' ('(at) home') also enters the top ten.

In April, 'anderhalve' and 'meter' appear in the word cloud as well, as people start using the term 'anderhalvemetersamenleving' ('one-and-a-half-metre society').

In May, facemasks became a topical issue and masks were exempted from the purchase tax BTW/VAT. As our word cloud for that month shows, websites selling facemasks mushroomed.

2020

Figure 4.3 | Popular words in .nl domain names in 2020.

Figure 4.4 | Popular words in .nl domain names in 2021.

4.7 Measurement methods

We identified the words making up domain names using a word splitter: an algorithm that breaks a character string into its component parts, such as 'merk' ('brand') and 'bewaking' ('guarding') in the domain name sidnmerkbewaking.nl, which we registered for our BrandGuard service.

In order to generate the word clouds, we ran the word splitter on all the domain names registered in the relevant period. We then filtered the results to remove articles, prefixes and adjectives, then counted how many times the other words occurred.

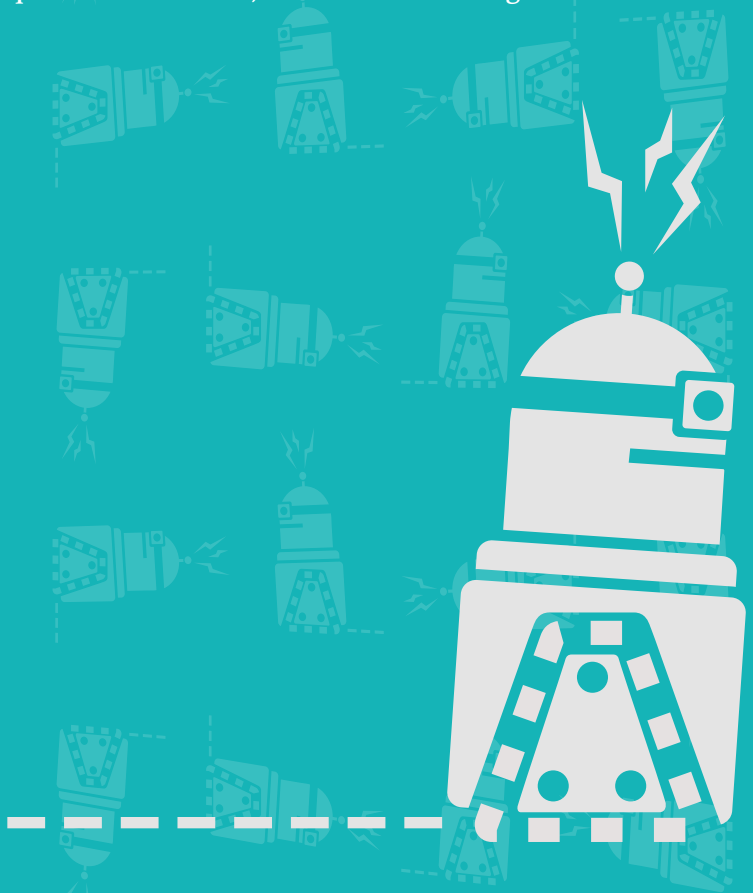
Trending words were identified by calculating ElasticSearch JLH scores.

We established the context in which words were used by automated scanning of the content of the websites for domain names. That was done using Word2vec to calculate 'embedding vectors': automatically generated word representations consisting of sequences of numbers defining the meaning of the words. We then calculated the cosine distance between pairs of number sequences, giving us a numeric expression of how dissimilar the two sequences were. The smaller the cosine distance between two sequences, the more closely related the two words are in their meaning.

05

DNS resolvers

In this section of the report, we examine data on the entities that look up information in the DNS on behalf of internet users: the DNS resolvers that send queries about .nl domain names (see section 2). Topics explored include the times of day that the .nl name servers receive most DNS queries, where the queries come from, and what technologies the DNS resolvers are using.



05

DNS resolvers

5.1 Patterns in resolvers' DNS queries

For insight into DNS resolvers' query patterns, we analysed all the incoming query traffic for the month of June 2021. The results are presented in figure 5.1. The time of day is plotted on the vertical axis (with 00:00 at the top and 23:59 at the bottom), and the day of the month on the horizontal axis (1 June on the left, 30 June on the right).

A number of patterns are visible. First, traffic is heavier in the daytime (from 05:00-06:00 to 20:00-21:00) than at night – reflecting the day-night rhythm of the people who use .nl domain names. Most of the weekends also stand out as relatively low-traffic periods, particularly Saturday 5 and Sunday 6 June.

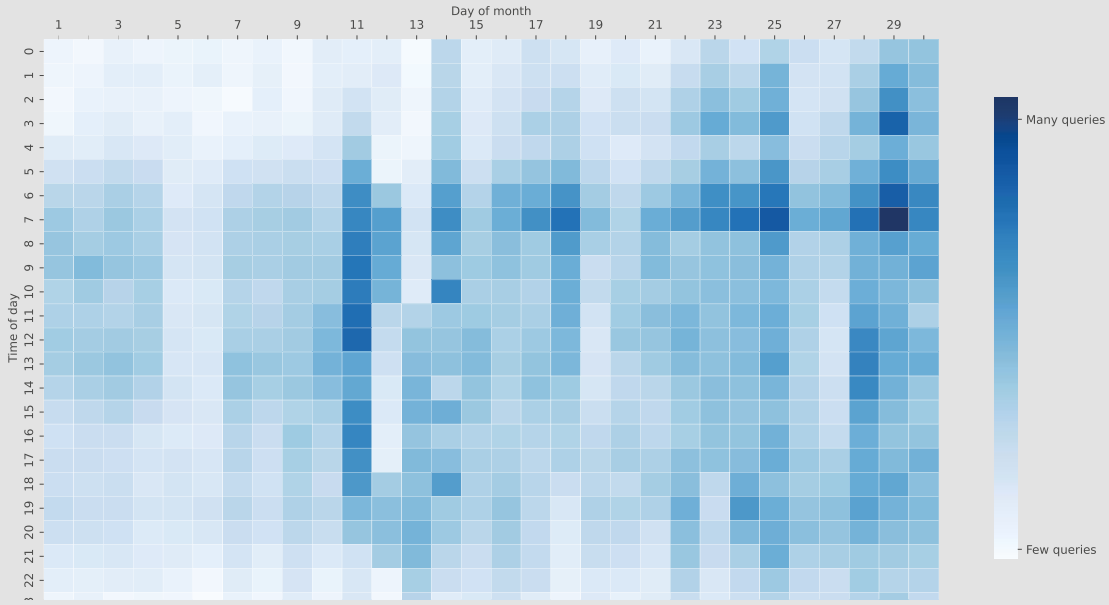


Figure 5.1 | Number of queries received from DNS resolvers per hour in June 2021.

Figure 5.2 is a 'zoomed out' version of figure 5.1, showing the number of queries received per day in 2021. The patterns are easy to discern.

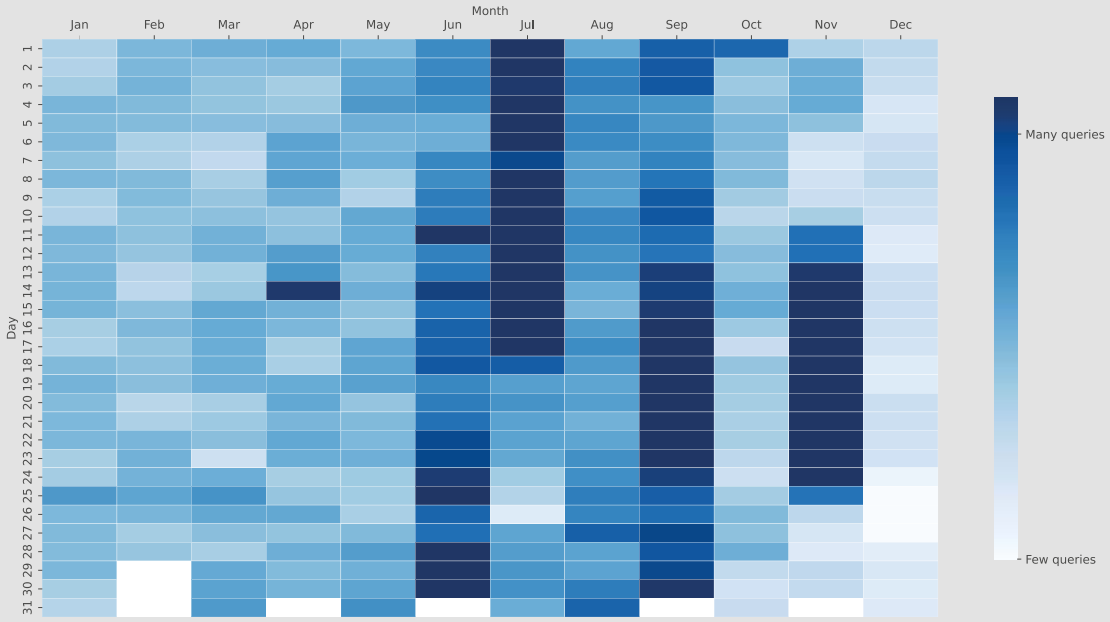


Figure 5.2 | Number of queries received from DNS resolvers per day throughout 2021.

5.2 Network protocols: IPv4 and IPv6

As well as relating to either an IPv4 address or an IPv6 address (see section 2), a query from a DNS resolver may be sent using either IPv4 or IPv6. Figure 5.3 shows the breakdown of incoming query traffic to the .nl name servers between IPv4 traffic and IPv6 traffic. The breakdown reflects the level of IPv6 adoption by DNS resolvers.

As you'll see, the percentage of the traffic accounted for by IPv6 gradually increased during the year, from just over 20 per cent at the start to nearly 30 per cent at the end. We regard that as a positive development, since the pool of IPv4 addresses is dry. However, the rate of increase is nowhere near sufficient for us to anticipate widespread adoption within a reasonable timeframe. More IPv6 statistics are presented in section 8.

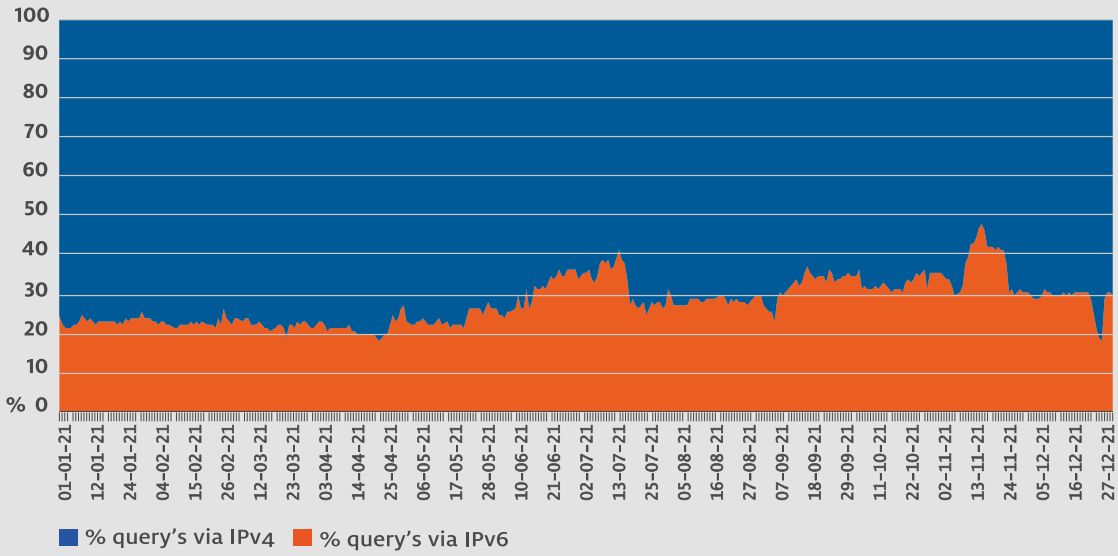


Figure 5.3 | Breakdown of incoming query traffic between resolvers with IPv4 addresses and those with IPv6 addresses.

5.3 Resolver locations

In order maximise the availability of .nl, we have several dozen authoritative servers distributed around the world to handle queries from 'nearby' DNS resolvers (see section 2). That's important, because the query traffic we receive isn't all from DNS resolvers in the Netherlands.

The international character of the queries handled by our servers is visualised in figure 5.4, where the locations of querying resolvers are plotted on a map. As one would expect, many of the resolvers asking about .nl domain names are in the Netherlands itself. However, the United States are also well represented, partly because many major DNS services are based there, including Google Public DNS and CloudFlare DNS.

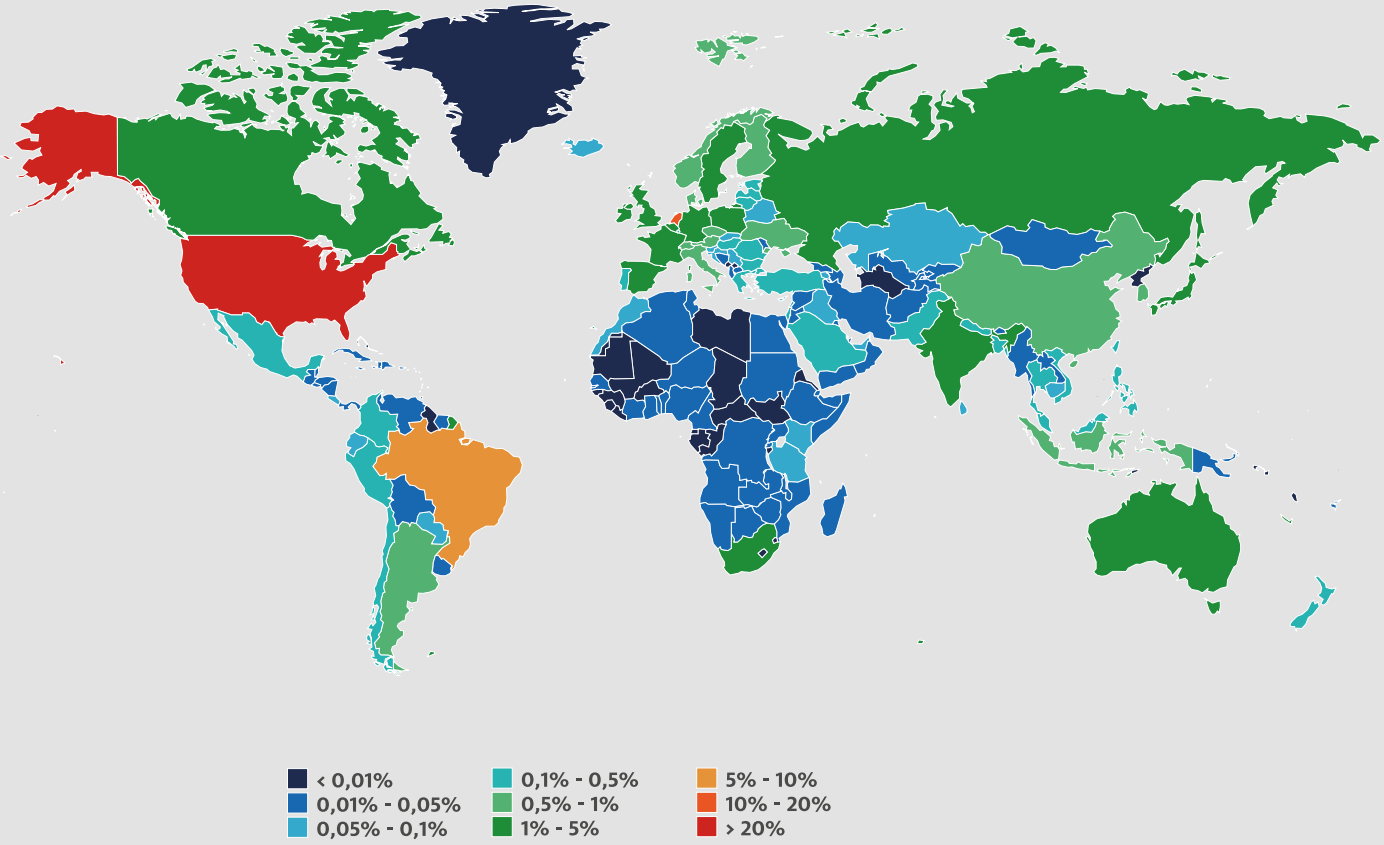


Figure 5.4 | Locations of DNS resolvers querying the .nl servers.

5.4 Routing security

Figure 5.5 shows the percentage of .nl domain names that pointed to web servers in networks using Resource Public Key Infrastructure (RPKI) security over the course of 2021. RPKI is an open standard designed to reduce the risk of network traffic getting hijacked. A route hijack can have serious implications. One scenario is that DNS queries from a web browser are hijacked and diverted to a malicious resolver, without the user or their browser detecting anything. The malicious resolver then directs the browser to a malicious server, thus enabling a phishing attack, for example.

In figure 5.5, the dark green part of the graph represents the percentage of domain names that were fully RPKI-secured, while the light green represents partially secured names. At the start of 2021, the fully secured proportion had just passed 50 per cent. However, it seems that growth of RPKI support has now plateaued. That's a pity, because RPKI can make an important contribution to the enhancement of internet security.

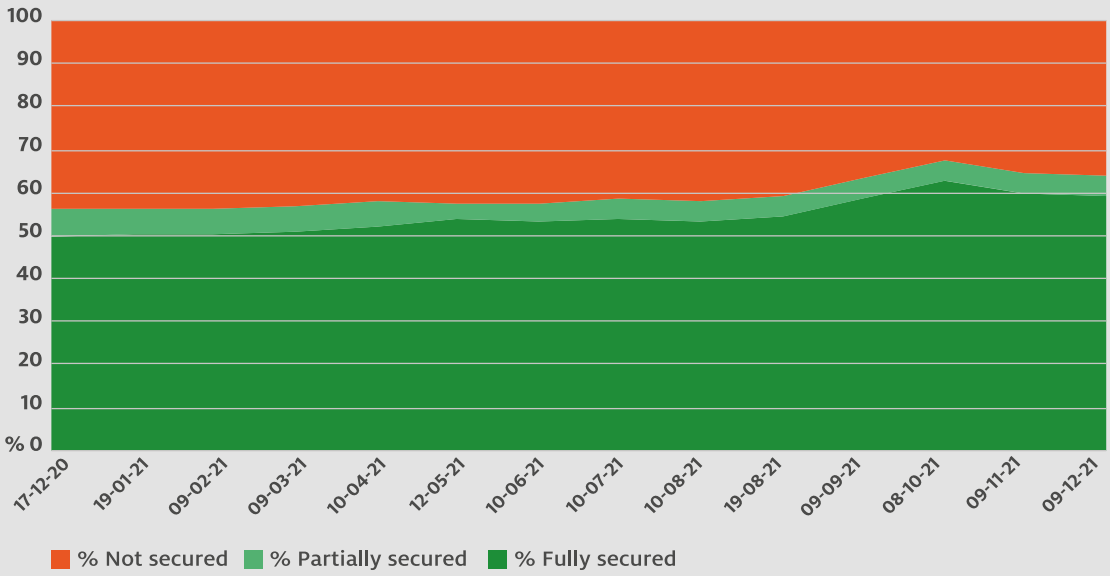


Figure 5.5 | Percentage of .nl domain names secured with RPKI.

5.5 Measurement methods

Most of the statistics presented in this section are drawn directly from ENTRADA (see subsection 2.4). The countries of origin of DNS queries were determined using a geoIP library: a resource that gives the geographical location of IP addresses and networks. The RPKI support graph was created by looking up the IP addresses of .nl websites and checking whether the addresses belonged to RPKI-secured networks.

06

Centralisation of the internet

Centralisation of the internet is the ongoing process whereby an increasingly small number of large players are responsible for an increasingly large proportion of services, data and knowledge. For example, more and more internet users now reply on major public resolvers, rather than resolvers operated by their internet service providers (ISPs). Centralisation is a topical issue because, although the big providers generally provide very good services, the concentration of control can introduce security and stability risks. If one provider has a technical problem, a relatively large proportion of the internet can be affected.

In this section of the report, we use our measurement data to gauge the potential centralisation-related risks to the security and stability of .nl domains.

06

Centralisation of the internet

6.1 Authoritative .nl name servers

One way of making online services more resilient is to use multiple authoritative DNS servers (see section 2). If one of your multiple servers is disabled by, say, a fault or a cyber-attack, DNS traffic is automatically routed to the others. However, the benefit of having multiple name servers is limited if all your servers are hosted by the same service provider. When the services of OHVcloud and Akamai recently went down, for example, thousands of domain names were temporarily unreachable.

International distribution of name servers

Figure 6.1 shows the international distribution of .nl name servers, other than SIDN's servers. Strikingly, a little more than half are located in the Netherlands itself. The US is the next most important country, accounting for nearly 20 per cent. The breakdown for 2021 was broadly similar to that for 2020.

The international distribution of servers is otherwise not an important indicator of centralisation. It's generally desirable to locate name servers close to the target user population, to minimise the processing time of DNS queries. The Netherlands is the predominant location of .nl users, so it's also the obvious place to locate the servers.

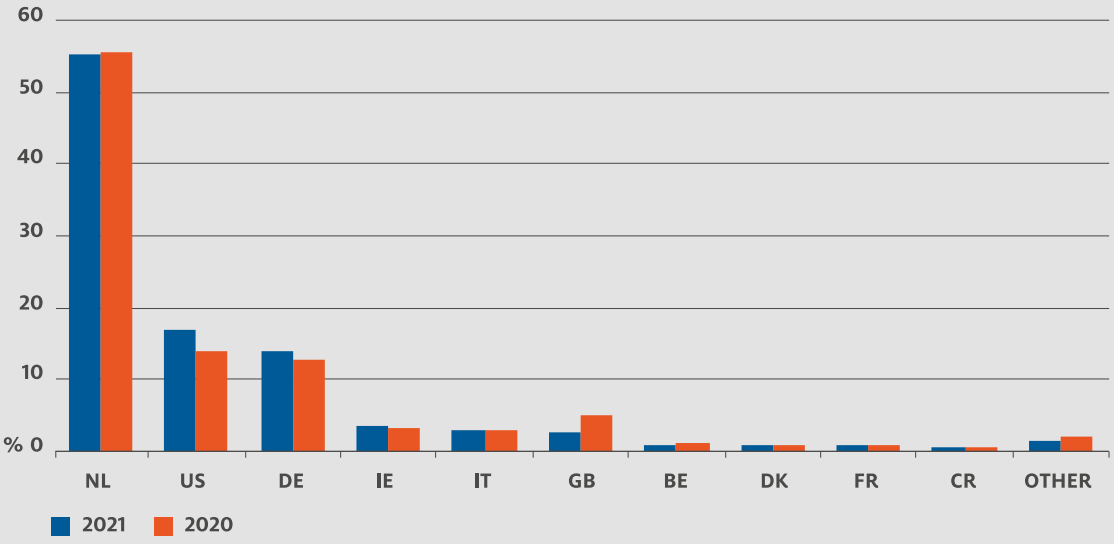


Figure 6.1 | International distribution of .nl name servers, excluding SIDN's servers.

Distribution of name servers across networks

By contrast, the distribution of name servers across networks is an important indicator of centralisation. Figure 6.2 shows how many distinct networks .nl name servers belong to. The main thing to observe is that nearly 50 per cent of all .nl domains have their name servers on a single network. The domain names in question are therefore vulnerable to any issue affecting that network. If the network goes down, as happened with Akamai, the name servers and therefore the domain name will be unreachable. The .nl domain names whose name servers are poorly distributed include many actively used names: 51 per cent of the domain names used for webshops have their name servers on a single network.

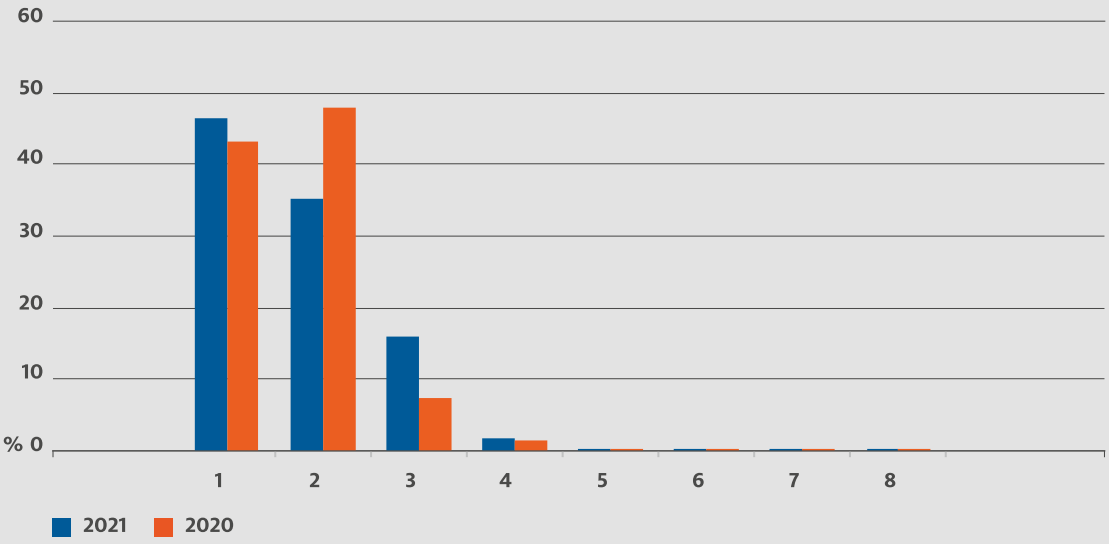


Figure 6.2 | Distribution of .nl name servers across distinct networks.

Increased reliance on single networks

One might expect the major outages mentioned above to have served as a warning to domain name operators about the dangers of reliance on a single service provider or network. However, it seems that the number of domain names with all their name servers on a single network actually increased by 3 percentage points in 2021.

Consolidation in the registrar community

One important explanation for the increased reliance on single networks is consolidation in the registrar community. The .nl registrars are businesses that register domain names for their customers, and usually also operate name servers. The number of .nl registrars has been declining for some years, with an increasingly large proportion of the .nl zone managed by a small number of actors. Another factor is that having multiple name server providers is often inconvenient. For example, enabling DNSSEC security (see subsection 2.4) is more complicated if you use several providers. Although protocols are fortunately now under development that should make multiple provider use more straightforward, that is unlikely to arrest the trend towards reliance on single providers.

Assuring full availability is hard

Another issue is that certain networks often host the name servers of thousands of domain names. Ten networks, most in the Netherlands and the US, host 64 per cent of all the name servers responsible for .nl domain names. That's up from 61 per cent just a year ago. Of course, network operators do their best to make sure their networks are always reachable. However, as the Akamai, OVH and Dyn incidents show, 100 per cent availability is impossible to assure, even for the biggest service providers.

6.2 Recursive resolvers

Recursive resolvers play an important role in the DNS (see section 2). If a major recursive resolver goes down, a large number of end users may be completely unable to reach any websites. Furthermore, the DNS queries and responses handled by such a resolver often give the operator detailed insight into end users' browsing behaviour.

Public resolver services

At the resolver level too, centralisation of the DNS infrastructure appears to be increasing. One driver of the trend is the growing proportion of query traffic originating from public resolver services, such as Google's 8.8.8.8 and van Cloudflare's 1.1.1.1.

Over the last year, we've observed a marked increase in queries from such providers. Figure 6.3 shows the percentage of incoming query traffic to the .nl name servers originating from public resolvers, with the trend line in blue. Roughly one query in four now comes from a public resolver service.

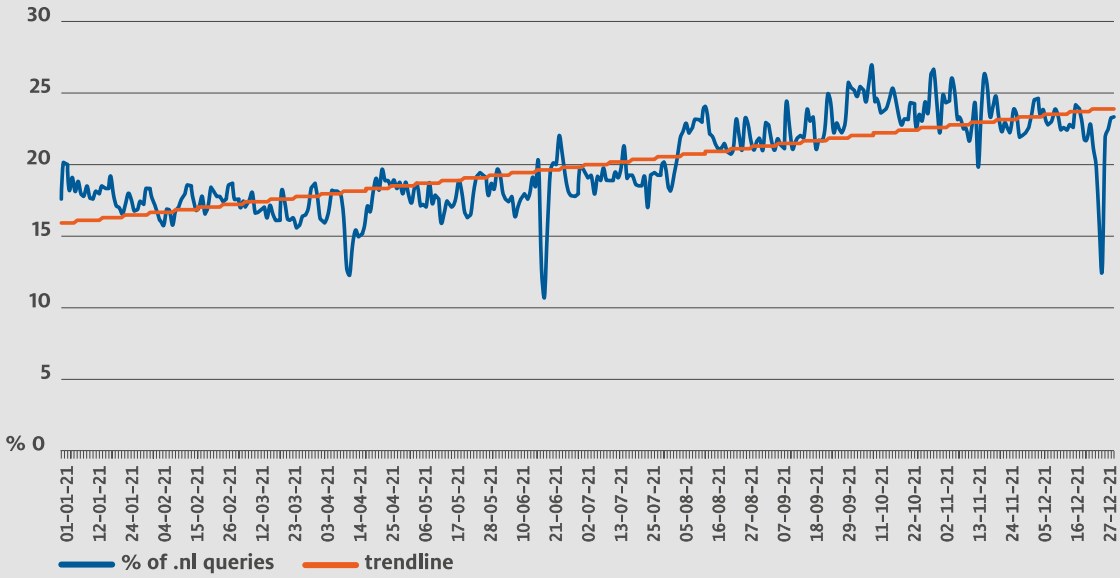


Figure 6.3 | Queries originating from public DNS resolvers.

Although one might expect most DNS queries about .nl domain names to come from the Netherlands, the dominance of US giants such as Microsoft and Google is again apparent. Of all the queries received in 2021, 33 per cent originated in the US and 26 per cent in the Netherlands. However, queries from US resolvers actually showed a downward trend: at the start of 2021 resolvers in America accounted for about 35 per cent of queries, but by the end of the year the figure was 31 per cent.

Local traffic handling

One reason for that trend was the rise of local DNS traffic routing in 2021. More local routing means that a DNS query from a user in the Netherlands that would have been sent to a US resolver at the start of the year is often now being handled by a resolver in Europe operated by the same service provider. Although we do measure 'anycast catchments', the findings are not considered in this report.

6.3 Web hosting

In the web hosting sector too, a number of big providers account for an ever-greater share of the services. The downside of that concentration is that, if one of the biggest providers were to suffer a service outage, a significant proportion of Dutch websites would be temporarily unavailable.

Actively used domain names

Figure 6.4 shows the ten hosting service providers responsible for the highest numbers of actively used .nl domain names. (Our definition of 'actively used' is explained in subsection 6.5.) We concentrate on actively used domain names because that is the most significant subset of names in the context of hosting sector centralisation. The reason being that a fault or a security problem affecting a service provider would have much more serious implications for active domain names than for, say, names used for parking pages.

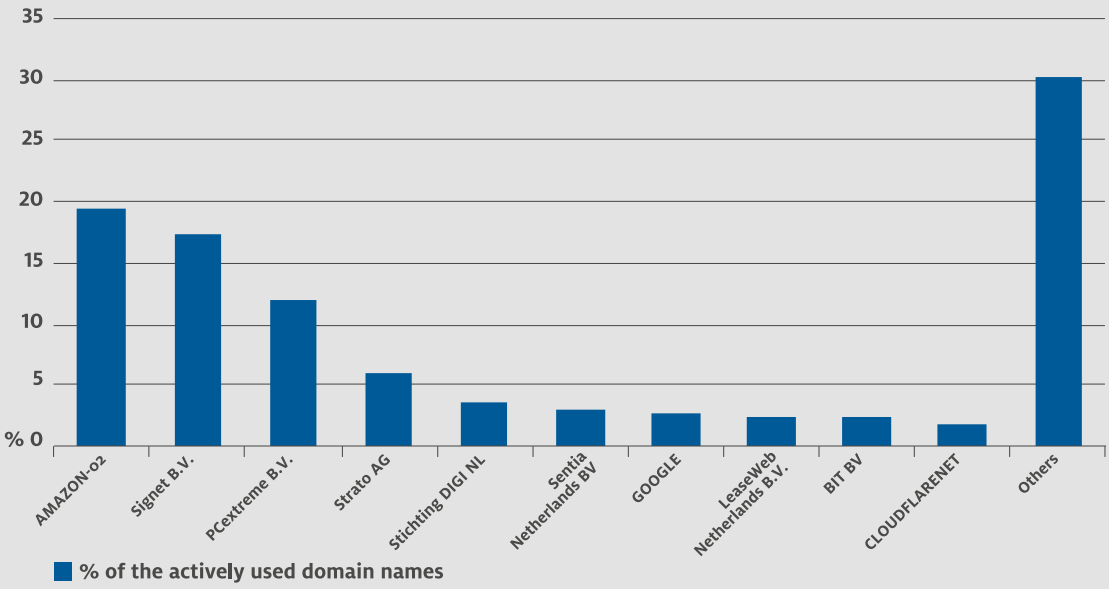


Figure 6.4 | Concentration in the .nl web hosting sector.

The right-most bar in figure 6.4 ('Other') indicates that 30 per cent of actively used .nl domain names are hosted by service providers outside the top ten. Within the top ten, the three biggest providers are well clear of the rest in terms of the number of active names hosted. They are Amazon (19 per cent of the total), Signet (17 per cent) and PCextreme (12 per cent). Together, those three companies provide web hosting for more than 48 per cent of actively used .nl domain names. After the top three, the percentage of domain names hosted per service provider soon falls away.

6.4 E-mail

Resolvers use the DNS to look up the mail servers linked to domain names in much the same way that they look up web servers (see section 2). In this field, concentration takes the form of numerous mail servers sharing the same infrastructure (networks, IP addresses, etc). Again, the risk is that availability problems or an outage affecting a single major mail service provider will have far-reaching consequences.

Many domains have a single IP address for mail

We have established that most domain names have one, two, three, four or ten IP addresses for their mail servers. As figure 6.5 shows, almost none have any other number of addresses. (NB: the y-axis is to a logarithmic scale.) Of the domain names using IPv4, 88 per cent have just a single IP address for mail. Amongst the domain names using IPv6, the situation is worse still: 98 per cent have a single IP address. It's also interesting to note that the fifth most common number of IP addresses to have is ten, not five, as one might imagine.

If a domain name has only one mail server IP address, the domain's mail server (or mail server group) is connected to the internet via a single network. A fault with that network would therefore inevitably

hit the domain's mail service. The risk of a network-related mail service interruption is significantly reduced if the server is connected via multiple networks, or if the domain has multiple mail servers on different networks.

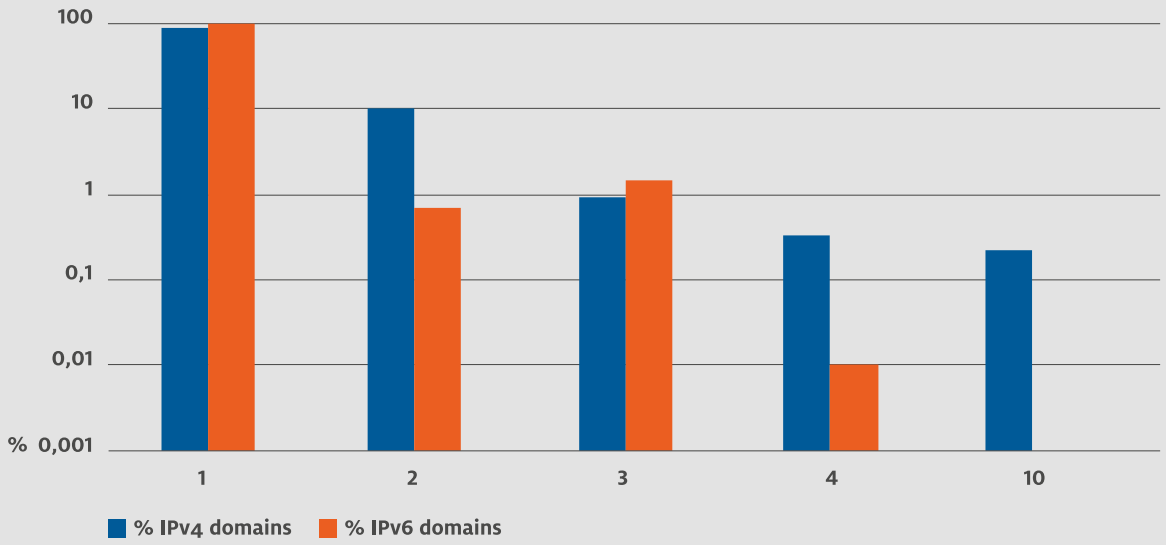


Figure 6.5 | Number of mail server IP addresses linked to .nl domain names.

Multiple IP addresses but a single network

It's common for .nl domains to have multiple mail server IP addresses, but on the same network (ASN).

That phenomenon is reflected in figure 6.6 (again with a logarithmic y-axis): 97 per cent of the domain names with IPv4 mail server addresses use a single network, as do 98 per cent of those with IPv6 addresses.

From the service reliability viewpoint, there is no advantage in having multiple e-mail server IP addresses if the addresses are on the same network. If the network goes down, both addresses will be unreachable.

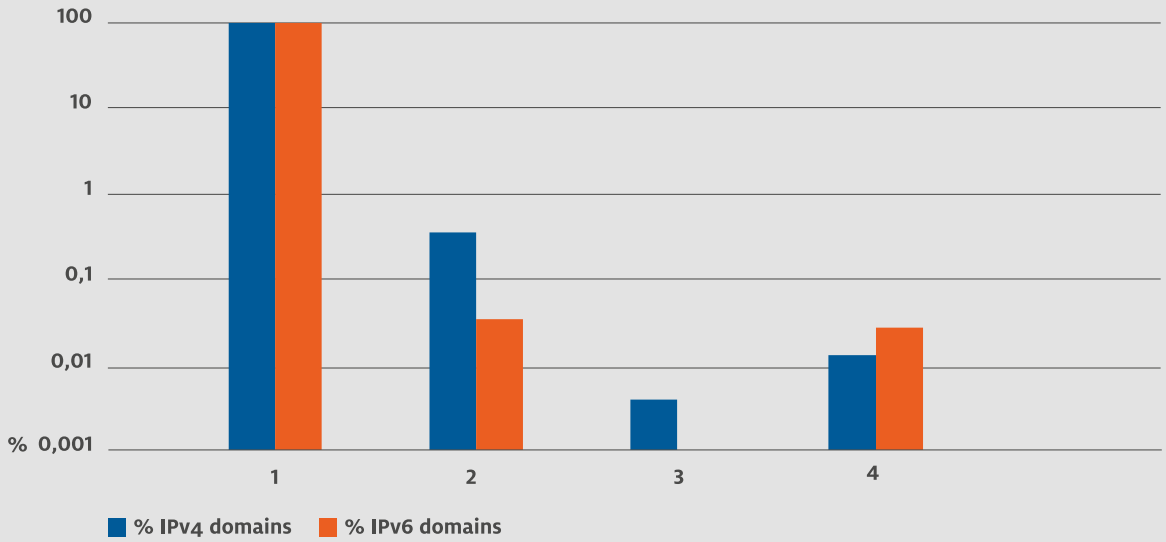


Figure 6.6: | Number of networks used for .nl mail servers.

6.5 Measurement methods

The .nl name server locations reported in subsection 6.1 were established using a GeoIP database. Network numbers can be obtained from the routing database, together with information about the network owner. We define an 'actively used domain name', as referred to in subsection 6.1, as a domain name linked to a website with proper content. That may be a business website, a forum or a small private website with text only. We do not count domain names linked to an internet service provider's parking page, or a 'reserved for use' notice or similar as actively used. We determined domain names' status from our DMAP dataset (see subsection 2.5).

Query traffic from public resolvers, as referred to in subsection 6.2, was quantified using a list of such resolvers' networks and IP addresses, in combination with information from the routing database.

The routing database was also used to identify web hosting networks, as referred to in subsection 6.3.

For the data presented in subsection 6.4, we drew on DMAP information to establish the number of networks used for .nl domains' mail servers, again in combination with data from the routing database.

07

Use of DNSSEC for domain name security

So far in this report, we have presented a variety of data illustrating how important the DNS is to society. Because the system is so important, it's vital that internet users can have confidence that, when they enter a domain name into a browser, they'll be taken to the right website. With a view to providing that confidence, the IETF standardised DNSSEC in 2005 (see section 2).

This section of the report explores various DNSSEC-related developments discernible in the .nl zone in 2021. Despite being largely hidden from end users, such developments have a major bearing on users' online safety.

07

Use of DNSSEC for domain name security

7.1 DNSSEC-adoption in .nl

DNSSEC helps to make the internet more trustworthy. We have therefore been promoting the use of DNSSEC for some years. We offer adoption incentives to .nl registrars, for example, as well as providing training and practical implementation guides. As a result, nearly 60 per cent of .nl domain names are DNSSEC-enabled (see figure 7.1). In absolute terms, no top-level domain has more signed domain names than .nl. However, as figure 7.1 shows, the growth of DNSSEC in the .nl zone has plateaued in recent years.

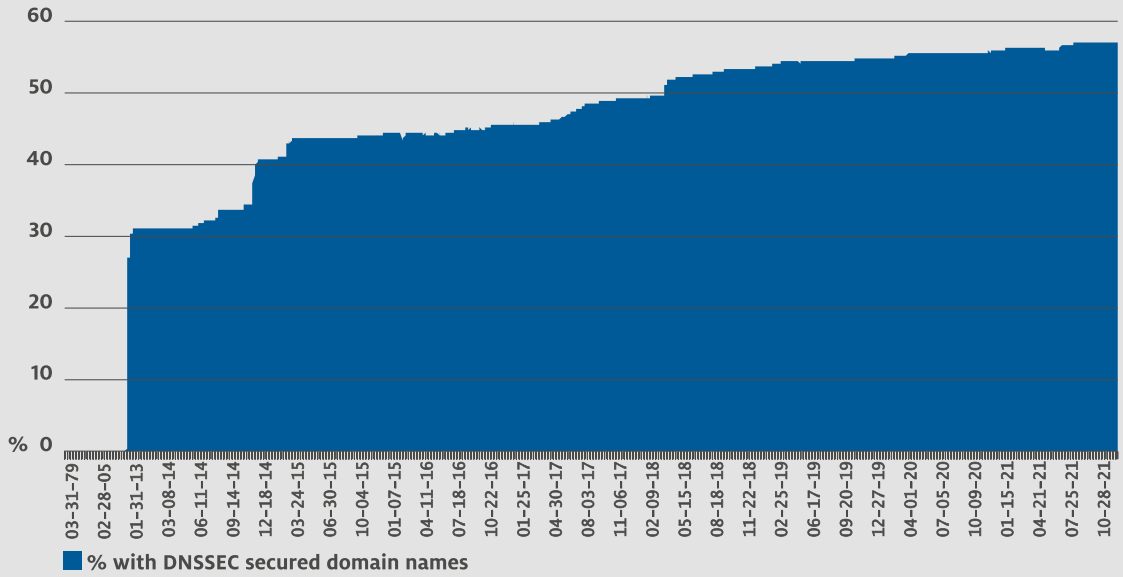


Figure 7.1 | Adoption of DNSSEC in the .nl zone.

7.2 DNSSEC drivers

The potential impact of a successful DNS cache poisoning attack, as described in section 2, is greater if the targeted domain name belongs to a bank than if it belongs to a private blogger, for example. It's therefore instructive to establish where DNSSEC protection is still lacking.

DNSSEC use by website type

As figure 7.2 shows, e-commerce websites are particularly unlikely to support DNSSEC. That's despite the fact that a successful cache poisoning attack on such a site could easily result in the theft of data from end users who are directed to a fake website without them realising. Business domain names are more likely than most to be DNSSEC-enabled, yet 40 per cent of them still lack this important layer of protection.

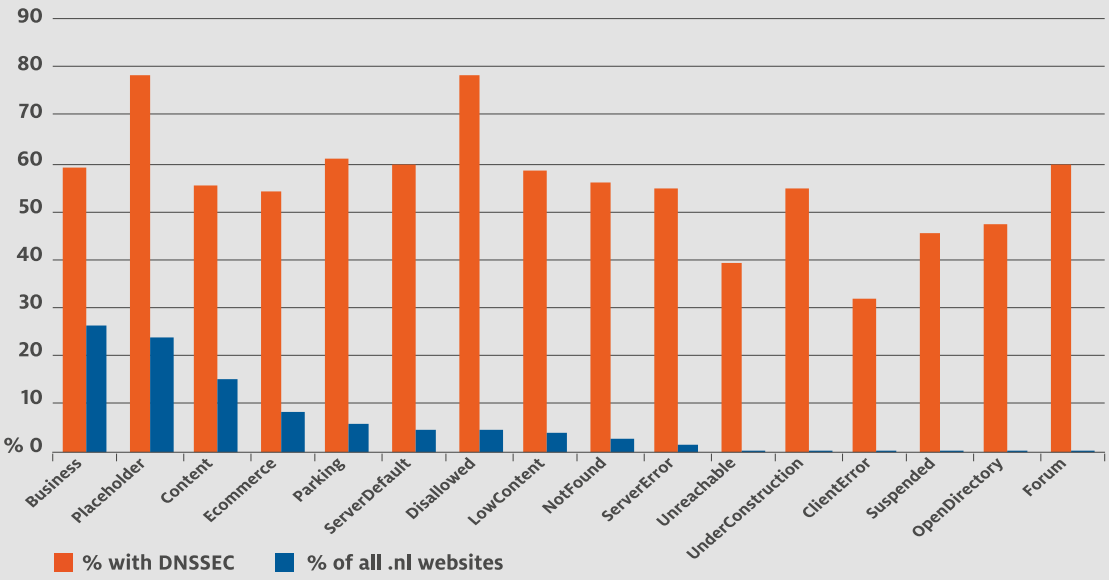


Figure 7.2 | .nl domain names with DNSSEC support, according to website type.

DNSSEC use by sector

Further analysis reveals that, amongst websites used for e-commerce or business websites, those belonging to Information and communication businesses often don't support DNSSEC (see figure 7.3). That's surprising, because one would expect such businesses to know more than most about rolling out DNSSEC. Generally speaking, however, support levels do not vary greatly from sector to sector, with the average just below 60 per cent.

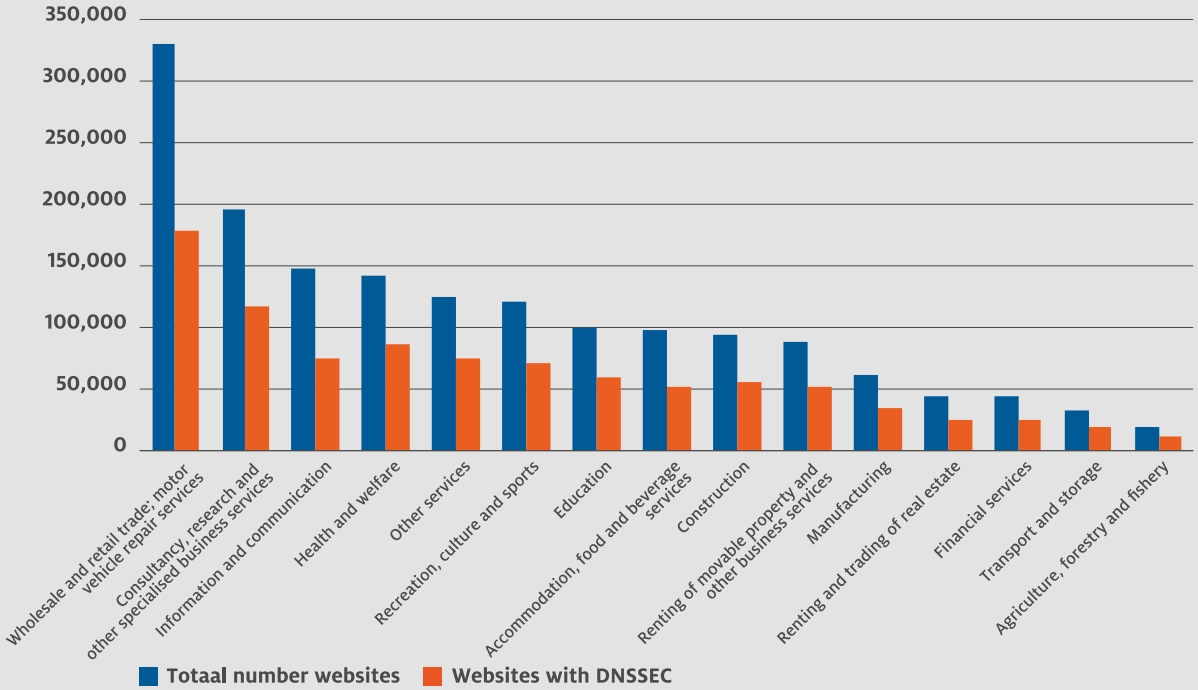


Figure 7.3 | DNSSEC use by sector.

DNSSEC use in the public sector

DNSSEC is commonly used for domain names belonging to local and national government entities. Some 90 per cent of national government websites are DNSSEC-enabled (according to the National Government Website Register). The corresponding figure for municipal websites is 98 per cent, and for provincial websites 92 per cent. The main reason for the high levels of support is that the government requires such services to be DNSSEC-enabled.

The effectiveness of the requirement becomes all the more clear if one compares the figures above to data on the private sector. Only 58 per cent of Dutch banks' domain names are DNSSEC-enabled, for example.

7.3 Adoption of new DNSSEC algorithms

DNSSEC has, of course, continued to evolve since SIDN first used it to secure a .nl domain name in 2012. As a result, some of the algorithms available for signing DNS records are now regarded as insecure (e.g. RSA/SHA1). It's therefore important to make sure that secure cryptographic algorithms are used instead.

Futureproofing

In the last year or so, two big service providers made significant moves in that context. At the end of 2020, Mijndomein went over to signing domain names exclusively with ECDSA P256. Shortly afterwards, TransIP opted to migrate away from the insecure RSA/SHA1 algorithm with an intermediate step to ECDSA as well. ECDSA P256 is a future-oriented algorithm that offers performance benefits on account of the very small signatures it generates.

Usage levels of the various available cryptographic algorithms are presented in figure 7.4. The plots show a clear migration from RSA-SHA256 to ECDSA in recent times.

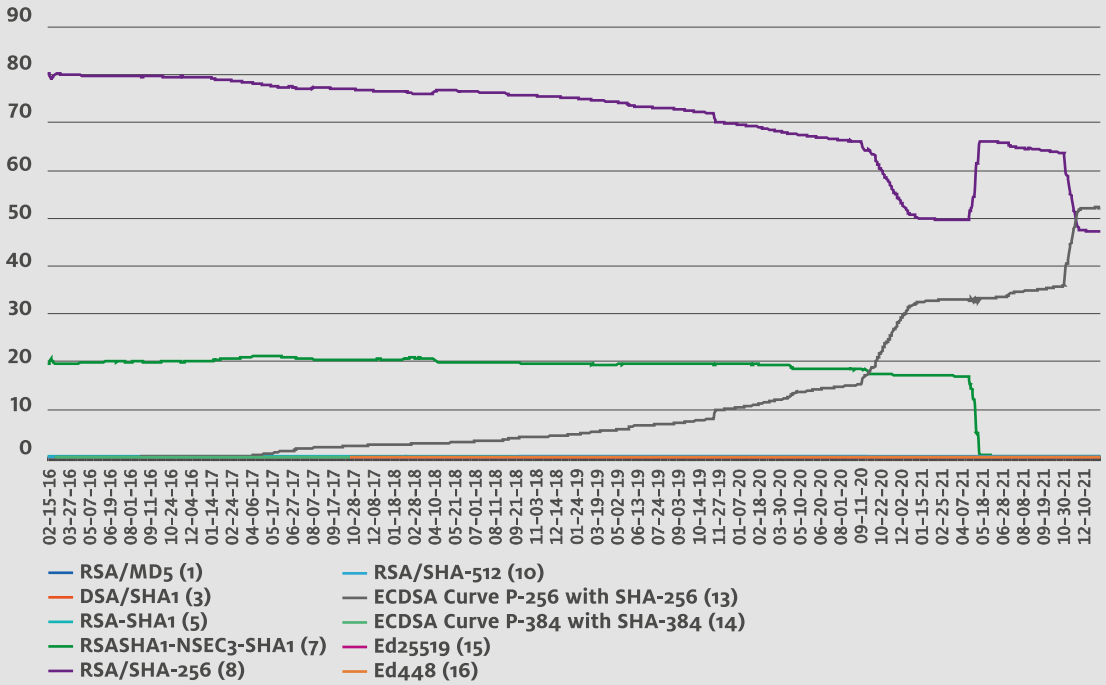


Figure 7.4 | DNSSEC algorithms used in the .nl domain since 2016.

7.4 Validation of DNSSEC signatures

DNSSEC can protect against attacks only if resolvers actually validate DNSSEC signatures (see section 2). It's therefore encouraging to note that, while the number of signed domain names remained broadly steady, the number of queries received from validating resolvers rose significantly last year, as illustrated in figure 7.5.

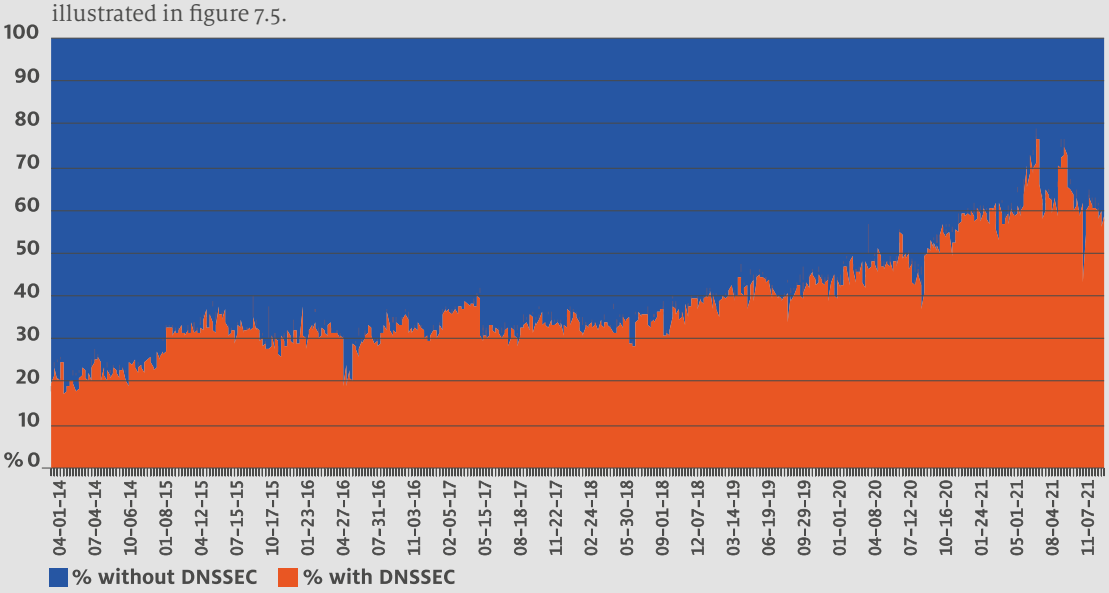


Figure 7.5 | DNS queries received from DNSSEC-validating resolvers.

Rising validation rates

In 2019, only 44 per cent of incoming query traffic came from validating resolvers. During the course of 2021, however, the figure reached 58 per cent. When a validating resolver receives a signed DNS response, it knows that the information is not false. One reason for the growth of validation is that Dutch internet access provider KPN enabled DNSSEC validation for its fixed-line and mobile customers at the start of 2020.

Problem-free technology

An argument that's often used against DNSSEC validation is that it prevents internet users reaching websites whose domain names are misconfigured for DNSSEC. By doing so, validation is liable to create more problems than it solves, the argument goes. However, as figure 7.6 shows, the number of domain names with DNSSEC configuration errors that might cause problems for resolvers is lower than ever. The rationale for enabling DNSSEC validation is therefore compelling.

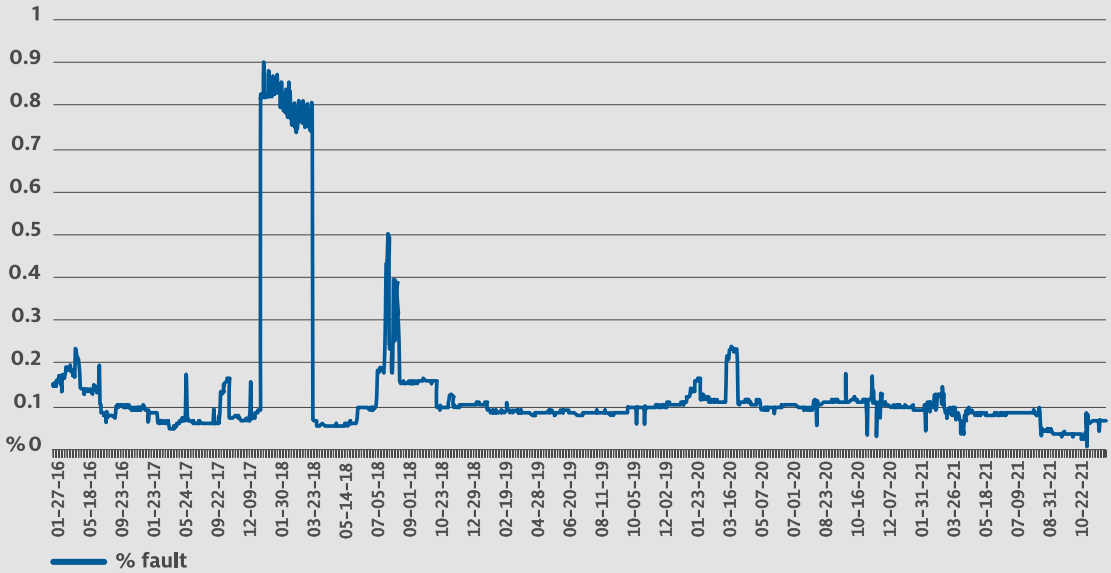


Figure 7.6 | Domain names with DNSSEC configuration errors. The percentage of all DNSSEC-enabled .nl domain names is plotted on the y-axis.

7.5 Adoption of DANE

Figure 7.7 shows that the number of .nl domain names with DANE-secured mail servers (see section 2) has risen rapidly in recent years. DANE is a good example of DNSSEC's added value as a security enabler in other parts of the internet and daily communication.

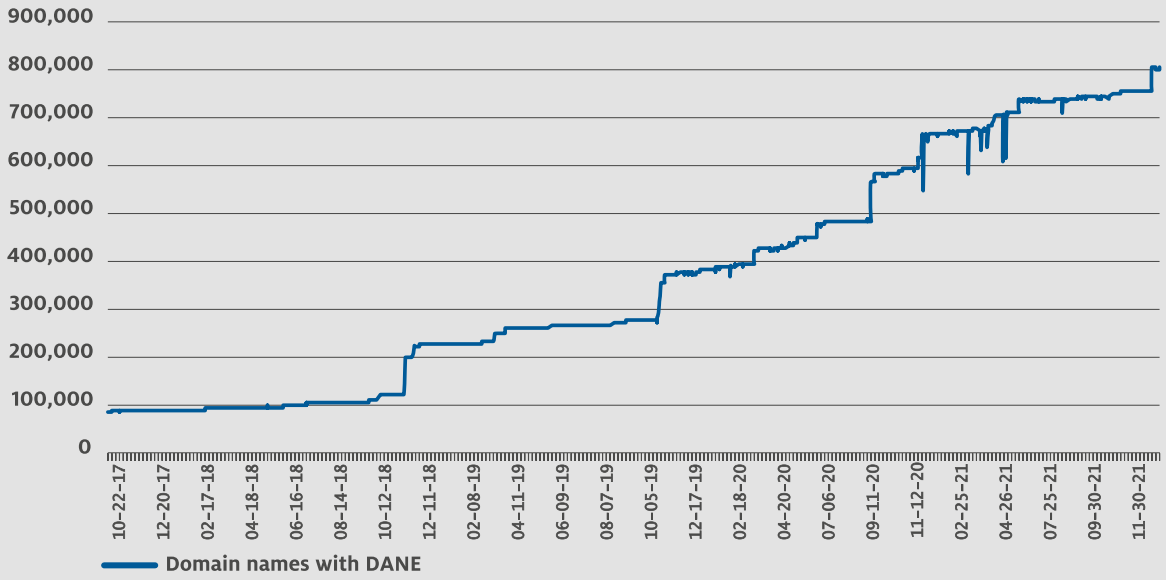


Figure 7.7 | .nl domain names with DANE-secured mail servers.

7.6 Measurement methods

We grouped websites into content-based types using data from our own DMAP crawler (subsection 7.2). We looked for things such as a Trade Register number or a notice saying that a domain is parked (temporarily or otherwise).

Sector classifications (subsection 7.2) are based on the Chamber of Commerce's standardised codes, which we've adopted as a DMAP dataset attribute.

Data on the DNSSEC algorithms used (subsection 7.3) is derived from the zones' own DNS information, as is the information about the use of DANE (subsection 7.5).

In order to establish whether a resolver validates DNSSEC signatures (subsection 7.4), we observe whether it actually requests DNSSEC records (e.g. signatures and public keys). The method is not 100 per cent reliable, but gives us a good general picture of resolver behaviour.

08

IPv6

As indicated in section 2 of this report, there are no IPv4 addresses left. Continued growth of the internet therefore depends on the widespread adoption of IPv4's successor, IPv6. Our measurements provide a good impression of how the transition to IPv6 is progressing, particularly within the .nl domain.

08

IPv6

8.1 DNS-based adoption indicators

For example, figure 8.1 illustrates the extent to which resolvers can contact the name servers for .nl domain names using IPv6. Figure 8.2 provides similar information, but in relation to mail servers only. What the two graphs show is that support for IPv6 is slowly but surely increasing.

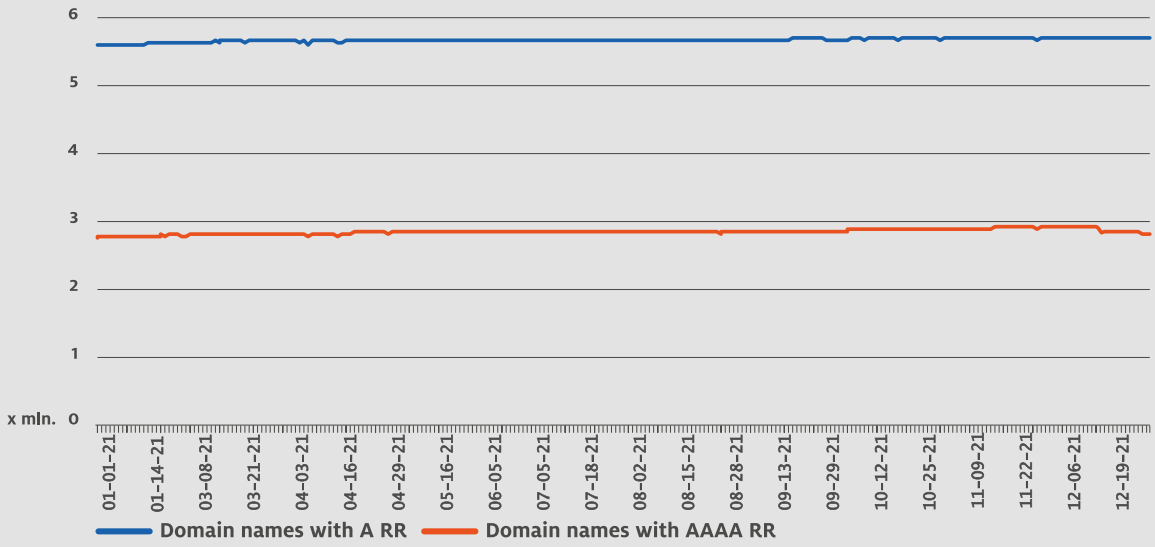


Figure 8.1 | IPv6-enabled .nl domain names.

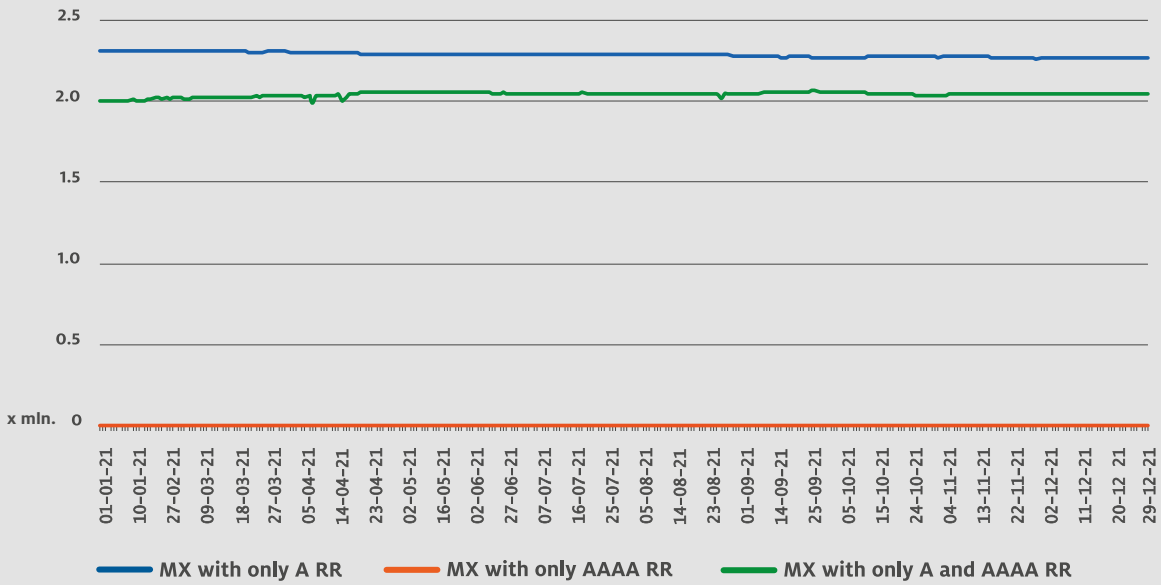


Figure 8.2 | IPv6-enabled mail servers.

Since the start of 2021, the number of domain names with AAAA records (implying support for IPv6) has increased by about 70,000. While that does mean IPv6 support has continued to grow, the rate of growth is so slow that, sadly, it's fair to say that adoption has virtually come to a halt in the Netherlands.

8.2 IPv6 address lookups

A similar pattern is apparent when we analyse IPv6 address lookups by resolvers. Figure 8.3 shows A record requests and AAAA record requests as percentages of the total number of queries about .nl domain names. An A record request is an IPv4 address lookup, while an AAAA record request is an IPv6 address lookup. The percentage of queries accounted for by IPv6 address lookups didn't increase in 2021 relative to the percentage accounted for by IPv4 address lookups.

However, the graph does highlight another interesting phenomenon in 2021, unrelated to IPv4 or IPv6: the peaks in other queries. The peaks reflect surges in query traffic from Google's public DNS resolver. For prolonged periods, the resolver was sending large numbers of requests for TXT records for apparently random domain names. The flow of queries has since stopped, but an explanation for what happened remains elusive.

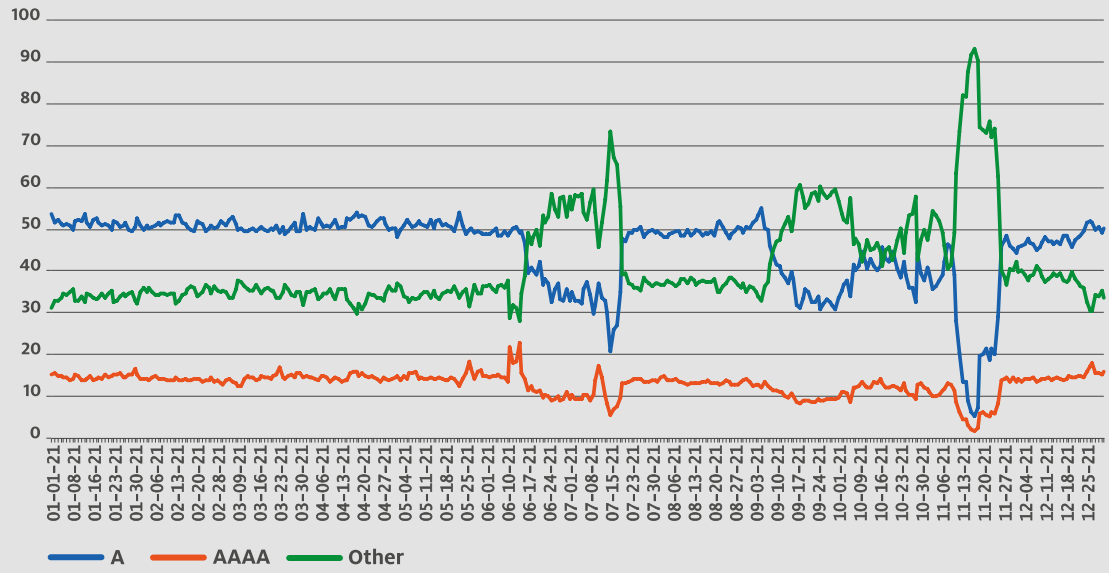


Figure 8.3 | IPv4 and IPv6 queries about .nl domain names.

8.3 Measurement methods

The statistics on domain names and mail servers (subsection 8.1) come from the OpenINTEL platform, which actively scans the .nl zone and records variables including support for IPv4 and IPv6.

The statistics on query types (subsection 8.2) come directly from our own ENTRADA platform.

Colophon

'The state of .nl' is a research report compiled by SIDN Labs and published by SIDN.

Contributors
Marco Davids, Marnie van Duijnhoven, Cristian Hesselman, Jelte Jansen, Elmer Lastdrager, Moritz Müller, Martin Sluijter, Thymen Wabeke and Maarten Wullink.

Concept and design
Lumen Ontwerpersnetwerk, Breda, The Netherlands

Translations
G & J Barker Translations, Worcester, United Kingdom
www.gandjbarker.co.uk

If you have any questions about the research, please mail sidnlabs@sidn.nl.

SIDN
Meander 501
6825 MD Arnhem
The Netherlands
PO Box 5022
6802 EA Arnhem
The Netherlands
www.sidn.nl
www.sidnlabs.nl