De staat van .nl

Inzicht in het Nederlandse internetdomein op basis van DNS-metingen

2021

SIDN LABS

Inhoud

Voorwoord

Voorwoord

Onze afhankelijkheid van het internet neemt almaar toe. Om er met een gerust hart steeds meer gebruik van te kunnen maken, moeten we erop kunnen vertrouwen dat het internet veilig is. Bij SIDN Labs onderzoeken we daarom hoe we de veiligheid van de internetinfrastructuur van onze samenleving verder kunnen verhogen. We baseren veel van ons onderzoek op grootschalige technische metingen aan het Domain Name System (DNS), zoals via de miljarden DNS-query's die SIDN dagelijks verwerkt voor .nl, metingen vanaf duizenden RIPE ATLAS-sensoren verspreid over het internet en regelmatige crawls van alle 6,2 miljoen .nl-domeinnamen.

'De staat van .nl' geeft lezers inzicht in de ontwikkelingen die we in onze .nl-data zien. We laten je kennismaken met statistische gegevens op basis van onze metingen en geven duiding aan onze doorlopende statistieken op stats.sidnlabs.nl. We kijken niet alleen naar de cijfers zelf, maar plaatsen deze ook in de bredere context van maatschappelijke ontwikkelingen.

Om alvast wat voorbeelden te geven van onze bevindingen in dit rapport:

- De gespannen situatie op de woningmarkt zie je terug in cijfers van de geregistreerde domeinnamen (hoofdstuk 4.2);
- Coronapersconferenties van de overheid en van grote bedrijven hebben direct merkbaar resultaat op domeinnaamregistraties (hoofdstuk 4.3);
- Een steeds kleinere groep van bedrijven verzorgt een steeds groter deel van de internetdienstverlening, en dit zijn niet altijd grote Amerikaanse bedrijven (hoofdstuk 6);
- Het gebruik van nieuwe technologieën voor veiligheid en stabiliteit stijgt, maar lang niet altijd even snel (hoofdstuk 7 en 8).

We hopen dat je het lezen van dit rapport net zo interessant vindt als wij het schrijven ervan vonden. Heb je na het lezen van 'De staat van .nl' vragen of feedback, neem dan vooral contact met ons op via sidnlabs@sidn.nl.

Het SIDN Labs-team

01

Inleiding

Het internet is een vast bestanddeel van ons dagelijks leven, zowel voor burgers als organisaties. We gebruiken het om contact te houden met onze familie en vrienden, voor werk, om zaken met de overheid te regelen, voor onze hobby's en voor nog veel meer. De coronapandemie maakte dat onze afhankelijkheid van het internet nog meer toenam. Zo 'Zoomen' en 'Teamsen' we iedere dag heel wat af en is online winkelen voor veel mensen gemeengoed geworden.

01

Inleiding

Onderkant van de ijsberg

En terwijl we er zo afhankelijk van zijn, staan maar weinig mensen stil bij de werking van de **infrastructuur van het internet**. Het fundament van de technische systemen (zoals routers, switches en DNS-servers) die het mogelijk maken dat internetapparaten, waar ook ter wereld, met elkaar kunnen communiceren. Dit vormt de onzichtbare 'onderkant' van de ijsberg die het internet voor de meeste mensen is en we merken vaak pas hoe afhankelijk we er van zijn als het ineens niet meer werkt.

Grootschalige metingen

SIDN Labs is de onderzoekstak van SIDN, de beheerder van het .nl-domein, en heeft als doel het verhogen van de betrouwbaarheid van de internetinfrastructuur van onze samenleving. Hiervoor doen we bijvoorbeeld veel technisch 'datagedreven' onderzoek naar de **veiligheid en stabiliteit** van de internetinfrastructuur, voor Nederland en het .nl-domein in het bijzonder. Dit doen we op basis van **grootschalige metingen**, bijvoorbeeld via de miljarden DNS-query's die SIDN dagelijks verwerkt, vanaf duizenden RIPE ATLAS-sensoren verspreid over het internet en via dagelijkse crawls van alle 6,2 miljoen .nl-domeinnamen. Een aantal van deze metingen publiceren we ook doorlopend volautomatisch op onze interactieve statistiekensite, stats.sidnlabs.nl.

Maatschappelijke ontwikkelingen

Ons doel met het rapport 'De staat van .nl' is lezers inzicht te geven in de ontwikkelingen die we in onze meetdata kunnen zien. De **aanleiding** is dat we in onze metingen soms de effecten van **maatschappelijke ontwikkelingen** zien. Om een voorbeeld te geven: we zagen in maart 2020 al de effecten van de coronapandemie in onze data. Ook zien we hoe de internetinfrastructuur evolueert om aan de steeds veranderende eisen van de maatschappij te voldoen. Denk daarbij bijvoorbeeld aan het toenemende gebruik van veiligheidsprotocollen, zoals routingsecurity of DNSSEC, die de veiligheid van de internetinfrastructuur vergroten en ervoor zorgen dat mensen en organisaties er met een gerust hart nog afhankelijker van kunnen worden.

Gebruikte datasets

Voor 'De staat van .nl' doken we in een aantal specifieke metingen die we in het afgelopen jaar uitvoerden en de analyses van de datasets die daaruit voortkwamen. Voor onze analyse gebruikten we als databron het **Domain Name System (DNS)**, het wereldwijde systeem dat domeinnamen naar IP-adressen vertaalt. De datasets die we gebruikten zijn: **ENTRADA** (passieve DNS-metingen, groeit met 2,7 miljard query's per dag), **DMAP** (website crawls, 50 miljoen metingen per maand) en **OpenINTEL** (actieve DNS-metingen, 4 miljard nieuwe datapunten per dag). ENTRADA en DMAP zijn onze eigen datasets, OpenINTEL is een gemeenschappelijke dataset van de Universiteit Twente, SURF, NLnet Labs en ons.

Domain Name System

Voordat we naar de cijfers en grafieken gaan, geven we in hoofdstuk 2 eerst een overzicht van het DNS. Het DNS vormt namelijk een van de pijlers van de internetinfrastructuur en is de bron voor de metingen die de basis vormen voor dit rapport. In dit hoofdstuk gaan we ook kort in op enkele andere technologieën, die van belang zijn voor het DNS en waarvan we verderop in het rapport cijfers bekijken. Daarnaast bespreken we in hoofdstuk 2 in meer detail de datasets die we gebruikten.

Wat je kunt verwachten

In de daaropvolgende hoofdstukken bespreken we metingen en analyses van:

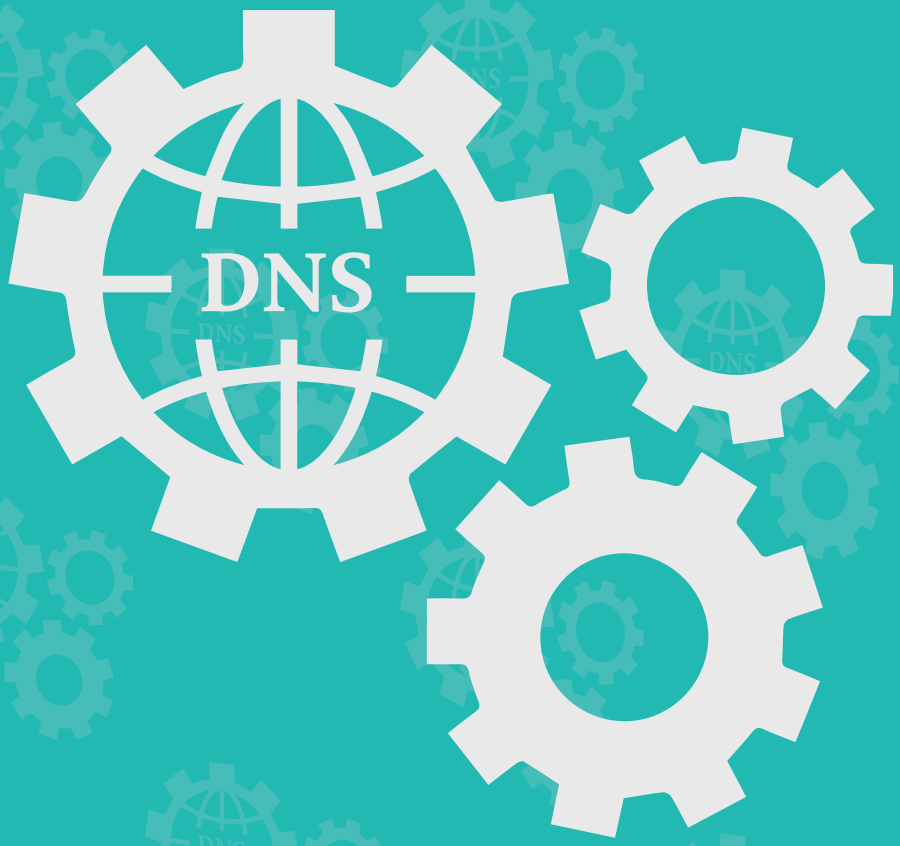
- .nl-domeinnamen, zoals actieve .nl-domeinnamen per registratiejaar en de lengte van domeinnamen (hoofdstuk 3);
- De woorden in domeinnamen die kenmerkend zijn voor 2020 en 2021, bijvoorbeeld de registratiesnelheid van coronagerelateerde domeinnamen (hoofdstuk 4);
- DNS-resolvers, de servers die DNS-informatie voor gebruikers opzoeken (hoofdstuk 5). Bijvoorbeeld de verdeling van resolvers over landen;
- Centralisatie van het internet, zoals de hoeveelheid DNS-verkeer van grote publieke DNS-resolvers als Google Public DNS (hoofdstuk 6);
- DNSSEC, bijvoorbeeld het gebruik ervan per sector en de adoptie van nieuwe DNSSEC-algoritmes in het .nl-domein (hoofdstuk 7); en
- IPv6, zoals het aantal zoekopdrachten van DNS-resolvers voor het IPv6-adres van een .nl-domeinnaam (hoofdstuk 8).

De methodologieën die we gebruikten voor onze metingen en analyses bespreken we in aparte paragrafen aan het eind van elk hoofdstuk. We hergebruiken soms grafieken van stats.sidnlabs.nl, waar we via dit rapport meer duiding aan geven.

02

Onze databron: het Domain Name System

De staat van .nl is gebaseerd op metingen aan en analyses van verschillende onderdelen van het Domain Name System (DNS), een wereldwijd systeem dat domeinnamen in enkele tientallen milliseconden naar IP-adressen vertaalt en dat onderdeel uitmaakt van de kern van de internetinfrastructuur. Het DNS werkt 'onder de motorkap', waardoor het voor de meeste gebruikers onzichtbaar is. Het komt bij bijna elke internettransactie in actie, bijvoorbeeld als je een website bezoekt of een e-mail verstuurt.



02

Onze databron: het Domain Name System

Om de grafieken in de rest van het rapport goed te kunnen begrijpen leggen we hieronder kort uit hoe het DNS op hoofdlijnen werkt. Ook gaan we nader in op de datasets die we gebruikten voor 'De staat van .nl'.

Een korte video van SIDN die de werking van het DNS uitlegt, vind je op www.supersokken.nl.

2.1 Domeinnamen

Een domeinnaam is een gemakkelijk te onthouden 'short hand' voor een IP-adres (zie § 2.2). Het is onderverdeeld in zogenaamde 'labels'. Zo bevat 'www.sidnlabs.nl' 3 labels; 'www', 'sidnlabs', en 'nl'. En binnen het DNS wordt er met de term 'domeinnaam' verder geen onderscheid gemaakt tussen een domein met veel labels zoals 'inloggen.sso.ict.oost.example.nl', of een domein met maar 1 label, zoals 'nl'.

Second-leveldomein

Als we het in het dagelijkse gesprek over een domeinnaam hebben, bedoelen we meestal een second-leveldomein: een domeinnaam met 2 labels, zoals je die registreert bij een registry als SIDN. Dus bijvoorbeeld example.nl. In dit rapport bedoelen we met de term 'domeinnaam' dan ook een second-leveldomeinnaam. Dit benoemen we verderop niet telkens expliciet.

Domeinnamen onder .nl bevatten tussen de 2 en 63 tekens (exclusief '.nl') en kan bestaan uit een combinatie van letters, cijfers en (behalve het eerste en laatste teken) een liggend streepje (-).

2.2 IP-adressen

Computers die zijn aangesloten op het internet hebben allemaal een uniek IP-adres om met elkaar te communiceren. Daar merken internetgebruikers in de regel niet veel van, want zij gebruiken meestal domeinnamen.

Er bestaan 2 smaken IP-adressen: IPv4- en IPv6-adressen. Een voorbeeld van een IP-versie 4-adres is `203.0.113.80`. IPv4 is een veelgebruikte, oude versie van het protocol. Een IPv6-adres heeft een ander format en ziet eruit als `2001:db8::7974:80`.

Vrijwel alle bestaande IPv4-adressen zijn inmiddels in gebruik, of gereserveerd voor toekomstig gebruik. Dankzij IPv6 zijn er veel meer IP-adressen beschikbaar. Omdat IPv4 en IPv6 een andere vorm van adressering gebruiken, zijn deze 2 varianten niet compatibel met elkaar. Iemand die alleen een IPv4-adres heeft, kan daarom niet communiceren met iemand die alleen een IPv6-adres heeft. Omdat IPv4 niet meer voor iedereen beschikbaar is, is het van groot belang dat iedereen IPv6 kan gebruiken.

2.3 Van domeinnamen naar IP-adressen

De taak van het DNS is om een domeinnaam zoals example.nl te vertalen naar een IP-adres. Op deze manier hoef je het IP-adres niet te onthouden, maar de makkelijker te onthouden domeinnaam.

Eigenlijk is het DNS dus een soort 'telefoonboek' voor internetadressen. Maar dat is niet het enige; het DNS kan ook informatie geven over bepaalde beveiligingsprotocollen. Denk daarbij bijvoorbeeld aan

e-mailbeveiliging, zoals welke e-mailservers er namens een bepaalde domeinnaam berichten mogen versturen.

Het daadwerkelijk gebruik van het DNS is een samenspel tussen 3 verschillende onderdelen: applicaties zoals webbrowsers en apps die gegevens van het DNS nodig hebben, zogenaamde DNS-resolvers die de informatie voor de applicaties kunnen opzoeken, en de autoritatieve DNS-servers die de informatie kunnen verstrekken. Dit samenspel kunnen we het best beschrijven aan de hand van een voorbeeld.

Een voorbeeld: de zoektocht naar het IP-adres van www.example.nl

Stel dat je in je browser (Chrome, Edge, Safari, etc.) voor het eerst naar www.example.nl gaat. Je browser vraagt dan eerst aan je besturingssysteem (Windows, MacOS, etc.) om het IP-adres achter deze website te vinden, zodat het de content van de website op kan halen. Het besturingssysteem neemt op zijn beurt contact op met de DNS-resolver.

De DNS-resolver gaat vervolgens voor jou achterhalen wat het IP-adres van www.example.nl is. Hij begint hierbij achteraan in de domeinnaam. Dus in ons voorbeeld zoekt de resolver contact met de DNS-rootservers, om te vragen waar 'nl' gevonden kan worden. De rootserver vertelt de resolver dat hij 'nl' kan vinden bij SIDN. Vervolgens gaat de resolver naar de DNS-server van SIDN om te vragen waar 'example.nl' staat. De laatste stap is dat de resolver verbinding maakt met de DNS-nameserver die bij example.nl hoort. De resolver vraagt hier het IP-adres van 'www.example.nl' op, wat `2a00:d78:0:712:94:198:159:35` is voor IPv6 en `94.198.159.35` voor IPv4. Met dit adres kan jouw browser de content van de website ophalen en aan je laten zien.

De DNS-resolvers die bovenstaande stappen uitvoeren, kunnen bij internetserviceproviders staan, maar het kunnen ook publieke resolvers zoals Quad9 of Google Public DNS zijn. In het eerste geval heeft jouw internetprovider de resolver vaak automatisch voor jou ingesteld.

E-mailprogramma's (bijvoorbeeld Outlook) gebruiken ook DNS-resolvers, maar dan om het IP-adres van mailservers op te zoeken. Het resultaat bestaat uit een of meer mailservers die het e-mailprogramma gebruikt om mail te versturen naar de betreffende domeinnaam.

De nameservers van .nl

SIDN beheert de autoritatieve nameservers van de .nl-zone. De zone bevat zogenaamde second-leveldomeinnamen (bijv. example.nl), de verwijzingen naar het volgende niveau voor resolvers (bijvoorbeeld www.example.nl) en cryptografisch materiaal voor DNSSEC (zie § 2.4). De servers van .nl staan op enkele tientallen plaatsen op de wereld en gebruiken een techniek genaamd 'anycast' om hun beschikbaarheid te maximaliseren. Hierdoor komen verschillende resolvers bij verschillende nameservers uit, afhankelijk van de netwerklocatie van de resolvers.
Onze rol als operator van het .nl-domein maakt dat wij via de .nl-nameservers toegang hebben tot de berichten die resolvers naar ons sturen om het IP-adres van domeinnamen op te zoeken.
De autoritatieve servers van SIDN verwerken dagelijks zo'n 2,8 miljard DNS-query's (peildatum 6 september 2021).

Maar wij zien niet alle individuele bevragingen. DNS-resolvers bewaren namelijk de antwoorden die ze van de .nl-nameservers krijgen een bepaalde tijd (caching). Het doel hiervan is de enorme aantallen bevragingen van internetgebruikers aan te kunnen en hen sneller het IP-adres te kunnen sturen.
De resolver hoeft dan namelijk niet nog een keer contact op te nemen met de autoritatieve server.
Hierdoor zien wij slechts momentopnames van al het DNS-verkeer, maar wel vanaf het hele internet.

Het verkeer tussen een computer en een website zien we nooit, want dit gaat buiten ons om nadat het DNS dan zijn zoekwerk heeft gedaan.

2.4 DNSSEC en DANE

Het DNS bestaat inmiddels 33 jaar en is niet met veiligheid in gedachten ontworpen. Daarom is het voor een aanvaller bijvoorbeeld mogelijk om antwoorden in het DNS te manipuleren. Aanvallers 'injecteren' hiervoor een malafide antwoord in de cache van een resolver. Dit noemen we een 'DNS cache poisoning'-aanval; de aanvaller 'vergiftigt' het geheugen (cache) van de resolver met vervalste gegevens.

Hierdoor verstrekt de resolver het verkeerde IP-adres en komen bezoekers ongemerkt uit bij bijvoorbeeld een malafide website die misschien heel sterk lijkt op de site die ze zochten. Er worden nog met enige regelmaat nieuwe manieren gevonden om dergelijke aanvallen te gebruiken.

DNSSEC

Gelukkig kunnen we dit soort aanvallen ontdekken met de hulp van de DNS Security Extensions (DNSSEC). Met DNSSEC kunnen domeinnaamhouders informatie die gekoppeld is aan hun domeinnamen, zoals het IP-adres van hun webserver, van een cryptografische handtekening voorzien. Resolvers controleren die handtekening (valideren) als ze een antwoord van een autoritatieve server ontvangen en geven het IP-adres pas door aan de applicatie als de handtekening klopt. Zo weten ze dat het antwoord van de autoritatieve servers ongewijzigd is en kunnen ze er zeker van zijn dat ze het kunnen vertrouwen. Dit proces verloopt natuurlijk volledig geautomatiseerd en razendsnel.

DANE

Het DNS bevat niet alleen informatie over IP-adressen van webservers of de namen van mailservers, maar kan in principe ook willekeurige informatie bevatten. DANE (DNS-based Authentication of Named Entities) maakt hier gebruik van om de communicatie tussen e-mailservers beter te beschermen. Een uitbreiding die zonder DNSSEC niet mogelijk zou zijn.

Standaard communiceren e-mailservers onderling onversleuteld, met als gevolg dat e-mails van derden gelezen of gemanipuleerd kunnen worden. Wel hebben mailservers de mogelijkheid om een versleutelde verbinding aan te vragen, maar een aanvaller kan deze aanvraag onderscheppen waardoor de communicatie alsnog onversleuteld is. DANE garandeert dat de verbinding versleuteld is, onder andere omdat het gebruik maakt van de beveiliging die DNSSEC biedt. Bekijk ook sidn.nl voor meer informatie over DNSSEC en DANE.

2.5 Datasets

Voor de analyses in de hoofdstukken 3 t/m 8 hebben we gebruikgemaakt van verschillende datasets. Natuurlijk is de .nl-zone zelf een belangrijke databron. Daarin kunnen we niet alleen zien welke domeinnamen er geregistreerd zijn, maar onder andere ook of een domeinnaam ondertekend is met DNSSEC. Daarnaast gebruiken we de datasets in tabel 2.1.

Dataset	Aantal meetpunten	Aantal datapunten	Frequentie
ENTRADA	24 .nl-nameservers, verspreid over de wereld	2.7 miljard query's per dag	continu
DMAP	6.2 miljoen .nl-domeinnamen	50 miljoen metingen per maand	maandelijks
OpenIntel	236 miljoen domeinnamen van o.m. .nl, .se, .com en .us	4 miljard per dag	dagelijks

Tabel 2.1 | Overzicht databronnen.

ENTRADA

Data over resolvers komt uit ons DNS-dataplatform ENTRADA. Hier slaan we de DNS-query's en -antwoorden op die we verwerken op de .nl-nameservers, voor onderzoek naar de veiligheid en stabiliteit van het internet. Om te bepalen of een resolver DNSSEC-handtekeningen valideert, meten we of een resolver ook daadwerkelijk DNSSEC-records (zoals handtekeningen en publieke sleutels) opvraagt. Deze methode is niet 100% nauwkeurig maar geeft ons wel een goede inschatting over het resolvergedrag.

DMAP

Om een website te categoriseren op inhoud of type verzamelen we data met onze crawler DMAP. Die loopt maandelijks automatisch alle .nl-domeinnamen langs. Zo kunnen we onder andere onderzoeken wat voor soort website aan een domeinnaam is gekoppeld.

OpenINTEL

Ten slotte werken we ook met data van OpenINTEL. Dit platform verzamelt dagelijks DNS-informatie van meer dan 236 miljoen domeinnamen, waaronder alle .nl-domeinnamen. Data over de gebruikte DNSSEC-algoritmen en het gebruik van DANE komt ook uit OpenINTEL. OpenINTEL is een samenwerking van SIDN Labs, NLnet Labs, SURF en de Universiteit Twente.

03

.nl-domeinnamen

We beginnen met een analyse van de informatie in de .nl-zone zelf: de .nl-domeinnamen zoals we die namens de domeinnaamhouders publiceren via de autoritatieve nameservers van .nl (zie hoofdstuk 2). We kijken naar het aantal .nl-domeinnamen, de lengte van domeinnamen, hoe lang .nl-domeinnamen bestaan, en de regio's waar ze geregistreerd worden.

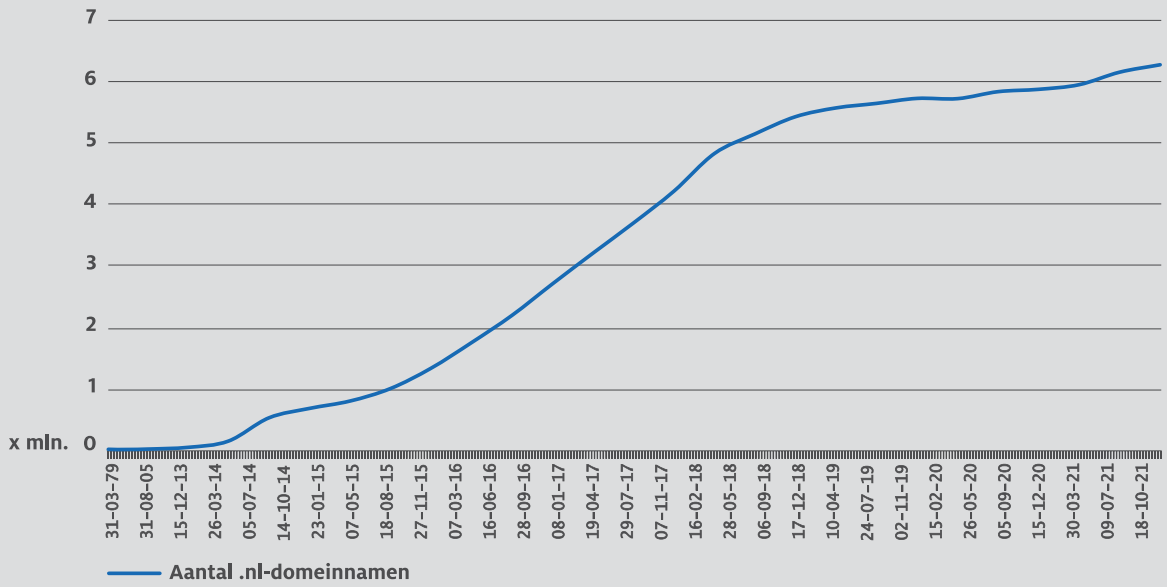
03

.nl-domeinnamen

3.1 Aantal registraties

Op 1 januari 2022 stond de teller van het aantal .nl-domeinnamen op 6.229.639 domeinnamen. Voor een relatief klein land als Nederland is dit aantal erg hoog: het staat gelijk aan 1 .nl-domeinnaam per 3 Nederlanders. Binnen Europa hebben alleen Duitsland (.de) en het Verenigd Koninkrijk (.uk) met respectievelijk 17.110.294 en 11.107.255 meer domeinnamen onder hun topleveldomein (stand per ultimo 2021, bron: CENTR. Wereldwijd staat .nl op de 5e plaats van grootste landendomeinen.

Waar in de beginjaren het beheer van .nl nog handmatig geschiedde, moest dat al snel geautomatiseerd worden om de groei bij te houden (zie ook Geschiedenis SIDN of beluister de podcast 'Het verhaal van .nl'). Vanaf eind jaren '90 tot 2014 groeide .nl flink, zoals in figuur 3.1 te zien is. Vanaf 2014 vlakte de groei in domeinnamen af, ook al bleef .nl gestaag verder groeien. Op 18 juni 2020 werd de 6 miljoenste domeinnaam geregistreerd.

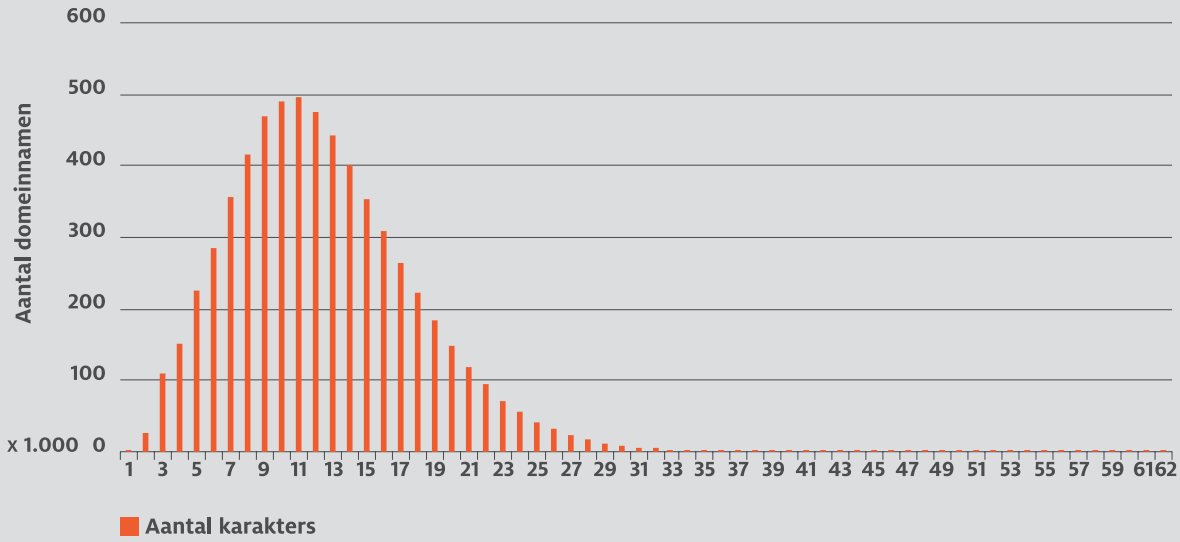


Figuur 3.1 | Overzicht aantal domeinnamen onder .nl door de jaren heen.

3.2 Lengte van .nl-domeinnamen

Domeinnamen met weinig tekens zijn relatief populair in .nl. Zo zijn er 1.296 mogelijke domeinnamen met 2 karakters ('.nl' zelf niet meegerekend) en deze zijn allemaal allang geregistreerd. Aan de andere kant van het spectrum zijn er 27 .nl-domeinnamen met het maximumaantal karakters geregistreerd (63), bijvoorbeeld: inhetverledenbehaalderesultatenbiedengeengarantievoordetoekomst.nl. Een paar jaar terug waren dit er nog 23 (Zie ook: Wat is de langste domeinnaam?).

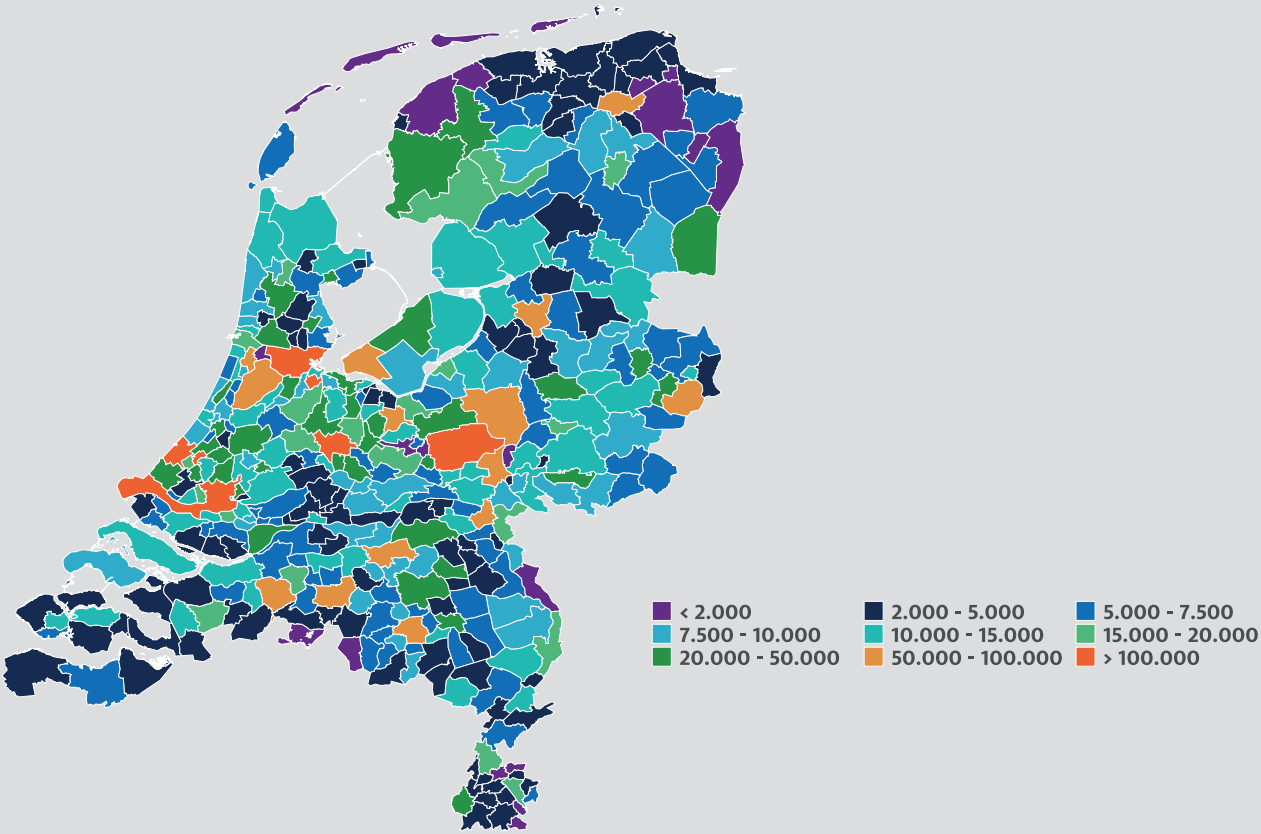
De meest voorkomende lengte van een .nl-domeinnaam is 12 karakters, waarvan er bijna een half miljoen zijn. De gemiddelde lengte van een .nl-domeinnaam is 13 karakters. In figuur 3.2 laten we voor elke mogelijke lengte zien hoeveel .nl-domeinnamen met dat aantal karakters geregistreerd zijn.



Figuur 3.2 | Aantal geregistreerde .nl-domeinnamen per aantal karakters.

3.3 .nl-domeinnamen per regio

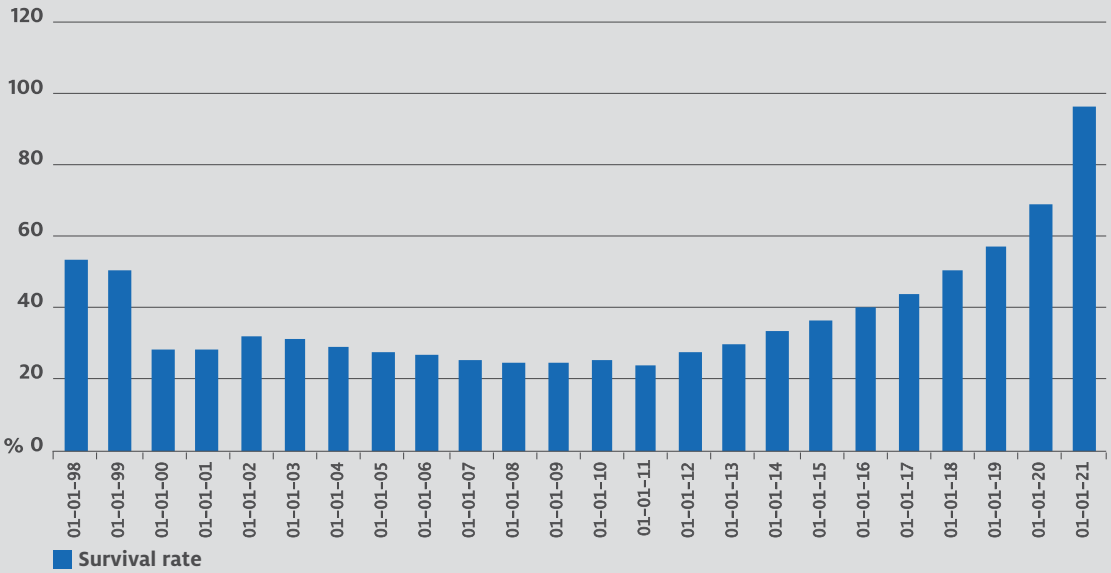
Als we kijken naar de steden waar domeinnamen geregistreerd zijn zien we ook flinke verschillen. Zo kunnen we bijvoorbeeld aan de postcodes in de registratiedata zien dat de meeste .nl-domeinnamen in Amsterdam zijn geregistreerd. Ook de grote (universiteits)steden zoals Rotterdam, Utrecht, en Wageningen zijn goed vertegenwoordigd. Figuur 3.3 laat een kaart van Nederland zien, met daarop ingekleurd hoeveel .nl-domeinnamen elke gemeente heeft. Een interactieve versie van deze kaart vind je op stats.sidnlabs.nl.



Figuur 3.3 | Aantal geregistreerde .nl-domeinnamen per Nederlandse gemeente, op basis van postcode.

3.4 Leeftijd van een .nl-domeinnaam

Ieder jaar worden er duizenden nieuwe domeinnamen geregistreerd. Een groot deel daarvan wordt een bepaalde periode gebruikt en daarna weer opgeheven. Er bestaat echter ook een categorie .nl-domeinnamen die langdurig geregistreerd blijft. In figuur 3.4 zien we welk percentage van de in een bepaald jaar geregistreerde domeinnamen eind 2021 nog steeds geregistreerd is.



Figuur 3.4 | Percentage nog actieve .nl-domeinnamen per registratiejaar per eind 2021.

De historische registratiegegevens waarover we beschikken gaan terug tot 1998. (Van de voorgaande periode is dit niet geadministreerd.) We zien dat in dat jaar en in 1999 geregistreerde domeinnamen, langer actief blijven dan in de periode tot ongeveer 2018. Dit is waarschijnlijk te verklaren doordat in die eerste paar jaar nog niet zoveel .nl-domeinnamen waren geregistreerd. Hierdoor was het makkelijker om een langdurig waardevolle domeinnaam te claimen.

Omdat we in deze grafiek kijken naar welke domeinnamen op dit moment nog actief zijn, is het percentage van domeinnamen dat in de afgelopen 5 jaar werd geregistreerd nog relatief hoog. Hier zitten immers ook nog domeinnamen tussen die misschien niet lang geregistreerd blijven, maar nog niet zijn opgeheven. Onze verwachting is dat over 5 jaar het aantal actieve domeinnamen uit deze periode flink is afgenomen, bijvoorbeeld omdat ze verwijzen naar concepten die dan niet langer actueel zijn.

3.5 Meetmethodes

Voor § 3.1 en § 3.4 hebben we de historische data over de .nl-zone gebruikt die we bij SIDN bijhouden.

Voor het bepalen van het aantal geregistreerde domeinnamen, en de lengte ervan, hebben we simpelweg de domeinnamen in de .nl-zone geteld.

De regiogegevens zijn gebaseerd op de postcodes in de registratie-informatie, zoals ze door registrars worden aangeleverd.

04

Woorden in nieuwe domeinnamen

In dit hoofdstuk onderzoeken we de populariteit van woorden die in nieuwe domeinnaamregistraties worden gebruikt. Worden er ineens veel registraties gedaan rondom een bepaald thema? Valt daaruit op te maken wat er in de Nederlandse maatschappij leeft? Zijn er nog andere ontwikkelingen geweest die terug te zien zijn in de registraties van domeinnamen? We richten ons hierbij vooral op de jaren 2020 en 2021.

04

Woorden in nieuwe domeinnamen

4.1 Het jaar 2021 in nieuwe registraties

De woordwolk in figuur 4.1 laat de woorden zien die voorkwamen in domeinnamen die in 2021 voor het eerst zijn geregistreerd. Hoe groter een woord is afgebeeld, hoe vaker dit woord voorkomt in domeinnamen die in 2021 geregistreerd werden. Op deze manier krijgen we inzicht in de populariteit van bepaalde woorden.

Figuur 4.1 | De meest populaire woorden in .nl-registraties in 2021.

De woorden ‘online’, ‘shop’, ‘straat’, ‘test’, ‘zorg’, ‘studio’ en ‘huis’ zijn het grootst afgebeeld in de woordwolk. Dit waren dus de meest populaire termen om te gebruiken in een nieuwe domeinnaam. Verderop bekijken we de populaire woorden ook per maand en gaan we iets dieper in op de vraag waarom deze woorden zo goed scoren.

4.2 Woorden die kenmerkend zijn voor 2021

De woordwolken geven aan hoe vaak een woord voorkomt, maar welke termen zijn kenmerkend voor 2021? Kunnen we deze kenmerkende woorden relateren aan zaken die spelen in de samenleving?

In tabel 4.1 zien we het lijstje van de top 10 van meest opvallende woorden. We hebben het samengesteld door het aantal keer dat een woord voorkomt in 2021 geregistreerde domeinnamen te vergelijken met het aantal keer dat een woord voorkomt in alle .nl-domeinnamen die in 2021 actief waren. Hierdoor verdwijnen termen die altijd al populair zijn naar de achtergrond, terwijl woorden die kenmerkend zijn voor 2021 eruit springen. Zie § 4.7 voor meer informatie over de berekening van opvallende woorden.

Positie	Woord	JLH-score	Aantal in 2021	Aantal in alle .nl-domeinnamen	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%

Tabel 4.1 | Opvallende woorden in 2021.

Trends in de samenleving zien we in tabel 4.1 duidelijk terug. De woorden ‘straat’ en ‘laan’ zijn bijvoorbeeld populair. We hebben deze domeinnamen nader onderzocht en kwamen erachter dat het hier in bijna alle gevallen om een domeinnaam gaat die een adres weergeeft. Dit is een populaire manier gebleken om huizen te verkopen. Zeker in het hogere segment wordt er voor de verkoop van een huis tegenwoordig vaak een eigen domeinnaam geregistreerd. De domeinnaam wordt gebruikt om er een website op te zetten met informatie over het huis.

We zien ook dat andere trends en gebeurtenissen invloed hebben. De term ‘meta’ komt bijvoorbeeld vaker voor sinds Facebook aankondigde de naam van haar holding te veranderen in Meta. Daarnaast is de populariteit van ‘crypto’ en de sport ‘padel’ duidelijk zichtbaar. Respectievelijk 37% en 56% van de domeinnamen met deze woorden erin, werd voor het eerst geregistreerd in 2021.

Ten slotte heeft COVID-19 natuurlijk ook een grote impact op domeinnaamregistraties. We zagen veel domeinnamen met de term ‘test’, hoewel we die niet allemaal aan COVID-19 kunnen relateren. Registraties met ‘pcr’ in de domeinnaam zijn ook erg populair en deze hebben wel een duidelijke link met de pandemie.

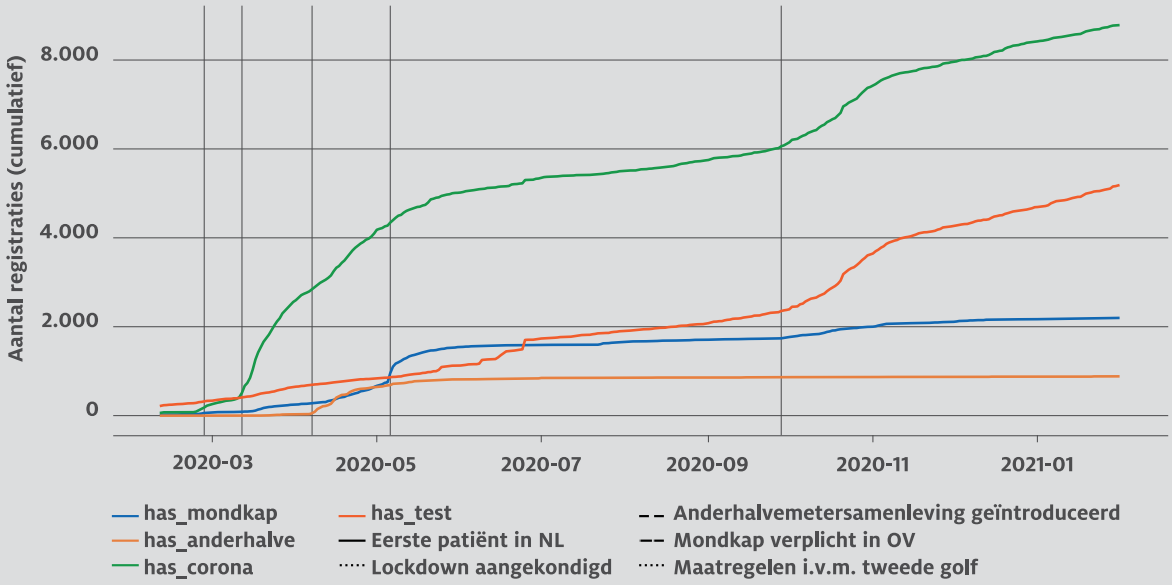
4.3 Snelheid van .nl-registraties

De vorige paragrafen lieten duidelijk zien dat thema’s die spelen in de samenleving een effect hebben op de domeinnamen die worden geregistreerd. We kunnen ook kijken naar de snelheid waarmee dit gebeurt.

Om dit te onderzoeken focussen we op het aantal registraties met woorden die te maken hebben met de coronapandemie. We kijken hierbij niet alleen naar 2021, maar nemen ook 2020 mee, zodat we het begin van de pandemie kunnen analyseren.

Figuur 4.2 laat zien wat het effect is van de coronapersconferenties op het aantal domeinnamen met een bepaald woord. De x-as toont de maanden februari 2020 tot februari 2021, waarbij de zwarte verticale lijnen een gebeurtenis aangeven, zoals de persconferentie waarbij de eerste coronabesmetting in Nederland werd aangekondigd. De y-as geeft het aantal registraties aan voor 3 verschillende COVID-19 gerelateerde woorden, namelijk ‘corona’, ‘anderhalve’ en ‘mondkap’.

Je ziet in deze figuur heel goed dat de persconferenties bijna onmiddellijk effect hadden op domeinnaamregistraties. Het aantal domeinnamen met het woord 'anderhalve' was vrij stabiel (oranje lijn). Begin april 2020 zien we een sterke toename nadat voor het eerst sprak over de 'anderhalvemeter samenleving'. Een vergelijkbaar effect zien we voor domeinnamen met het woord 'mondkap' (blauwe lijn). Het aantal registraties nam vrij stabiel toe vanaf de eerste coronabesmetting, maar explodeerde begin mei 2020 toen bekend werd dat mondkapjes verplicht werden in het openbaar vervoer. Domeinnamen met de term 'test' of 'pcr' (rode lijn) namen gestaag toe sinds het begin van de pandemie, maar we zien een sterke toename aan het begin van de tweede besmettingsgolf rond oktober 2020.



Figuur 4.2 | Snelheid van registraties na persconferenties.

4.4 Context van woorden in .nl-domeinnamen

Tot nu zoomden we in op de in .nl-domeinnamen gebruikte woorden. We kunnen ook kijken naar de woorden op webpagina's die gekoppeld zijn aan domeinnamen. Een webpagina bevat namelijk meer tekst waardoor we ook andere analyses kunnen uitvoeren.

We keken bijvoorbeeld naar de context waarin woorden worden gebruikt. Als 2 woorden dezelfde context hebben, dan wil dat zeggen dat ze in teksten vaak dicht bij elkaar staan. De woorden 'stoel' en 'tafel' worden bijvoorbeeld vaak in dezelfde zin of paragraaf gebruikt. Deze woorden delen dus dezelfde context en je kunt daarom zeggen dat ze qua betekenis aardig vergelijkbaar zijn.

De context van een woord kan natuurlijk ook veranderen. In tabel 4.2 zien we in de eerste kolom een aantal woorden, met in de tweede kolom vergelijkbare termen. We tonen hierbij de vergelijkbare woorden berekend in de zomer van 2018 en de vergelijkbare woorden berekend in het najaar van 2021.

Woord	Vergelijkbare context in 2018	Vergelijkbare context 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

Tabel 4.2 | Context van woorden.

Zoals je ziet is de context waarin 'corona' gebruikt wordt enorm veranderd. In 2018 kwam de context waarin het woord gebruikt werd sterk overeen met andere biermerken. In 2021 werd de term vooral gebruik in de context van COVID-19. Je ziet net zo'n verschuiving bij 'crisis'. De vergelijkbare woorden in 2018 hadden veelal een economische betekenis, terwijl we in 2021 woorden zien die te maken hebben met COVID-19.

4.5 Populaire woorden per maand in 2021

Hieronder zien we de woordwolken van de populairste termen in nieuw geregistreerde domeinnamen in 2020, opgesplitst per maand.

In dit overzicht valt allereerst op dat de woorden 'online' en 'shop' blijvend populair zijn. Dit is verder niet terug te leiden naar een specifieke gebeurtenis, maar kunnen we wel verklaren, omdat dit nu eenmaal standaard woorden zijn om toe te voegen aan een domeinnaam. Het is per definitie online, en 'shop' is een populair woord voor een webwinkel. We kunnen natuurlijk wel concluderen dat er in 2021 een groot aantal webwinkels is bijgekomen, maar dit is op zichzelf niet ongebruikelijk.
Het woord 'straat' viel in § 4.2 ook al op, maar we zien hier dat deze term het hele jaar door populair bleef. Datzelfde geldt, in iets mindere mate, voor 'crypto'. We zagen in § 4.2 ook het woord 'meta' al. In dit maandelijkse overzicht kunnen we zien dat dit ook echt pas na de aankondiging van Facebook naar boven kwam: pas in oktober verschijnt het woord in de wolk. Een kleiner, maar vergelijkbaar effect zagen we door het socialmediaplatform Clubhouse eerder in het jaar. In februari, maart, en april zien we ook 'club' even verschijnen als populaire term.

4.6 Populaire woorden per maand in 2020

In 2020 brachten we geen rapport als dit uit, maar we willen toch ook de techniek van deze paragraaf toepassen op dat jaar, aangezien hier wel een aantal interessante ontwikkelingen te zien zijn met betrekking tot de coronapandemie. In maart 2020 maakte het kabinet de eerste maatregelen om het coronavirus te bestrijden bekend. Hier is het besef van de ernst van de coronapandemie duidelijk terug te zien in de registraties: het woord 'corona' staat met stip op nummer 1. Ook 'thuis' staat ineens hoog in de top 10.

In april verschijnen ook de woorden 'anderhalve' en 'meter' in de woordwolk; dit is dan ook het moment geweest dat we begonnen te spreken over de anderhalvemetersamenleving.

In mei was er veel discussie over het gebruik van mondkapjes. Dit is ook de maand waarin de btw op mondkapjes werd afgeschaft, en in onze data zien we dan ook een grote stijging in het aantal websites dat mondkapjes verkoopt.

2020

Figuur 4.3 | Populaire woorden in .nl-domeinnamen in 2020.

Figuur 4.4 | Populaire woorden in .nl-domeinnamen in 2021.

4.7 Meetmethodes

Voor het bepalen van de woorden in een domeinnaam in dit hoofdstuk maken we gebruik van een woordsplitser. Om een voorbeeld te noemen: bij de domeinnaam sidnmerkbewaking.nl, die we zelf registreerden, komen de woorden 'merk' en 'bewaking' uit dit algoritme.

Om de woordwolken te maken, berekenden we voor alle in de betreffende periode geregistreerde domeinnamen de woorden. Uit het resultaat filterden we lidwoorden, voorzetsels, en bijvoeglijk naamwoorden en telden we de overgebleven woorden bij elkaar op.

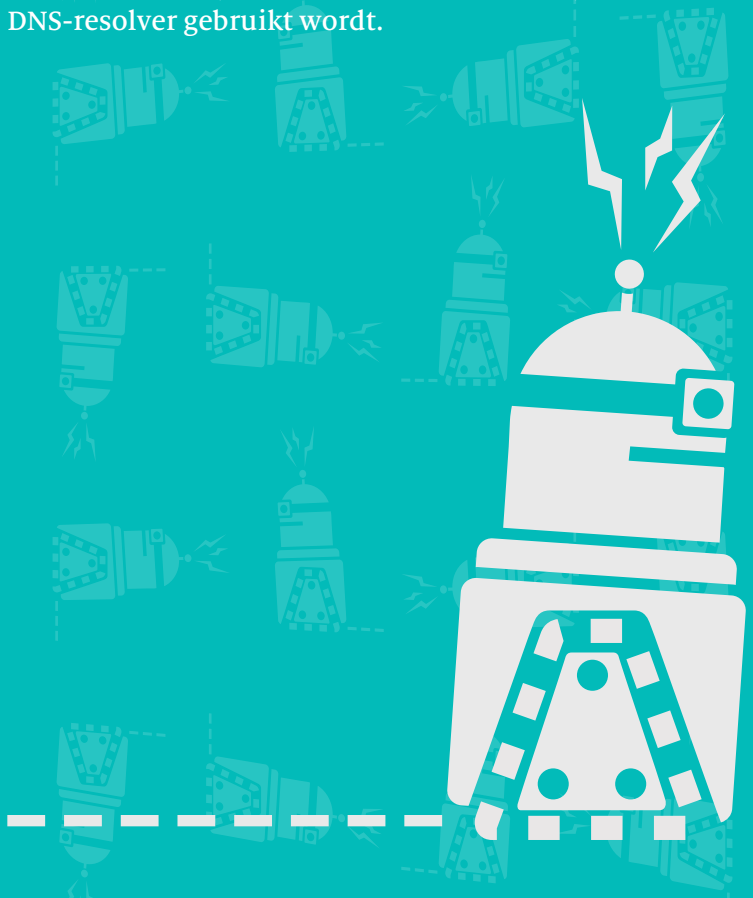
Het berekenen van de opvallende woorden doen we op basis van de JLH-score van ElasticSearch.

Om de context van termen te berekenen, onderzochten we geautomatiseerd de content van websites op domeinnamen. Daarbij maakten we gebruik van Word2vec om zogenaamde embedding-vectoren te berekenen. Dat zijn automatisch afgeleide representaties van woorden, die bestaan uit een lijst van getallen die de betekenis van dat woord omschrijven. Tussen 2 van zulke lijsten kun je de zogenaamde cosinusafstand berekenen, een waarde die aangeeft in hoeverre de lijsten van elkaar verschillen. Hoe kleiner de cosinusafstand tussen 2 lijsten is, hoe meer de 2 woorden een vergelijkbare betekenis hebben.

05

DNS-resolvers

In dit hoofdstuk kijken we naar de gegevens van de partijen die namens hun gebruikers informatie in het DNS opzoeken, oftewel de DNS-resolvers die .nl-domeinnamen opzoeken (zie hoofdstuk 2). We gaan onder andere in op wanneer we DNS-query's (DNS-zoekopdrachten) ontvangen op de .nl-nameservers, waar deze vandaan komen, en welke technologie er door de DNS-resolver gebruikt wordt.



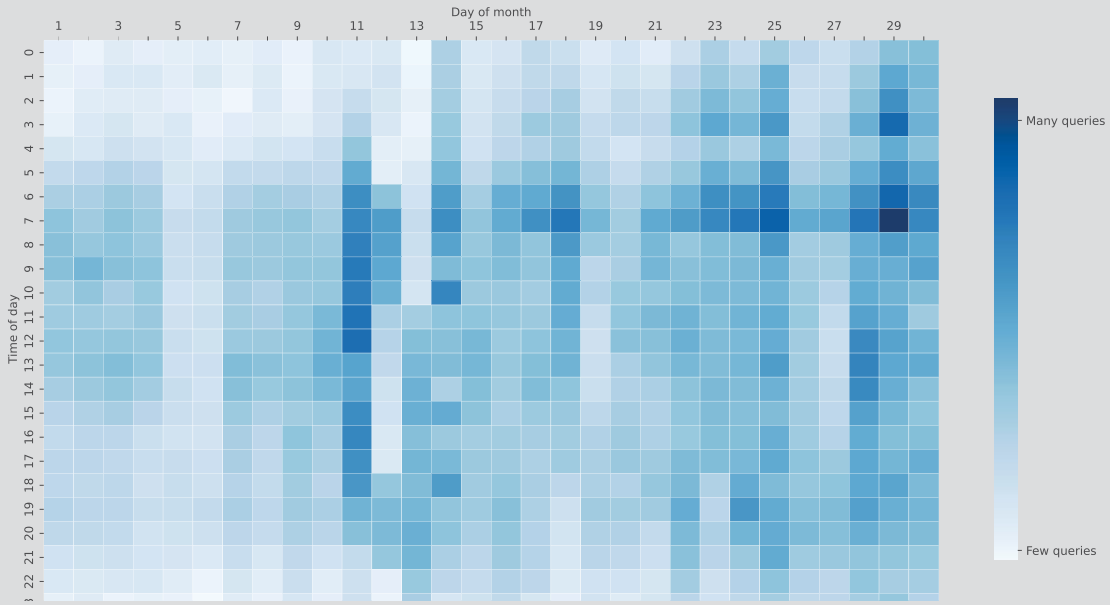
05

DNS-resolvers

5.1 Patronen in DNS-query's van resolvers

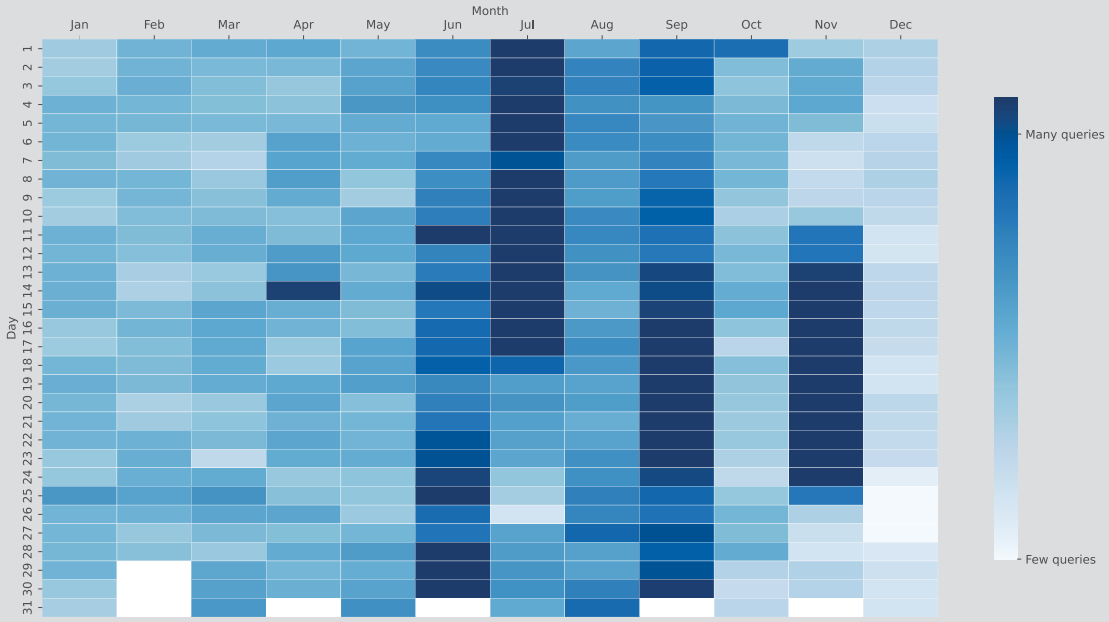
Om meer inzicht te krijgen in de querypatronen van DNS-resolvers, analyseerden we de hele maand juni in 2021 de query's die we van resolvers ontvingen. Figuur 5.1 toont het resultaat. De tijd van de dag staat verticaal (van middernacht bovenaan tot 23:59 onderaan) en de dag horizontaal (1 juni links, 30 juni rechts).

In het figuur zien we een aantal effecten. Allereerst valt op dat overdag (tussen 05.00 – 06.00 uur 's ochtends en 20.00 – 21.00 uur 's avonds) het aantal query's hoger ligt dan in de nacht. Dit heeft te maken met het dag- en nachtritme van de mensen die .nl-domeinnamen gebruiken. Ook zijn de meeste weekenden goed zichtbaar in het figuur, bijvoorbeeld zaterdag 5 en zondag 6 juni, omdat er dan minder query's zijn.



Figuur 5.1 | Aantal query's per uur van DNS-resolvers in juni 2021.

Figuur 5.2 toont het aantal query's per dag in 2021, een grafiek die je krijgt als je iets zou 'uitzoomen' in figuur 5.1. De opmerkelijkheden haal je er makkelijk uit.

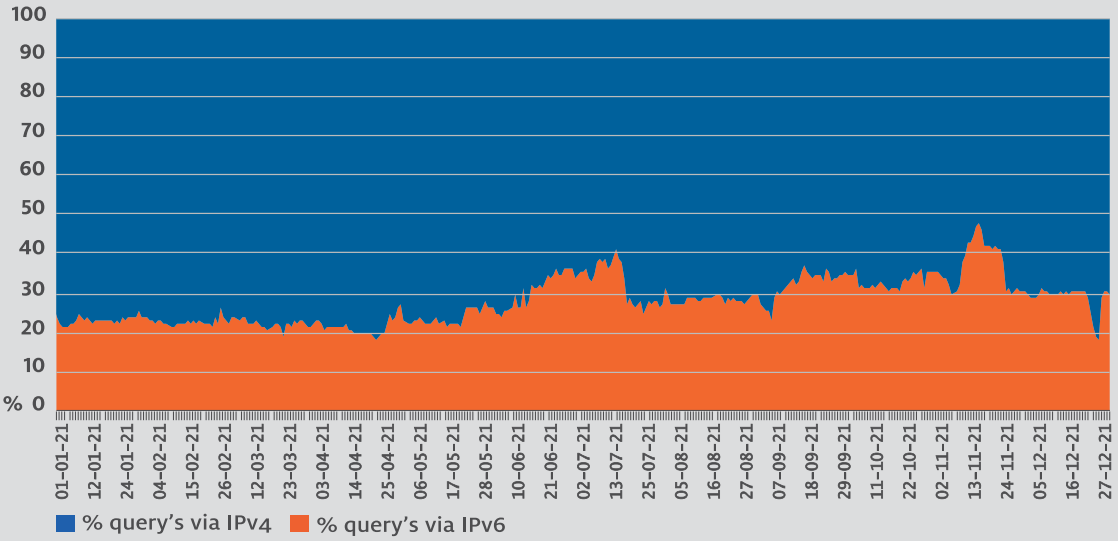


Figuur 5.2 | Aantal query's per dag van DNS-resolvers in geheel 2021.

5.2 Netwerkprotocollen: IPv4 en IPv6

Naast het opzoeken van IPv4- en IPv6-adressen (zie hoofdstuk 2), kunnen DNS-resolvers zelf ook communiceren via IPv4 of IPv6. In figuur 5.3 laten we het percentage van query's zien dat via IPv4 of IPv6 is binnengekomen op de .nl-nameservers. Dit geeft ons inzicht in de adoptie van IPv6 bij DNS-resolvers.

Er is een stijging in het percentage van IPv6 te zien, van iets boven de 20% begin 2021 naar iets onder de 30% van de query's aan het eind van het jaar. Dit is een ontwikkeling in de goede richting, omdat de pool van IPv4-adressen leeg is. Maar het stijgt nog lang niet snel genoeg om binnen afzienbare tijd te kunnen spreken van een hoge adoptiegraad. In hoofdstuk 8 kijken we nog naar andere cijfers over IPv6.

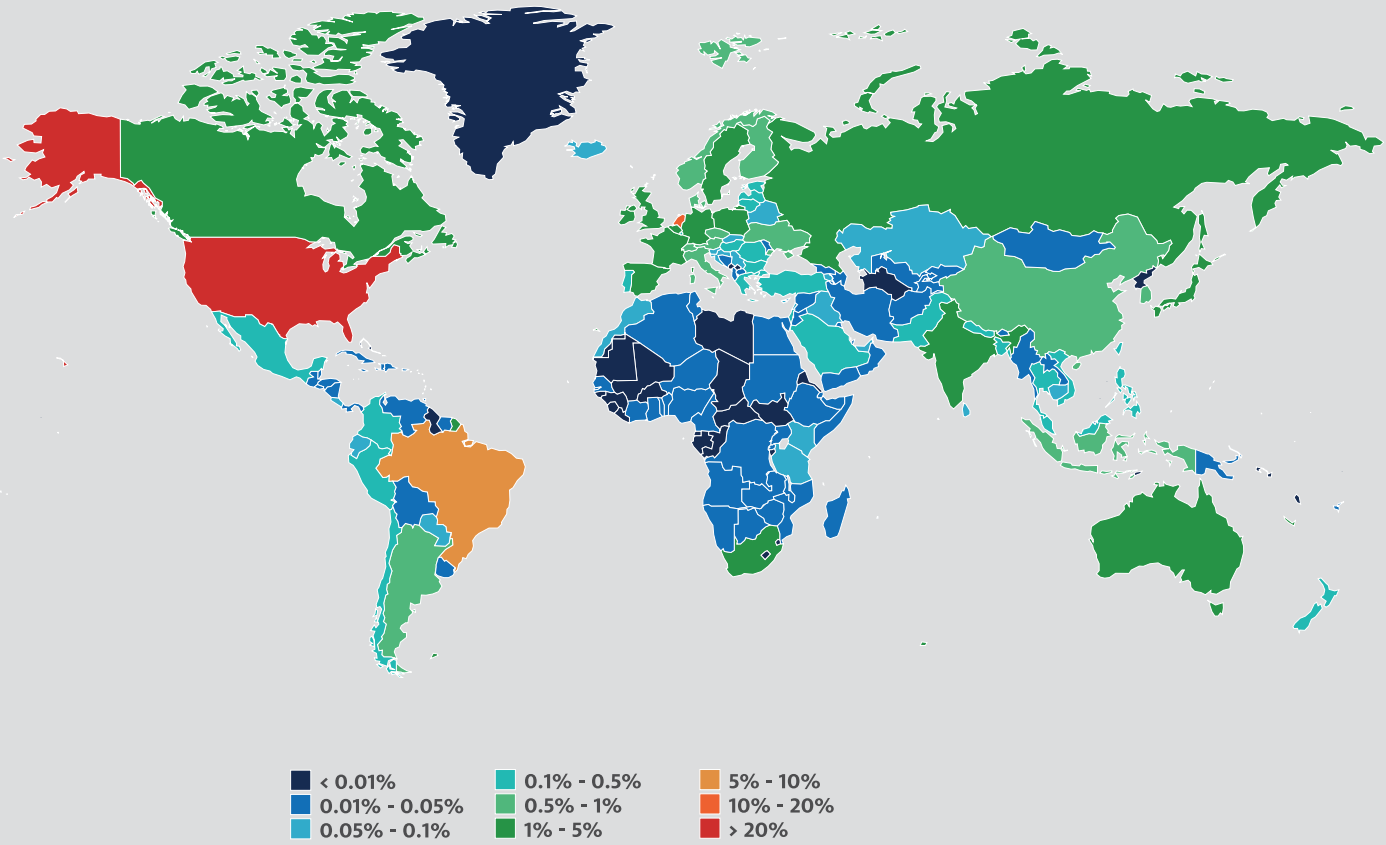


Figuur 5.3 | Percentage query's afkomstig van een resolver vanaf een IPv4- of IPv6-adres.

5.3 Resolverlocaties

Om de beschikbaarheid van .nl te maximaliseren, hebben we over de hele wereld enkele tientallen autoritatieve servers draaien om vragen van DNS-resolvers die daar 'in de buurt' staan te beantwoorden (zie hoofdstuk 2). We krijgen namelijk niet alleen vanuit Nederlandse DNS-resolvers query's binnen, ook vanuit het buitenland komen veel zoekopdrachten.

Om dit te visualiseren hebben we in figuur 5.4 de locatie van de DNS-resolvers op een wereldkaart gezet. Zoals te verwachten staan veel DNS-resolvers die query's naar ons sturen in Nederland zelf. Ook de Verenigde Staten zijn goed vertegenwoordigd, ook omdat hier veel grote DNS-diensten draaien, zoals Google Public DNS en CloudFlare DNS.

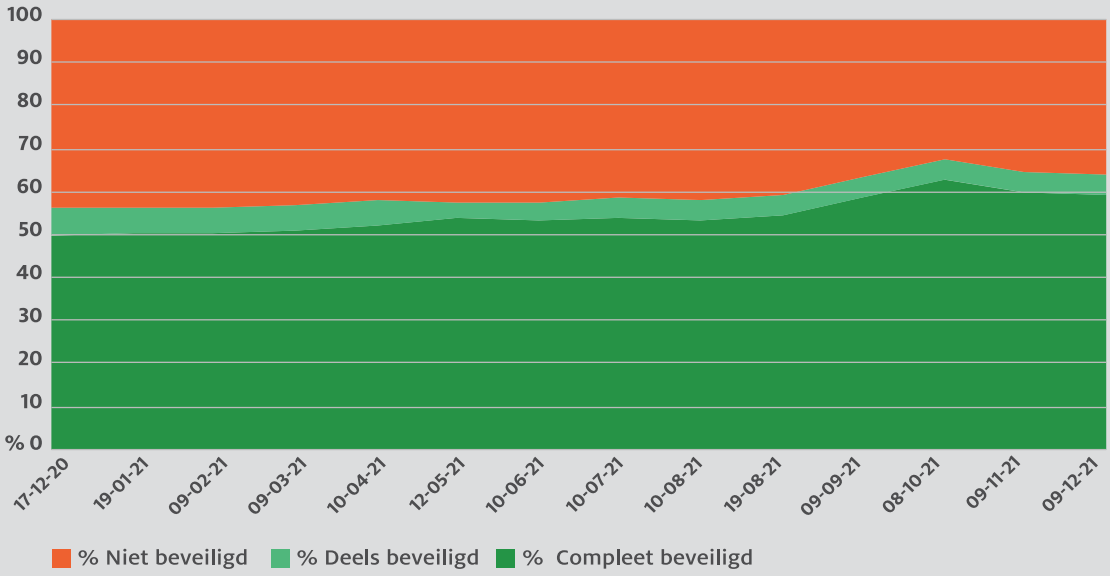


Figuur 5.4 | Locaties van de DNS-resolvers die de .nl-servers bevragen.

5.4 Routing security

Figuur 5.5 laat voor heel 2021 zien hoeveel .nl-domeinnamen wezen naar webservers die in een netwerk staan dat beveiligd is met Resource Public Key Infrastructure (RPKI). Met deze open standaard kunnen zogenaamde route hijacks (het kapen van netwerkverkeer) voorkomen worden. Een route hijack kan serieuze gevolgen hebben. Een scenario is dat DNS-vragen van een webbrowser door een hijack uitkomen bij een malafide resolver, zonder dat de browser of de gebruiker dat merkt. De malafide resolver leidt de browser vervolgens naar een malafide server, bijvoorbeeld voor een phishingaanval.

In figuur 5.5 staat het donkergroene deel voor 'compleet met RPKI beveiligd' en het lichtgroene deel voor 'deels beveiligd'. Begin 2021 is het percentage van volledig beveiligde routes de 50% gepasseerd, maar de grootste stijging lijkt inmiddels voorbij. Dat is jammer, want ook RPKI kan een belangrijke bijdrage leveren aan het vergroten van de internetveiligheid.



Figuur 5.5 | Het percentage .nl-domeinnamen dat beveiligd is met RPKI.

5.5 Meetmethodes

De meeste statistieken in dit hoofdstuk zijn cijfers die rechtstreeks uit ENTRADA komen (zie § 2.4). Het bepalen van het land van oorsprong hebben we gedaan met behulp van een geoIP-library, die voor een IP-adres of netwerk de locatie geeft. Voor de grafiek over het gebruik van RPKI hebben we de IP-adressen van .nl-websites opgezocht en vervolgens gekeken of deze adressen in een netwerk staan dat met RPKI is beveiligd.

06

Centralisatie van het internet

Centralisatie op het internet is de voortgaande ontwikkeling waarbij een steeds kleiner aantal grote internetpartijen steeds meer diensten, data en kennis naar zich toetrekt. Een voorbeeld daarvan is dat steeds meer partijen tegenwoordig gebruikmaken van grote publieke resolvers, in plaats van de resolvers van hun eigen internetserviceprovider (ISP). Dit is een veelbesproken onderwerp, want hoewel dergelijke grote partijen doorgaans heel goed zijn in wat ze doen, kan het bijvoorbeeld ook een stabiliteits- en veiligheidsrisico zijn. Een storing bij een enkele partij kan gevolgen hebben voor een relatief groot deel van het internet.

In dit hoofdstuk geven we aan de hand van onze metingen een indicatie van het potentiële risico van centralisatie voor de veiligheid en stabiliteit van de .nl-domeinen.

06

Centralisatie van het internet

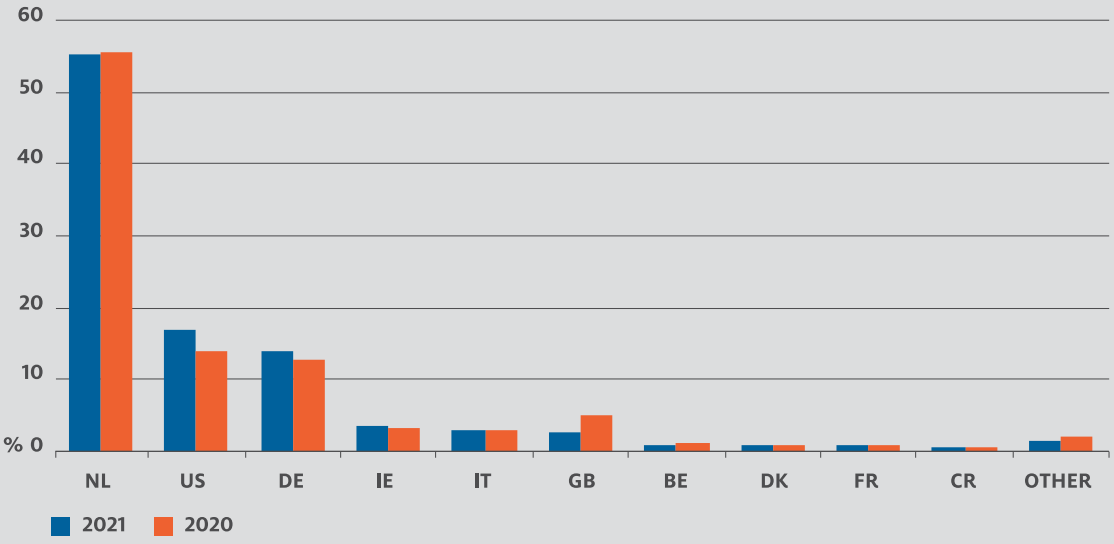
6.1 Autoritatieve .nl-nameservers

Een van de manieren om onlinediensten weerbaarder te maken is meerdere autoritatieve DNS-servers te gebruiken (zie hoofdstuk 2). Als een daarvan onbereikbaar wordt, bijvoorbeeld door een storing of een cyberaanval, dan zorgt het DNS er automatisch voor dat een andere nog wel bereikbare nameserver wordt benaderd. Maar meerdere nameservers zijn op zichzelf niet voldoende, als ze allemaal bij dezelfde provider staan. Dit was bijvoorbeeld de oorzaak van recente storingen bij de providers OHVcloud en Akamai. Hierdoor waren duizenden domeinnamen tijdelijk niet bereikbaar.

Spreiding van nameservers over landen

Figuur 6.1 laat de landen zien waarin .nl-nameservers staan, exclusief die van SIDN. Wat opvalt is dat iets meer dan de helft in Nederland zelf staat. De VS volgt met bijna 20%. Dit is vergelijkbaar met de situatie van een jaar geleden.

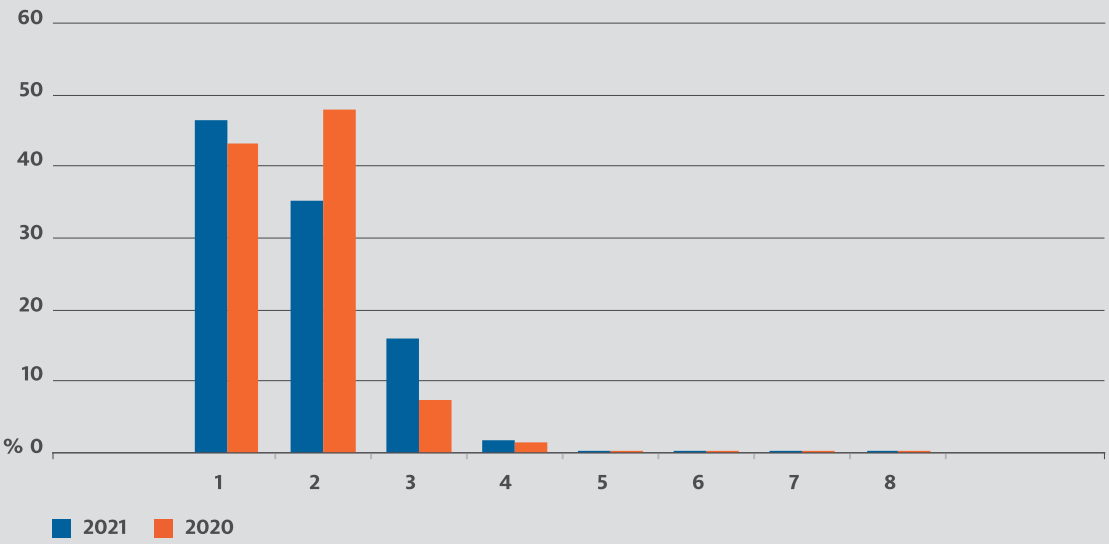
Een verspreiding over verschillende landen is verder geen belangrijke indicator van centralisatie. In de meeste gevallen is het wenselijk om nameservers dicht bij de doelgroep te plaatsen zodat ze zo snel mogelijk DNS-vragen kunnen beantwoorden. Voor .nl is dat toch doorgaans in Nederland.



Figuur 6.1 | .nl-nameserverlocaties, exclusief de nameservers van SIDN.

Spreiding van nameservers over netwerken

Wat wel een belangrijke indicator is, is de verspreiding van .nl-nameservers over netwerken. Figuur 6.2 laat zien in hoeveel verschillenden netwerken de .nl-nameservers staan. Hier valt op dat voor bijna 50% van alle .nl-domeinen al hun nameservers maar in 1 netwerk staan. Dit maakt deze domeinnamen gevoelig voor storingen. Wordt, zoals in het geval van Akamai, het netwerk of de nameserver onbereikbaar, dan wordt ook de domeinnaam onbereikbaar. Ook voor veel actief gebruikte .nl-domeinnamen zijn de nameservers niet verspreid: van 51% van de domeinnamen die worden gebruikt voor webshops, zijn de nameservers in een enkel netwerk geplaatst.



Figuur 6.2 | .nl-nameservers in verschillende netwerken.

Stijging nameservers in enkel netwerk

Je zou verwachten dat bovengenoemde grote storingen bij providers voor beheerders van domeinnamen een waarschuwing was om niet meer alleen op een enkele provider of een enkel netwerk te vertrouwen. We zien echter het tegenovergestelde: het aantal domeinnamen waarbij al de nameservers in een enkel netwerk staan, steeg in 2021 met 3 procentpunt.

Consolidatie registrarveld

Een goede verklaring hiervoor is de consolidatie die plaatsvindt in het registrarveld. De .nl-registrars zijn de partijen die domeinnamen registreren voor hun klanten en veelal ook de nameservers draaien. Het aantal registrars neemt al jaren af en een steeds groter deel van de .nl-zone wordt beheerd door enkele partijen. Daarnaast is het vaak lastig gebruik te maken van verschillende nameserverproviders. Een voorbeeld is het beveiligen van een domeinnaam met DNSSEC (zie § 2.4). Dit wordt ingewikkelder als er meerdere providers worden gebruikt. Gelukkig worden hiervoor al standaarden ontwikkeld die dit makkelijker moeten maken, maar of dat de trend stopt is twijfelachtig.

100% bereikbaarheid lastig

Een aanvullend risico is, dat een netwerk vaak de nameservers van duizenden domeinnamen host. 10 netwerken, vooral in Nederland en in de VS, hosten 64% van alle nameservers die verantwoordelijk zijn voor .nl-domeinnamen. Een jaar geleden was dat percentage nog 61%. Natuurlijk doen de beheerders hun best om te voorkomen dat hun netwerk onbereikbaar wordt, maar de voorbeelden van Akamai, OVH, maar ook Dyn laten zien, dat zelfs voor heel grote providers 100% bereikbaarheid onhaalbaar is.

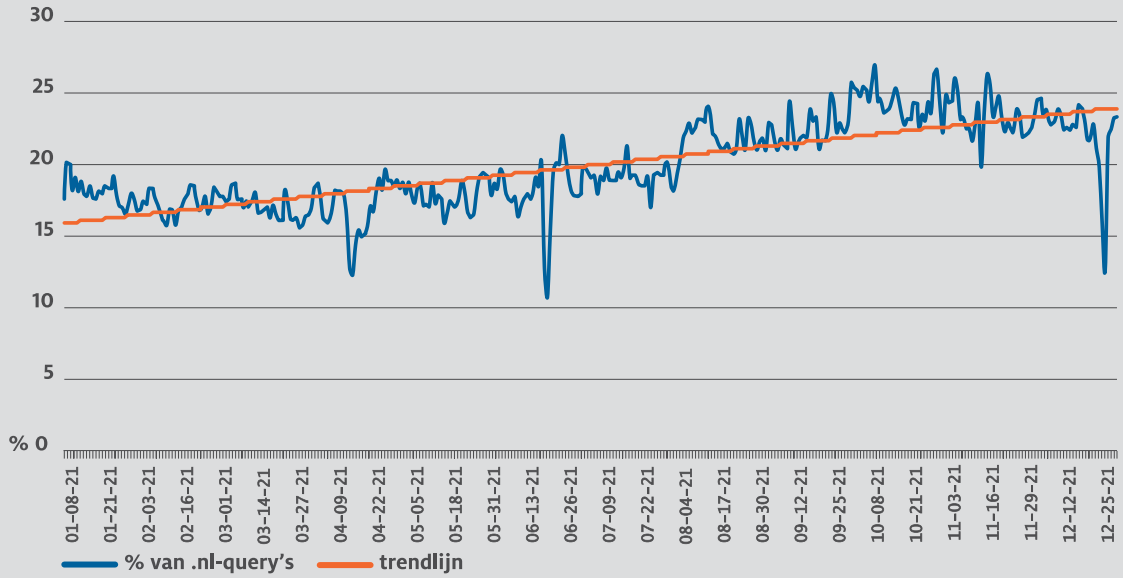
6.2 Recursieve resolvers

Recursieve resolvers (zie hoofdstuk 2) spelen een belangrijke rol in het DNS. Een grote onbereikbare recursieve resolver kan bijvoorbeeld als gevolg hebben dat veel eindgebruikers helemaal geen websites meer kunnen bereiken. Tegelijkertijd heeft een beheerder van zo'n resolver vaak diepgaand inzicht in het surfgedrag van eindgebruikers, omdat het hun DNS-query's en antwoorden afhandelt.

Publieke resolverdiensten

De centralisatie van de DNS-infrastructuur lijkt ook bij resolvers toe te nemen. Een reden hiervoor is dat er meer query's vanuit publieke resolverdiensten komen. Voorbeelden van deze diensten zijn 8.8.8.8 van Google en 1.1.1.1 van CloudFlare.

In het afgelopen jaar zagen we een sterke groei van query's vanuit deze providers. In figuur 6.3 laten we het percentage van query's van publieke resolvers zien die we afhandelden op de .nl-nameservers. De blauwe lijn is de trendlijn. Inmiddels komt ongeveer 1 op de 4 query's van een publieke resolverdienst.



Figuur 6.3 | Query's afkomstig van publieke DNS-resolvers.

Je zou misschien verwachten dat DNS-query's voor .nl-domeinnamen meestal vanuit Nederland komen, maar ook hier zien we de dominantie van grote Amerikaanse bedrijven zoals Microsoft en Google terug. 33% van alle query's in 2021 kwam uit de VS, tegen 26% vanuit Nederland. Maar we zien hier wel een teruglopende trend. Kwam aan het begin van het jaar 2021 nog rond de 35% van alle query's uit de VS, aan het einde van het jaar was dit nog 31%.

Verkeer lokaal afhandelen

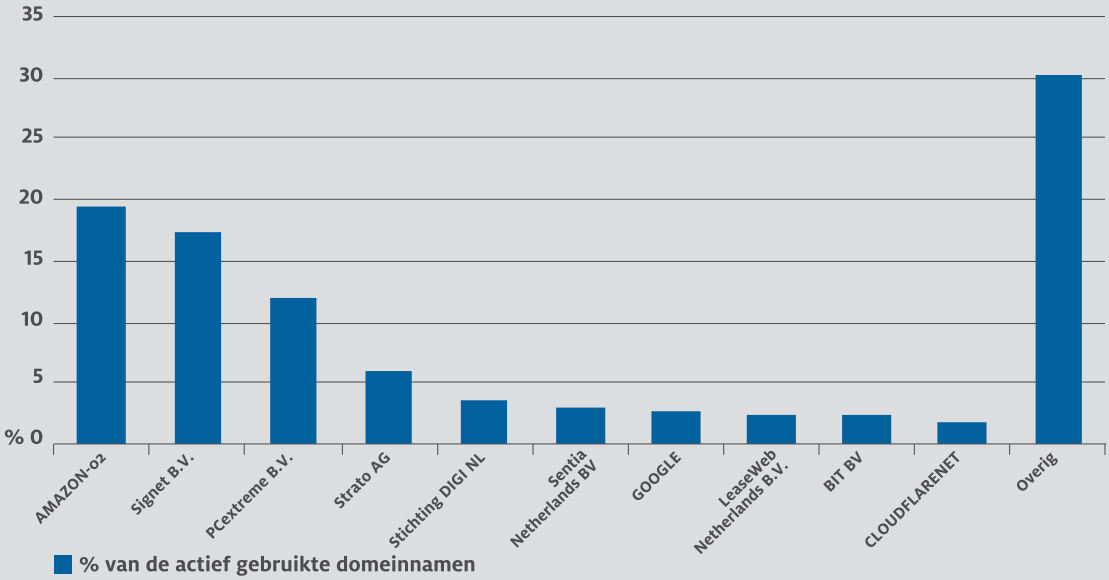
Een reden hiervoor kan zijn dat het DNS-verkeer in 2021 'lokaler' werd gerouteerd. Zo ging mogelijk begin van het jaar een DNS-aanvraag van een gebruiker in Nederland eerst naar een resolver in de VS, maar gaat dezelfde aanvraag nu naar een resolver van dezelfde aanbieder in Europa. Wij voeren hiervoor metingen uit naar zogenaamde 'anycast catchments', maar die bespreken we niet in dit rapport.

6.3 Webhosting

Ook bij webhosting is er een concentratie te zien naar een aantal grote hostingpartijen. De concentratie van zoveel domeinnamen bij enkele partijen kan bij uitval/storing van een van deze partijen tot gevolg hebben dat een niet te verwaarlozen deel van de Nederlandse websites tijdelijk niet meer bereikbaar is.

Actief gebruikte domeinnamen

Figuur 6.4 toont de top 10 hostingpartijen, die de meeste actief gebruikte .nl-domeinnamen hosten. In de figuur laten we alleen data voor 'actief gebruikte' domeinnamen zien (in § 6.5 leggen we uit wat we daaronder verstaan). Uit het oogpunt van centralisatie zijn dit de interessantste domeinnamen, omdat een storing of veiligheidsprobleem bij een actieve domeinnaam veel meer gevolgen heeft dan een storing bij bijvoorbeeld een parking page.



Figuur 6.4 | Concentratie in webhosting van .nl-domeinnamen.

De laatste staaf 'Overig' in figuur 6.4 laat zien dat 30% van de actief gebruikte domeinnamen wordt gehost door hosters buiten de top 10 voorkomen. In de top 3 staan, met flinke afstand tot de rest, de grootste hostingpartijen, deze verzorgen de webhosting voor respectievelijk 19% (Amazon), 17% (Signet) en 12% (PCextreme) van de actief gebruikte .nl-domeinnamen. Samen is de top 3 verantwoordelijk voor de webhosting voor ruim 48% van de actief gebruikte domeinnamen. Na de top 3 neemt het percentage domeinnamen gehost per individuele hostingpartij snel af.

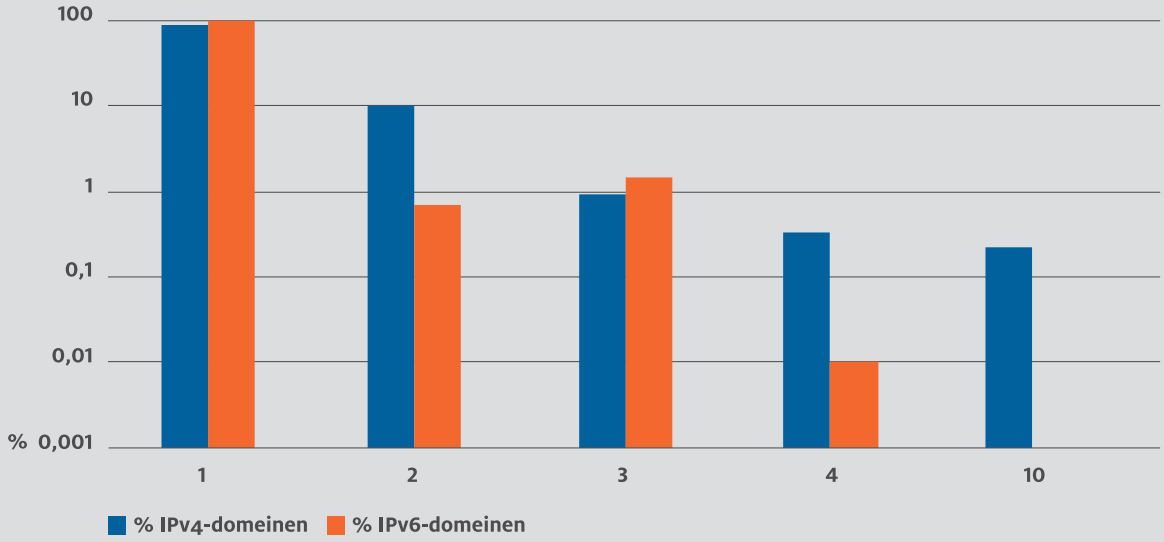
6.4 E-mail

Net zoals bij webservers zoeken resolvers aan een domeinnaam gekoppelde mailservers ook op via een DNS-query (zie hoofdstuk 2). Concentratie treedt op als mailservers dezelfde infrastructuur (netwerken, IP-adressen e.d.) gebruiken. Dit kan ook hier potentieel zorgen voor een verminderde beschikbaarheid of het helemaal uitvallen van een maildienst.

Vaak maar 1 IP-adres voor mail

Op basis van onze metingen vonden we dat de meeste domeinnamen 1, 2, 3, 4 of 10 IP-adressen voor hun mailservers gebruiken. Andere aantallen zien we niet of vrijwel niet. Dit laten we zien in figuur 6.5 (let op: de y-as is een logaritmische schaal). Voor IPv4 maakt 88% van de domeinnamen gebruik van maar 1 IP-adres. De situatie voor IPv6 is nog erger: hierbij maakt 98% van de domeinnamen gebruik van maar 1 IP-adres. Wat verder opvalt is dat op plaats 5 domeinnamen met 10 IP-adressen staan (in plaats van domeinnamen met 5 IP-adressen).

Als een domeinnaam maar 1 mailserver-IP-adres heeft, dan betekent dit dat de e-mailserver (of groep van e-mailservers) maar via 1 netwerk verbonden is met het internet. Een netwerkstoring zou dan ook meteen tot een e-mailstoring leiden. Deze kans is significant kleiner als een e-mailserver die via meerdere netwerken verbonden is of bij het gebruik van verschillende e-mailservers op verschillende netwerken.



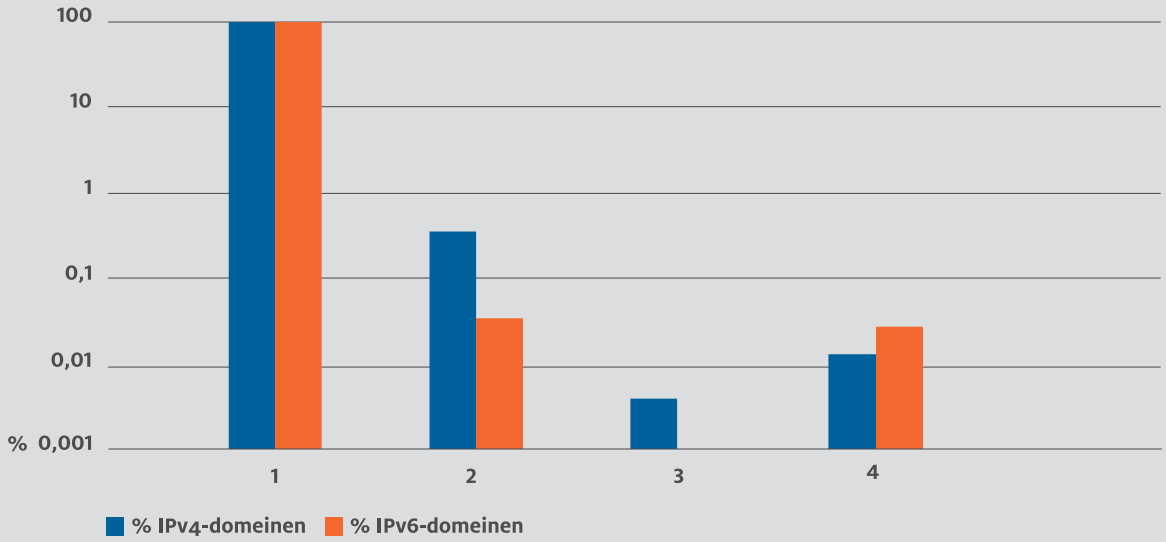
Figuur 6.5 | Aantal IP-adressen gebruikt voor mailservers onder .nl.

Wel meer IP-adressen maar toch in 1 netwerk

Bij mailservers onder .nl die wel gebruikmaken van meer dan 1 IP-adres komt het regelmatig voor dat de gebruikte IP-adressen zich toch in hetzelfde netwerk (ASN) bevinden.

Dit zien we in figuur 6.6 (ook weer met een logaritmische y-as): voor IPv4 maakt 97% van de domeinnamen gebruik van een enkel netwerk, voor IPv6 is dit 98%.

Voor de betrouwbaarheid van een e-mailserver heeft het gebruik van meerdere IP-adressen weinig zin, als deze IP-adressen in hetzelfde netwerk vallen. Een storing in dat betreffende netwerk zorgt er dan immers voor dat beide adressen onbereikbaar zijn.



Figuur 6.6 | Aantal netwerken gebruikt voor mailservers onder .nl.

6.5 Meetmethodes

De locaties van .nl-nameservers in § 6.1 zochten we op met behulp van een GeoIP-database. De netwerknummers zijn te vinden in routingdatabase, net als de informatie over de eigenaar van het netwerk. Met de term 'actief gebruikte domeinnaam' in § 6.1 bedoelen we een domeinnaam die daadwerkelijk inhoud heeft, mits er überhaupt een website is. Dat kan een bedrijfswebsite zijn, of een forum, maar ook een kleine privéwebsite met alleen wat tekst. Domeinnamen die we niet onder actief gebruikt scharen zijn domeinnamen waar een parking page van een internetprovider op staat, of bijvoorbeeld een melding 'gereserveerd voor gebruik'. De status van een domeinnamen bepalen we op basis van onze DMAP-dataset (§ 2.5).

We tellen de query's van publieke resolvers in § 6.2 aan de hand van een lijst van de netwerken en IP-adressen van deze resolvers, die we combineren met informatie uit de routingdatabase.

De routingdatabase gebruiken we ook voor het vinden van de webhosting-netwerken in § 6.3.

Het aantal netwerken van mailservers in § 6.4 bepalen we met informatie uit DMAP, ook weer in combinatie met gegevens uit de routingdatabase.

07

Een veilige domeinnaam met DNSSEC

In de voorafgaande hoofdstukken lieten we zien hoe belangrijk het DNS is voor onze maatschappij. Daarom is het ook erg belangrijk dat je er als internetgebruiker op kunt vertrouwen dat je uitkomt bij de juiste website, als je een domeinnaam intypt in je browser. Hiervoor standaardiseerde de IETF in 2005 DNSSEC (zie hoofdstuk 2).

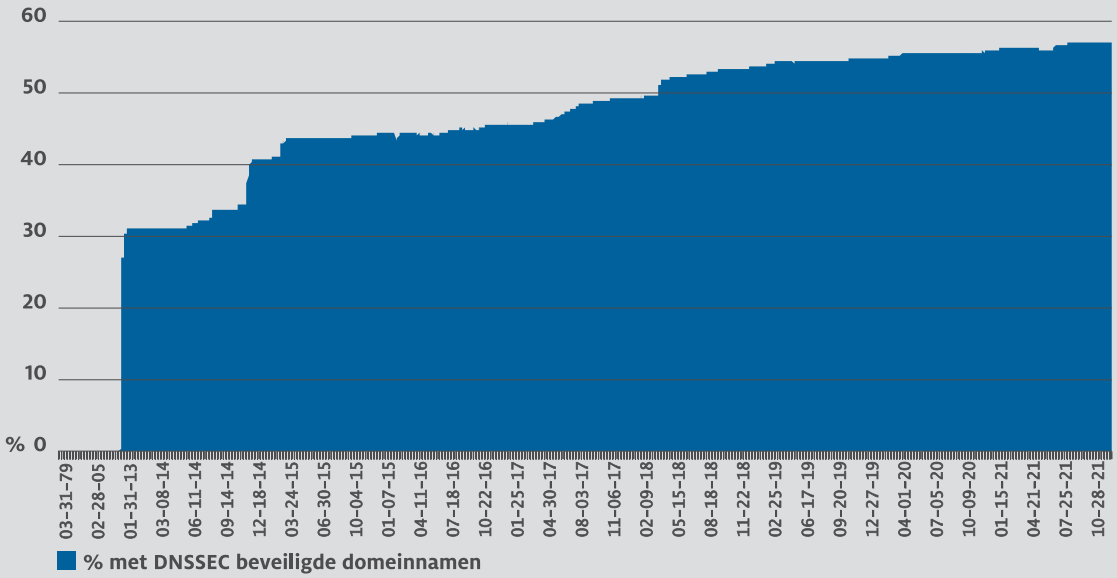
In dit hoofdstuk laten we zien hoe de .nl-zone in het afgelopen jaar in beweging was op het gebied van DNSSEC. Dit was niet direct zichtbaar voor eindgebruikers maar is toch cruciaal voor hun veiligheid op het internet.

07

Een veilige domeinnaam met DNSSEC

7.1 DNSSEC-adoptie in .nl

DNSSEC draagt bij aan de betrouwbaarheid van het internet. Om die reden stimuleren we al sinds jaren het gebruik van DNSSEC, onder andere met behulp van cursussen en incentives voor de .nl-registrars en praktische handleidingen over hoe je de standaard toepast. Hierdoor maakt bijna 60% van de .nl-domeinnamen gebruik van DNSSEC (zie figuur 7.1). Geen ander topleveldomein heeft in absolute aantallen meer gesigneerde domeinnamen. In figuur 7.1 is ook te zien dat de groei van DNSSEC onder .nl de laatste jaren afvlakt.



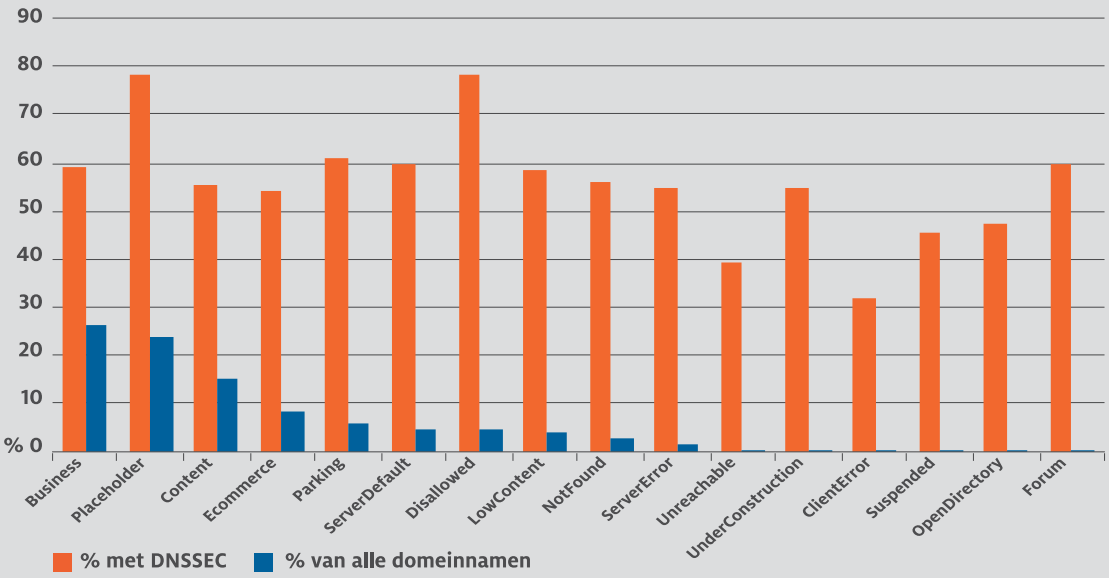
Figuur 7.1 | Ontwikkeling DNSSEC in de .nl-zone.

7.2 De drijfveren achter DNSSEC

Het effect van een succesvolle 'DNS cache poisoning'-aanval (zie hoofdstuk 2) op een domeinnaam van bijvoorbeeld een bank is potentieel groter dan een aanval op de domeinnaam van een privéblog. Daarom onderzoeken we waar DNSSEC-beveiliging nog ontbreekt.

DNSSEC per websitetype

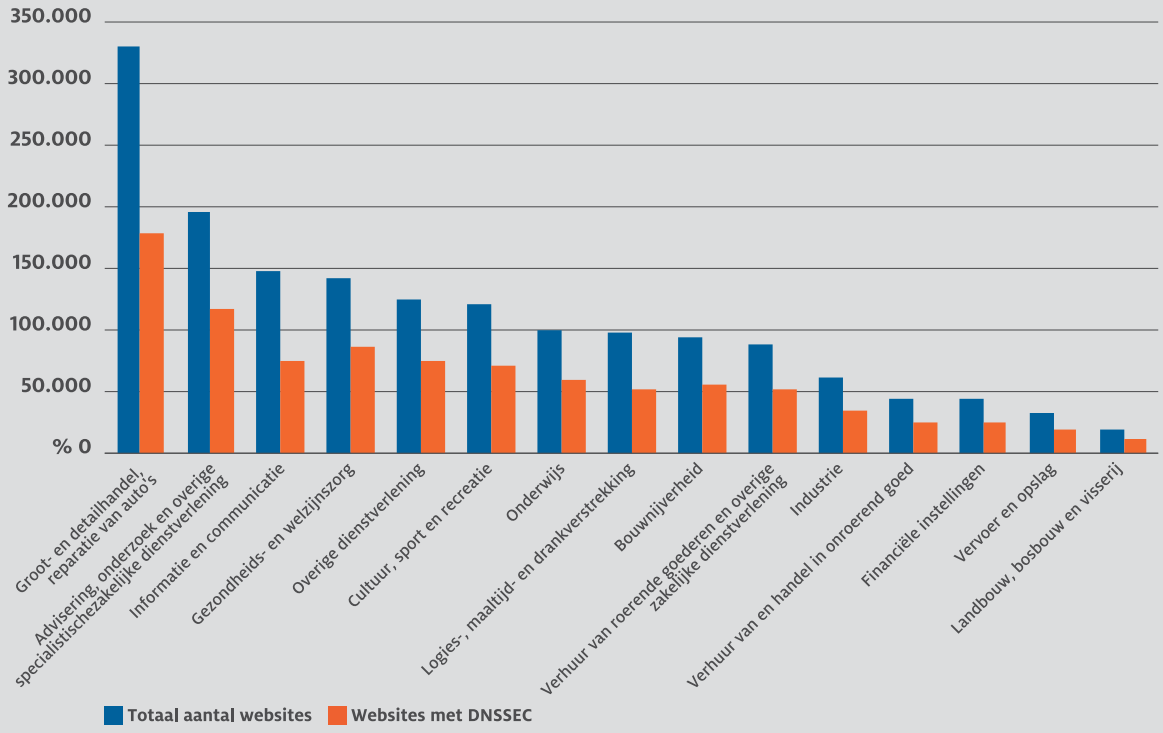
In figuur 7.2 zien we dat vooral e-commercewebsites vaak nog geen gebruikmaken van DNSSEC. Een aanval op dit soort websites kan echter wel als gevolg hebben dat gegevens van eindgebruikers worden gestolen, omdat gebruikers door een 'cache poisoning'-aanval zonder dat ze het merken worden omgeleid naar een nepwebsite. Domeinnamen van bedrijven (categorie Business) maken bovengemiddeld vaak gebruik van DNSSEC, maar toch mist 40% van deze domeinnamen die bescherming alsnog.



Figuur 7.2 | DNSSEC-gebruik per websitetype voor .nl-domeinnamen.

DNSSEC per sector

Als we iets dieper graven zien we dat van de websites die gebruikt worden voor e-commerce- of bedrijfswebsites, bedrijven uit de sector Informatie en communicatie vaak geen DNSSEC gebruiken (zie figuur 7.3). Dat is verrassend, aangezien deze bedrijven toch de meeste kennis zouden moeten hebben om DNSSEC uit te kunnen rollen. Tegelijk is er over het algemeen geen groot verschil te zien tussen de verschillenden sectoren en ligt het gemiddelde iets onder de 60%.



Figuur 7.3 | DNSSEC-gebruik per sector.

DNSSEC-gebruik door overheden

DNSSEC wordt wel vaak gebruikt bij domeinnamen van de rijksoverheid, provincies, en gemeentes. Bij de Rijksoverheid maakt 90% van de websites gebruik van DNSSEC (gebaseerd op het Websiteregister Rijksoverheid), bij gemeentes is dit 98%, en bij provincies 92%. Dit komt vooral omdat de overheid DNSSEC verplicht stelt voor dit soort diensten.

Dat deze verplichting effectief is, wordt nog duidelijker als we deze cijfers vergelijken met domeinnamen uit de private sector. Zo maakt maar 58% van de domeinnamen van Nederlandse banken gebruik van DNSSEC.

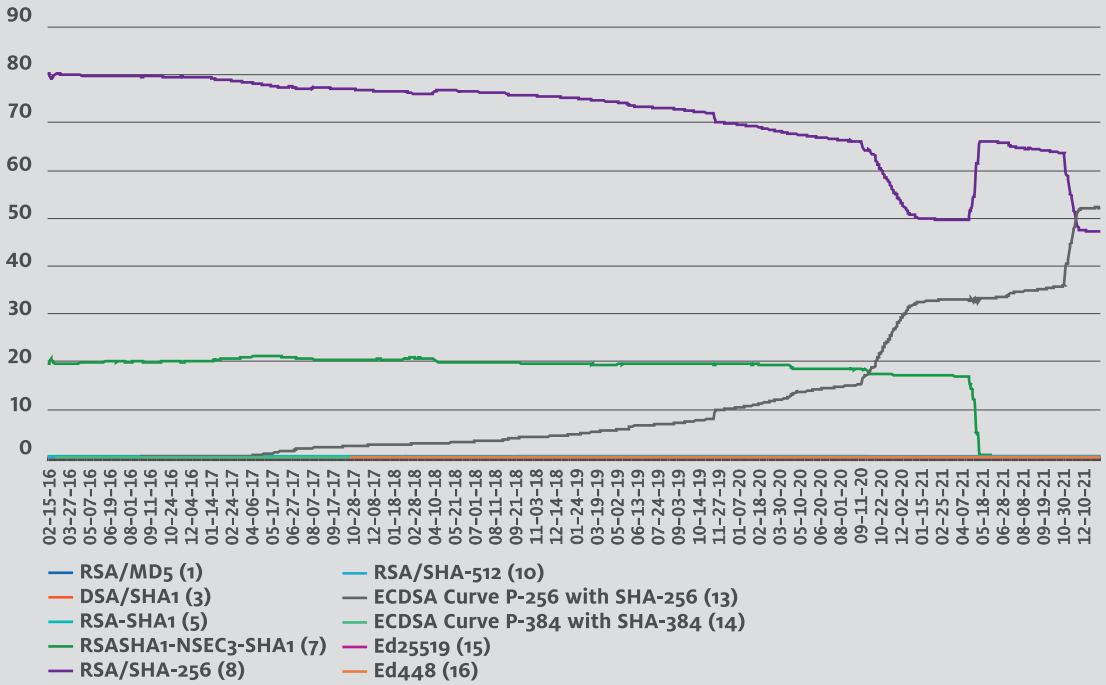
7.3 Veilig blijven met nieuwe DNSSEC-algoritmes

DNSSEC is natuurlijk verder geëvolueerd sinds SIDN de eerste .nl-domeinnaam in 2012 ermee beveiligde. Zo worden vandaag de dag sommige algoritmes voor het digitaal tekenen van DNS-records niet meer als veilig gezien (bijv. RSA/SHA1). Het is daarom belangrijk dat domeinnamen die voorzien zijn van een DNSSEC-handtekening dit doen op basis van een veilig cryptografisch algoritme.

Toekomstgericht

In het afgelopen jaar namen 2 grote providers hierin een grote stap. Eind 2020 is Mijndomein overgegaan om domeinnamen alleen nog maar met ECDSA P256 te ondertekenen en begin 2021 koos TransIP ervoor om van het onveilige algoritme RSA/SHA1 via een tussenstap over te stappen naar ECDSA. ECDSA P256 is een toekomstgericht algoritme dat vooral presteert door zijn erg kleine handtekeningen.

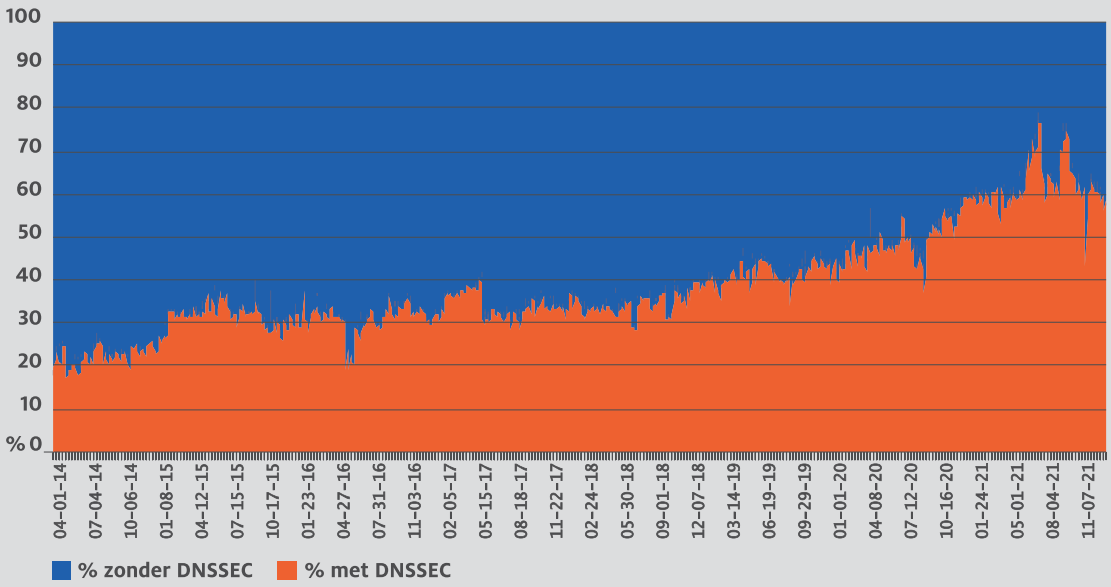
Figuur 7.4 laat het gebruik van de verschillende beschikbare cryptografische algoritmes zien. Er is een duidelijke overgang te zien van het gebruik van RSA-SHA256 naar ECDSA.



Figuur 7.4 | DNSSEC-algoritmen in gebruik bij .nl sinds 2016.

7.4 DNSSEC-handtekeningen valideren

DNSSEC beschermt alleen tegen aanvallen als resolvers de DNSSEC-handtekeningen ook valideren (zie hoofdstuk 2). In tegenstelling tot het aantal gesigneerde domeinnamen, steeg het aantal query's dat afkomstig is van validerende resolvers in het afgelopen jaar wel significant. Dit zie je in figuur 7.5.



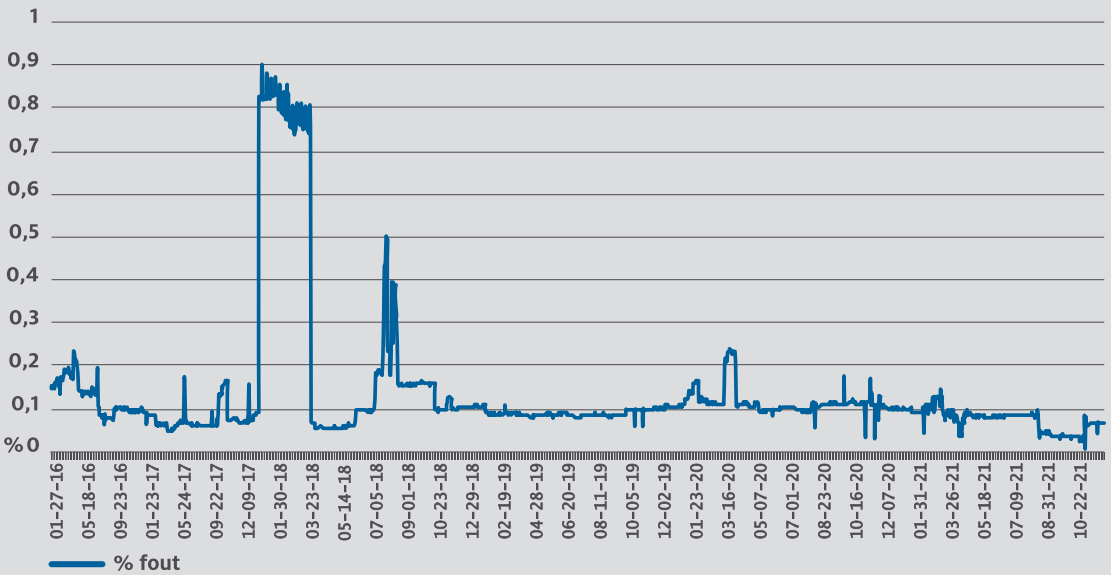
Figuur 7.5 | DNS-query's afkomstig van DNSSEC validerende resolvers.

Stijging validatie

In 2019 was maar 44% van de query's afkomstig van validerende resolvers. In 2021 steeg dit percentage naar 58%. Als deze resolvers een ondertekend antwoord krijgen kunnen zij zeker zijn dat de informatie niet afkomstig is van een aanvaller. Een reden voor deze groei is dat de Nederlandse internetaanbieder KPN sinds begin 2020 DNSSEC valideert voor zijn vaste en mobiele klanten.

Probleemloos

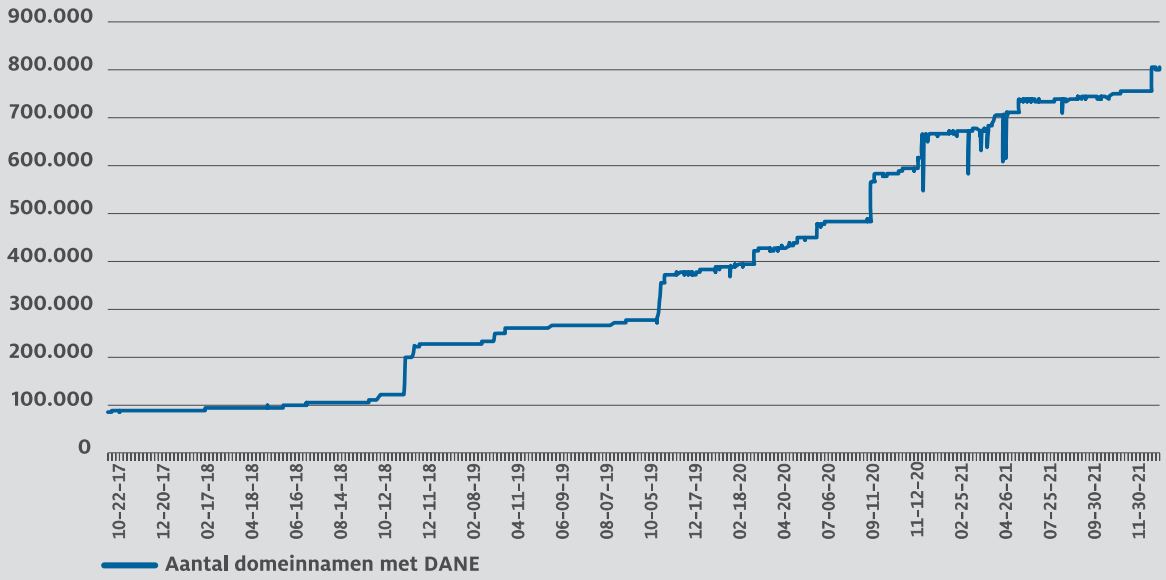
Een veelgebruikt argument om niet te gaan valideren is dat websites er onbereikbaar door zouden kunnen worden bij een verkeerde configuratie. Zo zou DNSSEC-validatie tot meer problemen leiden dan het oplost. Figuur 7.6 laat zien dat het aantal domeinnamen dat DNSSEC niet goed heeft geconfigureerd, en dus voor problemen bij resolvers kan zorgen, lager is dan ooit. Nog een reden dus om DNSSEC-validatie aan te zetten.



Figuur 7.6 | Domeinnamen waarbij DNSSEC niet goed is geconfigureerd. Y-as in percentages van alle .nl-domeinnamen met DNSSEC.

7.5 Adoptie van DANE

Figuur 7.7 illustreert dat het aantal .nl-domeinnamen met een door DANE beveiligde mailserver (zie hoofdstuk 2) de afgelopen jaren snel stijgt. DANE is een goed voorbeeld van de toegevoegde waarde die DNSSEC biedt om ook andere delen van het internet en onze dagelijkse communicatie te kunnen beveiligen.



Figuur 7.7 | .nl-domeinnamen met een door DANE beveiligde mailserver.

7.6 Meetmethodes

De websitestypes classificeren we met behulp van onze eigen DMAP-crawler en baseren we op de inhoud van de website (§ 7.2). Hierbij kijken we bijvoorbeeld naar het vermelden van een KvK-nummer, of de melding dat een domein al dan niet tijdelijk geparkeerd is.

De verschillende sectoren in de sectorclassificatie (§ 7.2) zijn gebaseerd op de Standaard Bedrijfsindelings (SBI)-codes van de Kamer van Koophandel, een attribuut in onze DMAP-dataset.

De gebruikte DNSSEC-algoritmes (§ 7.3) komen uit de DNS-informatie van de zones zelf, net als informatie over het gebruik van DANE (§ 7.5).

Om te bepalen of een resolver DNSSEC-handtekeningen valideert (§ 7.4) meten we of een resolver ook daadwerkelijk DNSSEC-records (zoals handtekeningen en publieke sleutels) opvraagt. Deze methode is niet 100% nauwkeurig maar geeft ons wel een goede inschatting over resolvergedrag.

08

IPv6

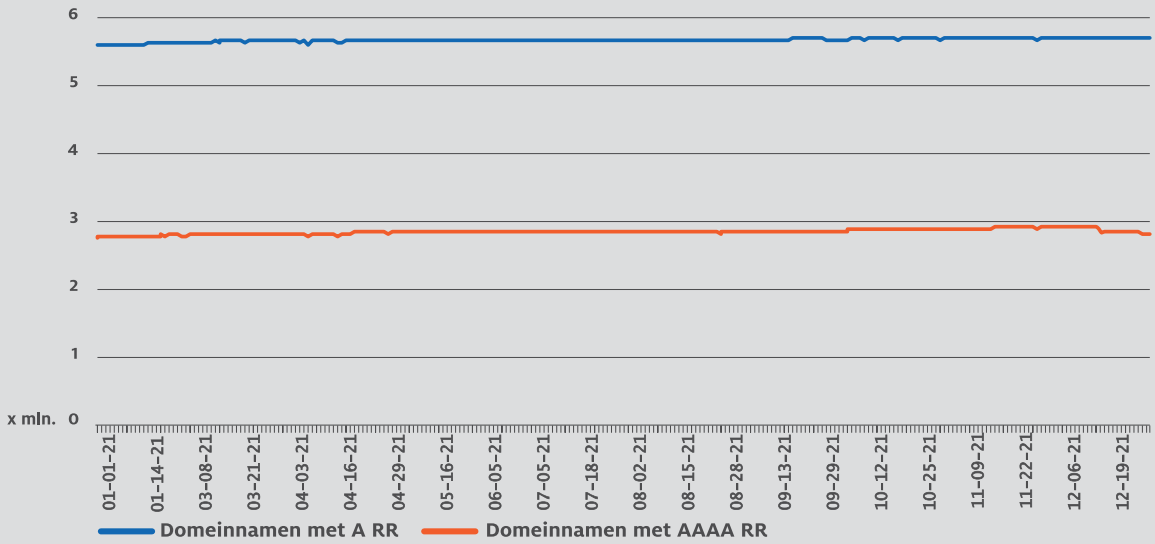
In hoofdstuk 2 schreven we dat de IPv4-adressen ‘op’ zijn. Het is daarom belangrijk dat de opvolger, IPv6, breed ondersteund wordt, zodat het internet kan blijven groeien. Onze metingen geven een goede indicatie van hoe de transitie naar IPv6 zich ontwikkelt – vooral binnen het .nl-domein.

08

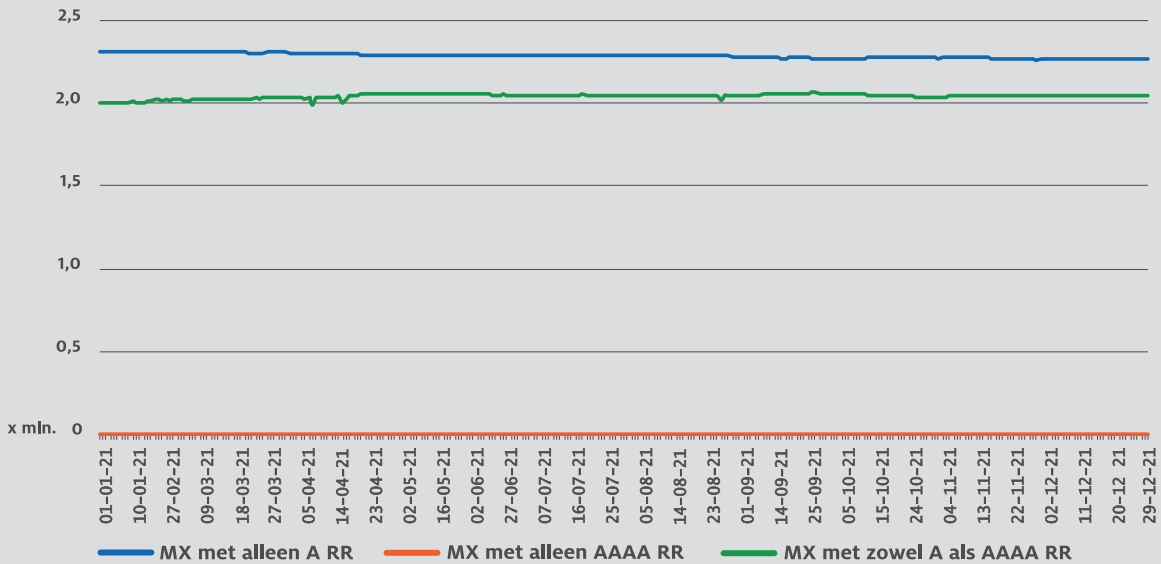
IPv6

8.1 Adoptie op basis van het DNS

Figuur 8.1 laat bijvoorbeeld zien in welke mate resolvers de nameservers van .nl-domeinnamen over IPv6 kunnen bereiken. Figuur 8.2 laat dit specifiek voor mailservers zien. We zien dat beide grafieken, weliswaar langzaam, maar geleidelijk stijgen richting meer IPv6.



Figuur 8.1 | IPv6-ondersteuning bij .nl-domeinnamen.



Figuur 8.2 | IPv6-ondersteuning bij mailservers.

Sinds begin 2021 zijn er zo'n 70.000 domeinnamen bijgekomen met een zogenaamd AAAA-record, wat betekent dat de domeinnaam bereikbaar is via IPv6. Dit is weliswaar een stijging, maar we kunnen helaas wel stellen dat de groei van IPv6 in Nederland nagenoeg tot stilstand is gekomen.

8.2 Opgezochte IPv6-adressen

Datzelfde patroon zien we ook terug bij resolvers die een IPv6-adressen opzoeken. Figuur 8.3 laat het percentage van A- en AAAA-query's zien ten opzichte van het totaal aantal query's voor .nl-domeinnamen. 'A' staat daarbij voor het opzoeken van een IPv4-adres, en 'AAAA' voor het opzoeken van een IPv6-adres. Het percentage van query's naar IPv6-adressen steeg in 2021 niet ten opzichte van het percentage query's naar IPv4-adressen.

We zien in deze grafiek wel een ander interessante gebeurtenis in 2021, die verder niet met IPv4 of IPv6 te maken heeft. Het gaat om de grote pieken van andere query's. Die waren afkomstig van de publieke DNS-resolver van Google die lange tijd een hoog aantal query's stuurde voor zogenaamde TXT-records met op het oog willekeurige domeinnamen. De query's zijn inmiddels gestopt, maar we weten nog niet wat de oorzaak was.

Figuur 8.3 | IPv4- en IPv6-query's voor .nl-domeinnamen.

8.3 Meetmethodes

De statistieken over domeinnamen en mailservers komen (§ 8.1) uit het meetplatform OpenINTEL, wat actief de .nl-zone scant en daarbij onder andere de ondersteuning van IPv4 en IPv6 opslaat.

De statistieken van querytypes (§ 8.2) zijn rechtstreekse cijfers uit ons eigen ENTRADA-platform.

Colofon

Het onderzoeksrapport ‘De staat van .nl’ is een uitgave van SIDN en is samengesteld door SIDN Labs.

Aan dit verslag werkten mee:
Marco Davids, Marnie van Duijnhoven, Cristian Hesselman, Jelte Jansen, Elmer Lastdrager, Moritz Müller, Martin Sluijter, Thymen Wabeke en Maarten Wullink.

Concept en vormgeving
Lumen Ontwerpersnetwerk, Breda

Vertalingen
G & J Barker Translations, Worcester, Verenigd Koninkrijk
www.gandjbarker.co.uk

Heb je vragen over het onderzoek, mail dan naar sidnlabs@sidn.nl.

Meld je aan voor onze nieuwsbrief
www.sidn.nl/nieuwsbrief

SIDN
Meander 501
6825 MD Arnhem
Postbus 5022
6802 EA Arnhem
T +31 (0)26 352 55 00
www.sidn.nl
www.sidnlabs.nl