

CSAM Hosting Monitor

Rapport september 2020

Qasim Lone, Carlos H. Gañán, Michel van Eeten

```
    } else if (a) {
      for (; o > i; i++)
        if (r = t.call(e[i], i, e[i]))
          return r;
    } else
      for (i in e)
        if (r = t.call(e[i], i, e[i]))
          return r;
  },
  trim: b && !b.call("\ufeff\u00a0") ? function(e) {
    return null == e ? "" : b.call(e)
  } : function(e) {
    return null == e ? "" : (e + "").replace(/^\s+|\s+$/, "")
  },
  makeArray: function(e, t) {
    var n = t || [];
    return null != e && (N(Object(e)) > n.length) ?
      Array.prototype.slice.call(e, 0, n.length + 1) : n
  },
  isArray: function(e, t, n) {
    var r;
    if (t) {
      if (n) return n.call(t, e, n);
      for (r = t.length, n = n ? 0 > n ? -n : r : r; r > n; r++)
        if (n in t && t[r] === e) return true;
    }
    return false
  }
};
```


Inhoudsopgave

Samenvatting	2
1 Introductie	3
2 Methodologie	5
2.1 CSAM data	5
2.2 CSAM Hosting Monitor	6
2.3 Meten van CSAM hosting over tijd.	7
2.4 Meting van CSAM verwijdering na NTD.	9
2.5 Beperkingen	9
3 Het CSAM landschap in Nederland	11
3.1 CSAM volume	11
3.2 Verdeling van CSAM over hostingbedrijven.	12
3.3 Verdeling van CSAM over domeinen	13
4 Snelheid van CSAM verwijdering	19
4.1 Meten van snelheid.	19
4.2 Verwijderingssnelheid per hostingbedrijf	20
4.3 Verwijderingssnelheid per domein	21
4.4 Verwijderingssnelheid van de URL's voor politie	22
4.5 Tot slot	22
5 Conclusie	25

Samenvatting

Sinds oktober 2018 werkt TU Delft aan een monitor die weergeeft waar Child Sexual Abuse Material (CSAM) gehost wordt in Nederland. Deze monitor is ontwikkeld in nauwe samenwerking met het EOKM en AbuseIO. De belangrijkste bevindingen zijn:

1. De hoeveelheid CSAM die is aangetroffen in Nederlandse hostingnetwerken is fiks gestegen in de 2020, maar de aantallen fluctueren hevig van maand tot maand. De groei is het gevolg van verschillende factoren die op dit moment niet van elkaar gescheiden kunnen worden. Er staat mogelijk meer materiaal online, maar er wordt wellicht ook beter gedetecteerd waardoor er meer materiaal gevonden wordt. Daarnaast wordt het gevonden materiaal nu efficiënter verwerkt en kunnen er sneller Notice & Takedown (NTD)-verzoeken verstuurd worden richting de hostingpartijen en domeineigenaren. De fluctuaties in de hoeveelheid CSAM komen ook door toename en afname in de data die EOKM krijgt uit het INHOPE-netwerk. Sinds maart 2020 is het Canadese meldpunt niet langer lid van INHOPE, waarmee die data wegvalt. Dit betekent dat IWF, het INHOPE-lid uit het Verenigd Koninkrijk, nu een belangrijkste bron van data is. De hoeveelheid URL's die door IWF zijn gemeld aan EOKM is fiks gedaald in de periode juni-augustus. Dit kan betekenen dat er een daling is in de hoeveelheid materiaal dat in Nederland wordt gehost, maar in het licht van alle fluctuaties is dat op dit moment nog niet vast te stellen.
2. Ondanks de fluctuaties in de hoeveelheid CSAM, is er wel een stabiel patroon in waar het materiaal gehost wordt. In 2019 en 2020 concentreerde het materiaal zich sterk bij vier hostingbedrijven. Deze zijn verantwoordelijk voor meer dan 95% van alle URL's. Onze korte dataset uit 2018 bevestigt dit beeld en toont drie van die vier providers (al werkte een van hen onder twee andere namen in 2018). Over het algemeen wordt het CSAM hosting landschap gedomineerd door een kleine set van bedrijven.
3. Er is veel minder stabiliteit als het gaat om de domeinnamen waar de CSAM wordt gehost bij de providers. Van de top 10 domeinen in 2020 waren er 5 ook in de top 10 in 2019 en geen enkele in de top 10 van 2018. De meeste domeinen met CSAM zijn legitieme image-hosting websites. Dit is niet uniek voor Nederland. Volgens INHOPE geldt dit voor de meeste CSAM wereldwijd waarover zij data hebben. Ook al verandert de lijst van domeinen met de meeste CSAM door de tijd, toch eindigen ook nieuwe domeinen weer bij dezelfde groep hosting providers. We hebben hiervoor geen goede verklaring op dit moment.
4. In de publiek-private werkgroep voor de bestrijding van online CSAM, gestart in 2018, is een norm overeengekomen: de hostingindustrie dient CSAM binnen 24 uur na een NTD verwijderd te hebben. In augustus 2020 heeft EOKM een steekproef van 2114 URL's gecontroleerd, in samenwerking met TU Delft. Deze URL's waren verdeeld over 15 hostingbedrijven en 70 domeinen. Ze werden bezocht door EOKM 24 uur na de NTD. Als het materiaal nog steeds online stond, werden ze opnieuw bezocht na 48 uur. Uit deze data blijkt dat 84% van alle CSAM verwijderd is binnen 24 uur, dat een verdere 12% verwijderd wordt tussen 24-48 uur en dat 4% online langer dan 48 uur online blijft – in sommige gevallen meer dan een week.
5. Alleen providers met een klein aantal URL's halen een 100% score in het verwijderen binnen 24 uur. De top 5 providers met de meeste URL's presteren zeer uiteenlopend. De slechtste haalt 0% offline binnen 24 uur, de beste 89%. We zien ook veel variatie op het niveau van de domeinen. Van de top 20 domeinen met de meeste URL's, hebben 11 domeinen meer dan 90% offline gehaald binnen 24 uur.

6. In aanvulling op het offline halen van het materiaal, nemen sommige providers met de meeste CSAM inmiddels meer preventieve maatregelen. Ze hebben sommige van hun klant-domeinen bewogen om zich aan te sluiten op de zogenaamde hashcheckservice van het EOKM. Deze dienst helpt image-hosting sites om het uploaden van bekende CSAM terug te dringen. Van de top 20 domeinen met de meeste CSAM in de afgelopen zes maanden (maart – augustus 2020), hebben er vier de hashcheckservice ingebouwd op hun website. Van deze vier zitten er drie bij NForce en een bij Leaseweb. De vier domeinen waren verantwoordelijk voor ongeveer 13% van alle URL's in die periode. De effecten van deze preventieve maatregelen zijn nog niet goed vast te stellen vanwege de vele factoren die van invloed zijn op de hoeveelheid CSAM die gehost wordt in Nederland.

1

Introductie

Al jaren staat Nederland hoog op de lijst van landen waar het meeste “child sexual abuse material” (CSAM) wordt gehost op internet.¹ Data over de locatie van CSAM wordt verzameld via de nationale meldpunten in het INHOPE-netwerk. Gebaseerd op de data van 2019 noemde INHOPE Nederland als het tweede land in de wereld qua hosting van CSAM en het leidende land in Europa.² Mondiaal hebben Nederland en de Verenigde States het grootste aandeel, elk met meer dan 20% van het materiaal. In Europa werd 79% van alle URL's die INHOPE had gevonden gehost in Nederland.

Dit probleem wordt niet direct veroorzaakt door de hostingbedrijven. Hun klanten zijn bijvoorbeeld legitieme image-hostingsites die op hun beurt weer legitieme gebruikers trekken, maar ook gebruikers die de site misbruiken voor het uploaden en verspreiden van CSAM. Ook zijn hostingbedrijven niet de veroorzakers van dit probleem, ze worden wel in toenemende mate gezien als een partij die een belangrijke verantwoordelijkheid heeft in de bestrijding ervan. Op zijn minst worden bedrijven geacht snel en adequaat te reageren op zogenaamde “Notice & Takedown” (NTD)-verzoeken. Deze verzoeken worden verstuurd aan het hostingbedrijf en de domeineigenaar en vragen hen de CSAM te verwijderen. De reacties van hostingbedrijven zijn uiteenlopend, van snel tot traag tot geen enkele actie. In extreme gevallen zijn er zelfs bedrijven die zogenaamde ‘bullet-proof hosting’ voor criminelen aanbieden en moedwillig meewerken aan de hosting van illegale content.

In reactie op de prominente rol van Nederland in het hosten van CSAM zijn er in 2018 rondetafelbijeenkomsten georganiseerd met een brede coalitie van vertegenwoordigers van overheden, private partijen en wetenschap.³ Het doel van dit publiek-private initiatief is de bestrijding van CSAM hosting in Nederland te verbeteren. Het overleg heeft geleid tot de formulering van enkele gedeelde “ambities”. Een van deze ambities is om transparanter te maken welke hostingbedrijven betrokken zijn bij CSAM hosting en hoe snel deze bedrijven dit materiaal uit hun netwerken verwijderen zodra ze een NTD hebben ontvangen.

Dit rapport geeft mede invulling aan deze ambitie. We analyseren de hosting van CSAM in Nederland op basis van data van het Expertisebureau Online Kindermisbruik (EOKM), het Nederlandse INHOPE-meldpunt. We rapporteren over volume, locatie, hostingbedrijven en domeinen. We beantwoorden daarbij de volgende onderzoeksvragen:

¹CSAM is een van de voorkeursbegrippen die internationaal gehanteerd wordt in plaats van de termen kinderpornografie, kinderpornografisch beeldmateriaal en kinderporno. Zie: <https://www.eokm.nl/kennisbank/eokm-terminologie/>.

²INHOPE Annual Report 2019, online raadpleegbaar via: https://www.inhope.org/media/pages/the-facts/download-our-whitepapers/009c452175-1595854476/annualreport_inhope_2019.pdf

³Zie: <https://zoek.officielebekendmakingen.nl/kst-31015-150.pdf>

1. Welke Nederlandse bedrijven hosten CSAM in hun netwerk? Het eerste doel is te identificeren waar CSAM wordt gehost en door wie. We combineren de data van het EOKM met IP adressen, Autonomous System-informatie en WHOIS-registratiedata om te achterhalen welke bedrijven het netwerk beheren waar de CSAM is aangetroffen.
2. Hoe verandert door de tijd de locatie van CSAM in de Nederlandse markt? We volgen de plekken waar CSAM zich concentreert over op basis van drie momentopnames die verzameld zijn in de periode van oktober 2018 tot en met augustus 2020.
3. Hoe kan de verdeling van CSAM over hostingbedrijven en domeinen verklaard worden? We nemen de antwoorden op de eerste twee vragen en proberen te interpreteren hoe deze passen in het plaatje van de Nederlandse hostingmarkt.
4. Hoe snel reageren hostingbedrijven en hun klanten (de domeineigenaren) op NTD-verzoeken om CSAM te verwijderen? We rapporteren over welk gedeelte van het materiaal binnen 24 uur verwijderd is en hoe verschillende bedrijven presteren ten aanzien van deze 24-uurs norm. Hiervoor heeft het EOKM in augustus 2020 handmatige controles uitgevoerd op een steekproef van CSAM URL's, in samenwerking met TU Delft.

De rest van dit rapport is als volgt opgebouwd. We bespreken eerst de methodologie van de CSAM hosting monitor (Hoofdstuk 2). Daarna presenteren we de resultaten over de locatie van CSAM in de Nederlandse hostingmarkt en we bespreken hoe dit patroon zich ontwikkeld heeft de afgelopen jaren. (Hoofdstuk 3). Vervolgens analyseren we de resultaten van de snelheid waarmee NTD-verzoeken zijn uitgevoerd en CSAM verwijderd is (Hoofdstuk 4). We eindigen met enkele conclusies over de volgende stappen en de verdere ontwikkeling van de monitor.

2

Methodologie

De ontwikkeling van de monitor is gestart in november 2018. Het doel is om de locatie van CSAM te bepalen door de tijd heen en om te meten hoe snel hostingbedrijven en domeineigenaren reageren op NTD-verzoeken. In dit hoofdstuk beschrijven we eerst de data die we gebruiken voor de monitor en vervolgens hoe we die data koppelen aan hostingbedrijven en domeinen.

2.1. CSAM data

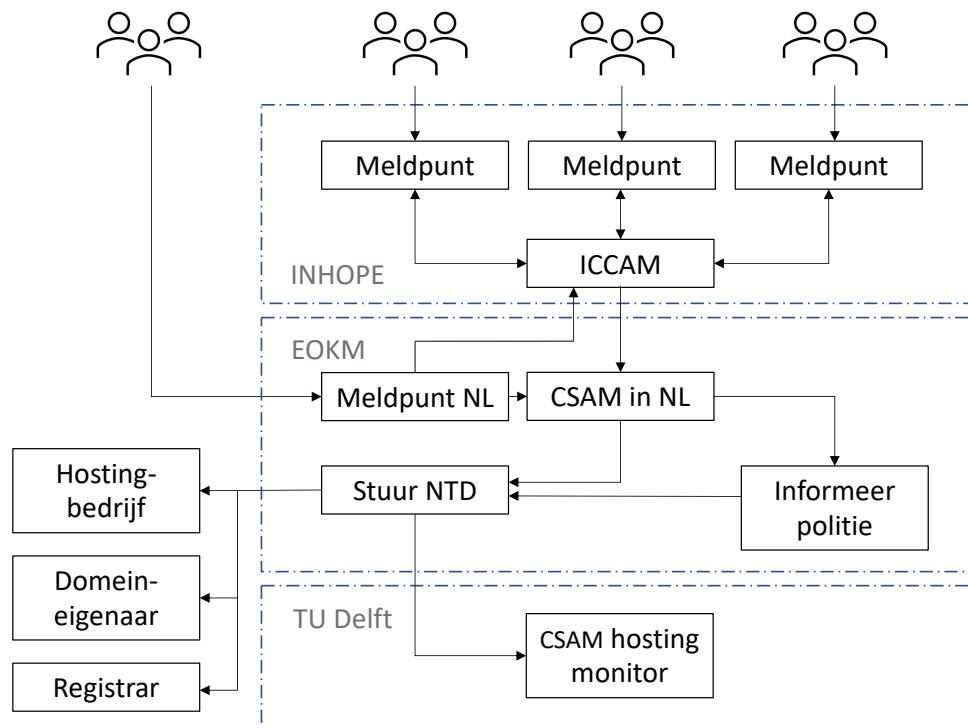
De belangrijkste data over de locatie van CSAM komt van het EOKM. EOKM ontvangt meldingen van vermoedelijke CSAM van vrijwilligers via het Meldpunt Kinderporno en van buitenlandse meldpunten in het INHOPE-netwerk. De INHOPE-data komt binnen via een systeem dat ICCAM heet. De meldingen hebben normaliter de vorm van een URL (Uniform Resource Locator) – simpel gezegd, een link naar een webpagina met CSAM of rechtstreeks naar een foto. De URL kan verwijzen naar een enkele foto of naar een verzameling foto's op bijvoorbeeld een pagina in een forum.

EOKM-medewerkers controleren de gemelde URL's en verifiëren dat de inhoud inderdaad CSAM betreft en illegaal is. Daarna controleert men dat het gehost wordt in Nederland. Zo niet, dan wordt de URL via ICCAM gestuurd naar het meldpunt in het land waar het wel gehost wordt. Als geverifieerd is dat het materiaal CSAM is en in Nederland wordt gehost, dan beslist EOKM of men de politie dient te informeren over de link of of dat men meteen een NTD-verzoek uitstuurt naar het hostingbedrijf, de domeineigenaar of, in een klein deel van de gevallen, naar de registrar waar de eigenaar de domeinnaam heeft geregistreerd.

Meestal wordt er meteen een NTD-verzoek gestuurd. In een klein deel van de gevallen wordt eerst de politie geïnformeerd. Daarna stuurt EOKM alsnog een NTD uit. Dit vindt echter plaats met een zekere vertraging. Er kunnen verschillende redenen zijn om een link naar de politie te sturen. Het kan gaan om CSAM die niet eerder is gezien, om materiaal met een mogelijke Nederlandse connectie qua slachtoffer of locatie of het kan gaan om een link die relevant zou kunnen zijn voor bepaalde lopende politieonderzoeken.

Uit de EOKM data haalt TU Delft de relevante gegevens voor de monitor. Tot januari 2020 werd het NTD-proces bij EOKM handmatig uitgevoerd, waarbij medewerkers e-mails maakten met de relevante links en die stuurde naar de betreffende bedrijven. Vanaf januari 2020 is een geautomatiseerd systeem in gebruik genaamd SCART.¹

¹SCART is gebaseerd op AbuseIO. Zie: <https://abuse.io/>



Figuur 2.1: Overzicht van het NTD-proces voor CSAM en de TU Delft monitor

Voor de monitor hebben we verschillende methoden en technieken ontwikkeld om de data te verwerken van zowel het handmatige proces als het geautomatiseerde systeem van het EOKM. Met steun en feedback from EOKM, hebben we scripts ontwikkeld om de CSAM URL's te verrijken en het aantal URL's per hostingbedrijf en domein te tellen. Deze scripts draaien binnen het EOKM netwerk, zodat geen CSAM URL's en domeinnamen het netwerk verlaten. Alleen de uiteindelijke statistische resultaten worden naar TU Delft verstuurd en daar verwerkt.

2.2. CSAM Hosting Monitor

De data die door de scripts worden verwerkt, bevatten het IP-adres waar de CSAM wordt gehost, het domein, de geschatte geografische locatie (op basis van GeoIP), het Autonomous System (AS) waar het IP-adres toe behoort, de organisatie aan wie het IP-adres is toegewezen volgens WHOIS data en het tijdstip waarop de NTD is verstuurd.

WHOIS data is afkomstig uit databases met informatie over de eigenaren en gebruikers van IP-adressen, domeinnamen en de Autonomous System Numbers die gebruikt worden bij de routing van verkeer. De WHOIS databases voor domeinnamen worden beheerd door de zogenaamde 'registries' van de verschillende top-level domeinen. Voor het .nl domein, bijvoorbeeld, is dat SIDN. Via SIDN's WHOIS kun je dus informatie opzoeken over de eigenaar en de contactgegevens voor een domeinnaam die eindigt op .nl.

De WHOIS databases voor IP-adressen worden beheerd door de vijf zogenaamde 'Regional Internet Registries' (RIR). Voor Europa is RIPE NCC (Réseaux IP Européens Network Coordination Centre) de RIR. WHOIS data over IP-adressen geeft informatie over de organisatie aan wie de betreffende adressen zijn toegewezen door de RIR, zoals hostingbedrijven. Die hostingbedrijven kunnen ervoor kiezen om op hun beurt zelf die adressen weer verder toe te wijzen aan een van hun klanten, dat zelf ook een bedrijf kan zijn. In termen van de verantwoordelijkheid voor wat er gebeurt op die adressen, zijn er twee scenario's. Als de klant en het hostingbedrijf overeenkomen dat de klant alle klachten en 'abuse'-

problemen afhandelt, dan past het hostingbedrijf de WHOIS-data bij de RIR aan en meldt hierin de contactgegevens van de klant voor abuse- en NTD-meldingen. Dit heet een 'sub-allocatie'. Als er in WHOIS geen sub-allocatie te zien is, dan betekent het dat het hostingbedrijf zelf de verantwoordelijkheid houdt voor het oplossen van abuse-problemen op die IP-adressen.

In Figuur 2.2 vatten we het proces samen waarmee we URL's toewijzen aan hostingbedrijven. We beginnen met het selecteren van de URL's die geverifieerd zijn als zijnde CSAM en illegaal en voor welke NTD-verzoeken zijn verstuurd naar de hostingbedrijven en/of domeineigenaren. We tellen geen URL's mee waarvoor geen NTD is verstuurd omdat de bedrijven dan vermoedelijk niet op de hoogte zijn van de aanwezigheid van het materiaal. Dit komt weinig voor en is meestal het gevolg van een tekort aan medewerkers bij het EOKM tijdens piekmomenten, waardoor niet alle URL's gecontroleerd konden worden of ze inderdaad CSAM bevatten. Als die controle wel verricht is, wordt er in beginsel altijd een NTD verstuurd.

Vervolgens zoeken we het IP-adres op in de IP WHOIS-database bij RIPE NCC.² Dit geeft ons de naam van de organisatie aan wie het IP-adres is toegewezen. De WHOIS-informatie bevat ook het e-mailadres voor abuse-meldingen en NTD's. We vergelijken het abuse-contact van WHOIS met het contactadres waarheen de NTD is gestuurd. Als deze niet hetzelfde zijn, dan tellen we die URL's niet mee voor dat hostingbedrijf, omdat we niet zelf kunnen verifiëren dat het bedrijf inderdaad de NTD heeft ontvangen op het adres dat vermeld staat in WHOIS. Dit gold alleen voor een minieme fractie van alle URL's.

In een klein aantal gevallen vonden we geen organisatie in RIPE NCC WHOIS, omdat het IP-adres is uitgegeven door een andere RIR. In die gevallen hebben we de organisatienaam achterhaald via de MaxMind IP2ORG database.³ Die mapping hebben we vervolgens handmatig gecontroleerd door de organisatienaam te vergelijken met de domeinnaam in het e-mailadres dat als abusecontact is gebruikt voor de NTD. Als dat overeenkomstig is, wijzen we de URL toe aan die organisatie.

Naast het identificeren van het hostingbedrijf, kijken we ook op welk domein de CSAM is aangetroffen. We halen de zogenaamde "Fully Qualified Domain" (FQDN) uit de URL. De URL bestaat uit een subdomein en het "parent" domein. Bijvoorbeeld: het sub-domein `abc` onder het parent domein `voorbeeld.nl` heeft als FQDN `abc.voorbeeld.nl`. We gebruiken FQDNs om URL's aan toe te wijzen, en niet de parent-domeinen, omdat in sommige gevallen verschillende subdomeinen gehost worden door verschillende hostingbedrijven. Na voltooiing vervangen onze scripts elke FQDN door een random nummer dat buiten EOKM verwerkt kan worden zonder de locatie van de CSAM te onthullen.

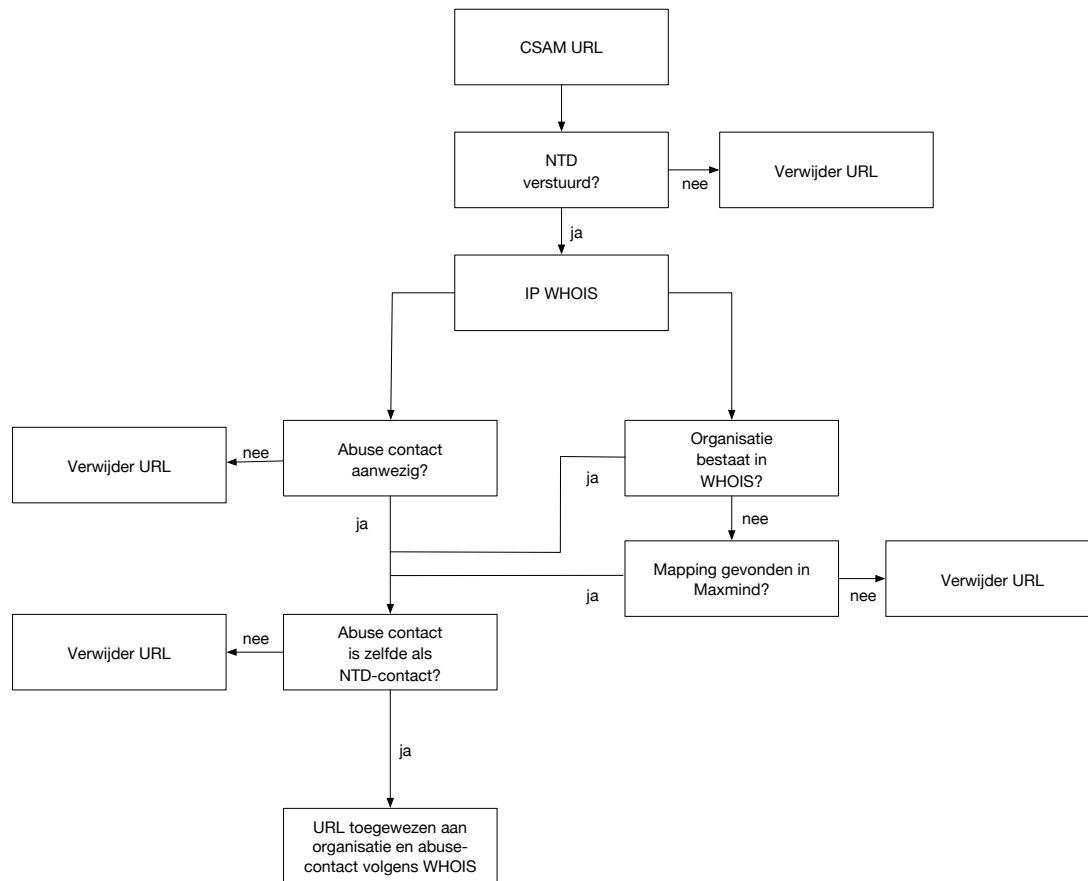
2.3. Meten van CSAM hosting over tijd

Idealiter zou de monitor door de tijd heen op een consistente manier de locatie van CSAM in Nederland meten. Dit zou ons dan een continue tijdlijn geven die voor elk moment toont waar het materiaal zich bevindt. Die momentopname zouden we dan kunnen vergelijken met eerdere en latere momenten. De monitor is in de loop der tijd dit ideaal gaan naderen, maar hij is nog in ontwikkeling en dus ook aan verandering onderhevig geweest sinds de start in november 2018.

Verschillende factoren hebben een rol gespeeld bij de wijze waarop dataverzameling is geëvolueerd. Allereerst diende het juridisch en technisch mogelijk gemaakt worden dat de data geanalyseerd kon worden binnen de regels en het netwerk van EOKM. De logbestanden van het handmatige proces tot januari 2020 waren lastig geautomatiseerd te verwerken. Om die reden konden we alleen de CSAM data

²Zie: <https://apps.db.ripe.net/db-web-ui/query>

³Zie: https://dev.maxmind.com/geoip/legacy/csv/#GeoIP_Organization_Edition_CSV_Database_Fields



Figuur 2.2: Toewijzen van CSAM URL's aan hostingbedrijven

reconstrueren voor twee specifieke periodes: oktober-november 2018 and augustus-november 2019. In december 2019 werd het nieuwe SCART-systeem in productie genomen. Dit heeft veel problemen met de dataverwerking opgelost. De maand december heeft slechts beperkte data vanwege de overgang naar het nieuwe systeem. Vanaf januari 2020 is er echter continue longitudinale data beschikbaar.

Een andere verandering is dat verschuivingen plaatsvonden in de gegevens die via ICCAM binnenkwamen. Aangezien de meeste EOKM-data over CSAM afkomstig is van ICCAM, hebben veranderingen in de manier waarop INHOPE-data gedeeld wordt een duidelijk effect op wat er in Nederland wordt waargenomen. De gegevens in ICCAM zijn afkomstig van andere INHOPE-meldpunten. De meldpunten in het V.K. en Canada zoeken proactief naar CSAM met webcrawlers en vinden daardoor relatief veel CSAM, vergeleken met meldpunten zoals EOKM die alleen meldingen krijgen van derden. Wanneer deze crawlers meer domeinen bezoeken die in Nederland worden gehost, bijvoorbeeld omdat men daar eerder CSAM heeft aangetroffen, neemt de kans toe dat er meer CSAM worden gevonden en dan komen er dus meer URL's binnen via ICCAM bij het EOKM. De tegenovergestelde verandering treedt ook op. In maart 2020 is het Canadese meldpunt uit INHOPE gestapt en dus ook gestopt met het delen van de in Nederland ontdekte CSAM met EOKM. In plaats daarvan besloot de Canadese hotline hun eigen NTD's te sturen. Verder kwam er vanaf juni 2020 veel minder data binnen via IWF. Het is niet bekend of dit komt door wijzigingen in de opsporing van CSAM of doordat er minder materiaal in Nederland wordt gehost.

In het licht van deze complexiteit nemen we de CSAM-data op in drie delen: een momentopname van oktober-november 2018, een van august-november 2019 en de continue meting vanaf januari 2020. Aangezien deze periodes gebaseerd zijn op verschillende datageneratieprocessen, kunnen we ze niet eenvoudig in dezelfde tijdlijn plaatsen. In plaats daarvan maken we verschillende tabellen om de belangrijkste overeenkomsten en verschillen in CSAM-hosting in 2018, 2019 en 2020 te bekijken. Voor 2020 geven we vervolgens een dieper beeld, inclusief trends in de tijd.

2.4. Meting van CSAM verwijdering na NTD

Samen met EOKM hebben we ook een meting gedaan van de snelheid waarmee NTD's werden opgevolgd. De norm die met de industrie is afgesproken in de publiek-private werkgroep was dat CSAM binnen 24 uur na de NTD door providers moet worden verwijderd. We presenteren de meting en de onderliggende methodologie in meer detail in hoofdstuk 4.

2.5. Beperkingen

Tot slot bespreken we enkele van de methodologische uitdagingen voor de monitor. We houden rekening met deze beperkingen bij het interpreteren van de resultaten. Ten eerste zijn veel van de gegevens afkomstig vanuit ICCAM. Twee van de meldpunten, die in Canada en het V.K., zoeken proactief naar CSAM en zij leverden veel van de ICCAM data. Dit heeft verschillende implicaties. Ten eerste is het onduidelijk hoe hun crawlers en handmatige zoekopdrachten de nieuwe inhoud detecteren en hoe hun aanpak verandert over tijd. Om deze reden weten we niet of de daling in het volume van gedetecteerde CSAM sinds juni 2020 daadwerkelijk betekent dat er minder materiaal online is in Nederland. Daarnaast is het mogelijk dat de crawlers hun zoektocht richten op hostingbedrijven en domeinen waar al eerder materiaal is gedetecteerd, in plaats van elders. Dit kan resulteren in een vertekend beeld van het probleem, aangezien materiaal bij die hostingbedrijven eerder wordt ontdekt dan bij andere providers. Het feit dat andere bedrijven wellicht minder gecrawld worden laat natuurlijk onverlet dat de bedrijven die meer aandacht krijgen wel degelijk het betreffende materiaal hosten.

Verder zijn er fluctuaties in de data die worden veroorzaakt door wijzigingen in het INHOPE-netwerk. Sinds maart is het Canadese meldpunt geen lid meer van INHOPE. Dit betekent dat IWF, het U.K. INHOPE-lid, nu een belangrijke bron van gegevens is. Het aantal van IWF ontvangen URL's is de afgelopen drie maanden aanzienlijk gedaald. Het is niet duidelijk waardoor deze veranderingen in het CSAM-volume optreden. We kunnen niet zeggen of dit een weerspiegeling is van veranderingen in de aanwezigheid van CSAM in Nederlandse netwerken of veranderingen in de detectie van CSAM. Waarschijnlijk is het een combinatie van deze effecten.

Een ander probleem is dat sommige domeinen reageren met meerdere IP-adressen wanneer ze worden opgevraagd via DNS. Met andere woorden, een domeinnaam heeft identieke kopieën van dezelfde website die op meerdere servers staan, in sommige gevallen verspreid over verschillende hostingbedrijven. De domeineigenaren gebruiken vaak een techniek genaamd 'round-robin DNS'. Het biedt domeineigenaren de mogelijkheid om de belasting van hun website over verschillende hostingbedrijven te verdelen en het verbetert de fouttolerantie van hun website in geval van downtime bij een van de bedrijven. Deze opzet betekent dat hetzelfde domein, en dus hetzelfde materiaal, wordt gehost bij verschillende bedrijven. Voor de monitor hebben we de URL toegekend op basis van de DNS-lookup die werd uitgevoerd op het moment van de NTD, aangezien dit ook de basis is voor het verzenden van de NTD. Met andere woorden, al onze tellingen zijn gebaseerd op de NTD's.

Bij het versturen van de NTD's treedt ook een kleine vertekening op. EOKM houdt een 'groene lijst' van coöperatieve domeinen bij. Ze verwijderen de inhoud snel. EOKM stuurt ze meteen een NTD, zonder extra controles uit te voeren op de gerapporteerde CSAM. Voor de andere domeinen vindt er meer onderzoek plaats voor elke URL. In piekmomenten kan dit betekenen dat het even kan duren voordat EOKM tijd heeft om dit onderzoek uit te voeren. In die periode is de URL mogelijk offline gegaan. Als het onderzoek dan start, is de URL niet meer geldig en wordt er dus geen NTD verzonden. Dit zou betekenen dat deze domeineigenaar en provider mogelijk iets minder NTD's krijgen dan domeinen en providers op de 'groene lijst' bij dezelfde hoeveelheid materiaal. Gegeven dat we rekenen op basis van NTD's, betekent dit dat de 'groene lijst' domeinen en providers potentieel een iets hogere telling krijgen dan die niet op de groene lijst staan.

Een laatste en belangrijke uitdaging is om geautomatiseerde uptime-metingen uit te voeren van de URL's die door EOKM worden ontvangen. Geautomatiseerde metingen zouden ideaal zijn, omdat het ons fijnmazige gegevens zou opleveren over hoe lang een URL online blijft na een NTD. Dit is echter niet mogelijk door verschillende technische complicaties. Een daarvan is het gebruik van CAPTCHA's door de domeineigenaren. CAPTCHA's zijn tests die zijn opgenomen zijn op een webpagina en die een bezoeker vragen een eenvoudige taak uit te voeren om te bepalen of de gebruiker al dan niet een mens is. De huidige trackingscripts kunnen de CAPTCHA's niet automatisch oplossen en hierdoor kan de pagina-inhoud niet worden opgehaald. In ons rapport is daarom een handmatig uitgevoerde uptime-meting opgenomen, uitgevoerd door EOKM in nauwe samenwerking met de TU Delft. We beschrijven dit in meer detail in hoofdstuk 4. Enerzijds hielp het handmatige proces ons om CAPTCHA's te omzeilen. Aan de andere kant is de timing minder nauwkeurig. De controle na 24 uur vond in sommige gevallen eigenlijk wat later plaats, zeg na 26 uur. Nooit werd de controle eerder dan 24 uur uitgevoerd. De fout in de meting is dus altijd in het voordeel van het hostingbedrijf en de domeineigenaar. Anders geformuleerd, we ronden de verwijderingssnelheid af naar beneden. Een gerelateerd probleem zijn de NTD's die op vrijdag zijn verzonden. Omdat EOKM in het weekend geen controles heeft uitgevoerd, zijn deze URL's maandag opnieuw bezocht. Als ze op dat moment zijn verwijderd, hebben we ze binnen 24 uur als verwijderd geteld. Nogmaals, we zijn voorzichtig geweest in het beoordelen van de bedrijven.

Samenvattend staat de monitor voor bepaalde uitdagingen bij het meten van de locaties met CSAM-data. Hoewel deze problemen een effect hebben op de exacte hoeveelheden CSAM die worden berekend per bedrijf, zijn deze effecten niet zo groot dat ze de getrokken conclusies ondermijnen.

3

Het CSAM landschap in Nederland

In dit hoofdstuk onderzoeken we welke providers in Nederland CSAM hosten en waar het materiaal geconcentreerd is. Het belangrijkste doel van dit hoofdstuk is om de huidige staat van CSAM hosting te schetsen alsook hoe dit is geëvolueerd in de afgelopen twee jaren.

Zoals uitgelegd in hoofdstuk 2, hebben we drie datasets die verschillende tijdsperiodes bestrijken. Van 2018 en 2019 hebben we alleen twee kortere momentopnames. Van januari tot augustus 2020 hebben we longitudinale gegevens. Het belangrijkste onderscheid zijn de gegevens van voor en na het ingebruiknemen van het geautomatiseerde systeem, genaamd SCART, door EOKM. Hierdoor is het aantal URL's waarvoor EOKM NTD's kan verzenden toegenomen. Het viel ook samen met een grotere toestroom van URL's vanuit ICCAM. Zie hoofdstuk 2 voor meer details. Het doel van dit hoofdstuk is de volgende vragen te beantwoorden:

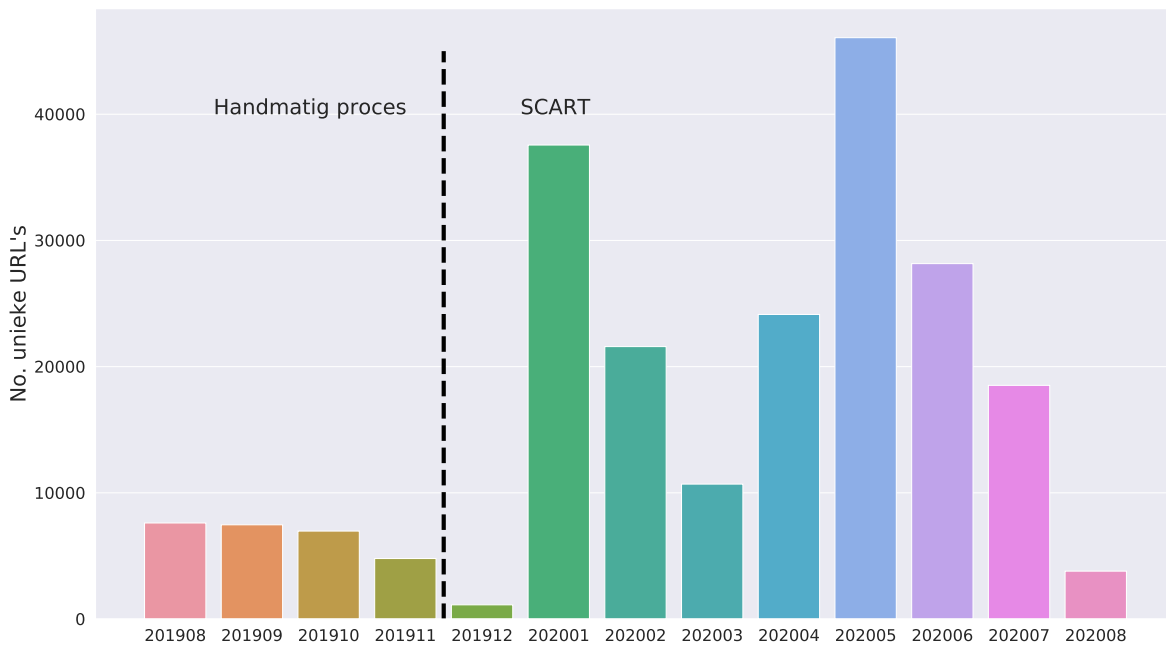
- Hoe is het volume van CSAM in de loop van de tijd veranderd?
- Hoe is CSAM verdeeld over hostingbedrijven? Hoe is dit veranderd?
- Hoe is CSAM verdeeld over domeinen? Hoe is dit veranderd?

We bespreken eerst patronen in het volume van CSAM in Nederland en dan hoe dit materiaal verspreid is over het hostinglandschap door de tijd heen. Vervolgens kijken we naar de domeinen waar dit materiaal is gehost en de dynamiek hierin.

3.1. CSAM volume

In figuur 3.1 tonen we het aantal URL's dat per maand gemeld is aan hostingbedrijven in de periode augustus – november 2019 en januari – augustus 2020. Te zien is dat het aantal URL's in 2019 rond de 5.000 tot 7.000 URL's per maand schommelde. Dit is vergelijkbaar met de volumes die we ook zagen in de momentopname van gegevens van 2018. In december 2019 is het geautomatiseerde systeem SCART getest en is het EOKM-personeel getraind in het gebruik ervan. Die overgang verlaagde tijdelijk het aantal URL's dat naar providers werd gestuurd. Om deze reden laten we deze maand buiten beschouwing in een deel van de analyse. Vanaf januari 2020 was het nieuwe SCART-systeem volledig operationeel.

Figuur 3.1 laat zien dat het aantal URL's waarvoor EOKM NTD's verzond aanzienlijk is toegenomen in 2020. Veel maanden zagen rond de 20.000 URL's of meer, maar het fluctueert sterk. We zullen de pieken in januari en mei bespreken als we naar de domeinen kijken.



Figuur 3.1: Aantal URL's waarvoor NTDs zijn verstuurd per maand, voor en na ingebruikname van SCART

De algehele groei is het resultaat van verschillende op elkaar inwerkende factoren die niet kunnen worden ontward. Er is misschien meer materiaal online beschikbaar, maar mogelijk ook betere detectie gaande, waardoor er meer materiaal wordt ontdekt. Er is ook een verbeterde efficiëntie bij het verwerken en verzenden van NTD's voor het gedetecteerde materiaal. Schommelingen worden bovendien veroorzaakt door stijgingen en dalingen in de ontvangst van gegevens uit het INHOPE-netwerk. Sinds maart is het Canadese meldpunt geen lid meer van INHOPE. Dit betekent dat IWF, het U.K. INHOPE-lid, nu een belangrijke gegevensbron is voor ICCAM.

Van juni tot en met augustus 2020 zien we een neerwaartse trend. Dit is voornamelijk te wijten aan een afname van het volume van de INHOPE ICCAM-gegevens, met name van IWF. We weten niet of deze daling een weerspiegeling is van een daadwerkelijke vermindering van de hoeveelheid online CSAM-materiaal in Nederland of dat er veranderingen waren in de manier waarop CSAM-materiaal werd gedetecteerd en gerapporteerd door IWF. Alleen als in de loop van het najaar van 2020 de aantallen laag blijven, kunnen we meer vertrouwen krijgen in de conclusie dat de hoeveelheid materiaal inderdaad is afgenomen.

3.2. Verdeling van CSAM over hostingbedrijven

In tabel 3.1 laten we de verdeling zien voor het jaar 2019 en 2020. Ondanks de fluctuaties in de CSAM-gegevens is de verdeling van URL's over hostingbedrijven opmerkelijk stabiel. Dezelfde vier bedrijven hebben de overgrote meerderheid van alle URL's: 98,03% in 2019 en 99,5% in 2020. Bovendien wordt de meeste inhoud gehost door domeinen bij één hostingbedrijf: NForce. Naast deze vier providers is er een lijst met bedrijven die een klein aantal gerapporteerde URL's hebben. In onze momentopname voor 2019 ontvingen in totaal 28 hostingbedrijven minimaal één NTD. Voor 2020 was dit 52 (gemeten over een langere periode).

Dit beeld is ook stabiel als we verder teruggaan in de tijd en het vergelijken met onze gegevens uit 2018. In tabel 3.2 hebben we de top 5 hostingbedrijven met de meeste URL's vergeleken in onze datasets uit 2018, 2019 en 2020. Zoals gezegd, toont het beeld voor 2019 en 2020 dezelfde set van providers. Er lijkt wat verandering te zijn ten opzichte van 2018, maar dit is vooral cosmetisch. IP Volume is nauw verwant aan Incrediserve en Quasi Networks. Als we de IP-adressen met CSAM uit

Aug 2019 – nov 2019		Jan 2020 – aug 2020	
Hostingbedrijf	No. URL's	Hostingbedrijf	No. URL's
NFOrce	19,393 (72.31%)	NFOrce	174,652 (93.57%)
IP Volume	5136 (19.15%)	KnownSRV	4581 (2.45%)
LeaseWeb	1013 (3.77%)	IP Volume	4420 (2.36%)
KnownSRV Ltd	752 (2.80%)	LeaseWeb	2013 (1.07%)
Overigen	750 (1.95%)	Overigen	1049 (0.56%)
Totaal	26,819	Totaal	186,715

Tabel 3.1: Aantal CSAM URL's bij Nederlandse hostingbedrijven

Hostingbedrijf	2020	2019	2018
NFOrce	1	1	1
KnownSRV	2	4	
IP Volume	3	2	
Leaseweb	4	3	5
ISPIRIA Networks	5		
Serverius Holding		5	
Incrediserve			2
Swiftwill			3
Quasi Networks			4

Tabel 3.2: Top 5 hostingbedrijven van afgelopen jaren

2018 voor Incrediserve en Quasi Networks opzoeken in de huidige RIPE WHOIS, dan zien we dat ze nu zijn toegewezen aan IP Volume.¹ Als we meenemen dat die verschillende namen verwijzen naar min of meer hetzelfde bedrijf, namelijk het huidige IP-volume, dan was de top 4 in 2018 in wezen dezelfde set providers als in 2019 en 2020, afgezien van één wijziging: Swiftwill verdween en werd door KnownSRV.

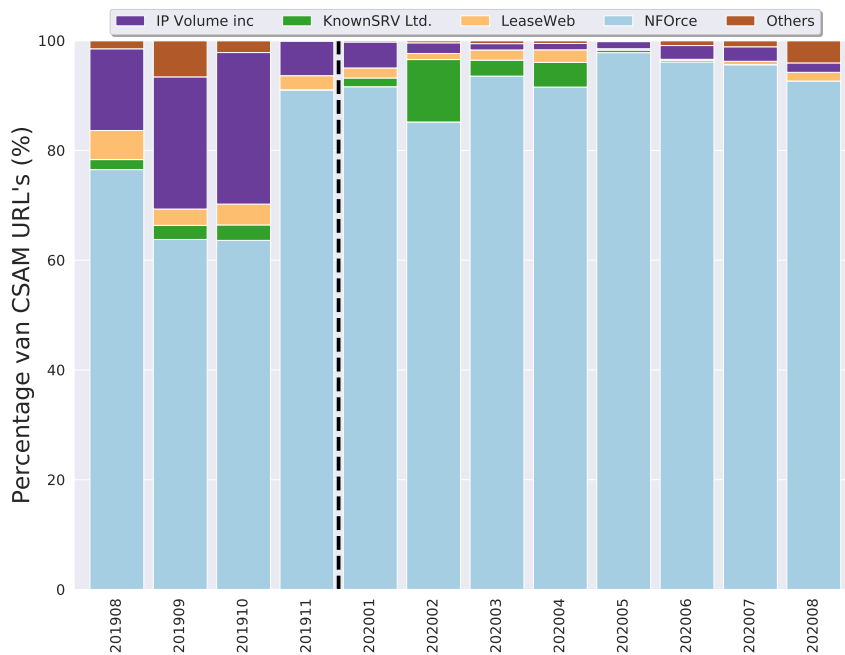
Figuur 3.2 laat zien dat het beeld ook van maand tot maand grotendeels stabiel is. In de laatste maanden van 2019 was NFOrce al hoster van het grootste deel van alle CSAM. Deze verdeling is in 2020 nog verder scheefgetrokken. In de afgelopen maanden is meer dan 90% van de URL's gevonden op domeinen die worden gehost bij NFOrce. In deze periode was de 'traditionele' top 4 niet meer duidelijk zichtbaar op maandniveau. Bijna alles werd gedomineerd door NFOrce. De resterende fractie was erg klein en verdeeld over meer aanbieders dan de drie andere aanbieders uit de top 4. We vermoeden dat deze verandering in het patroon een kortetermijneffect is dat verband houdt met de plotselinge daling van de gegevens afkomstig van het INHOPE-netwerk.

3.3. Verdeling van CSAM over domeinen

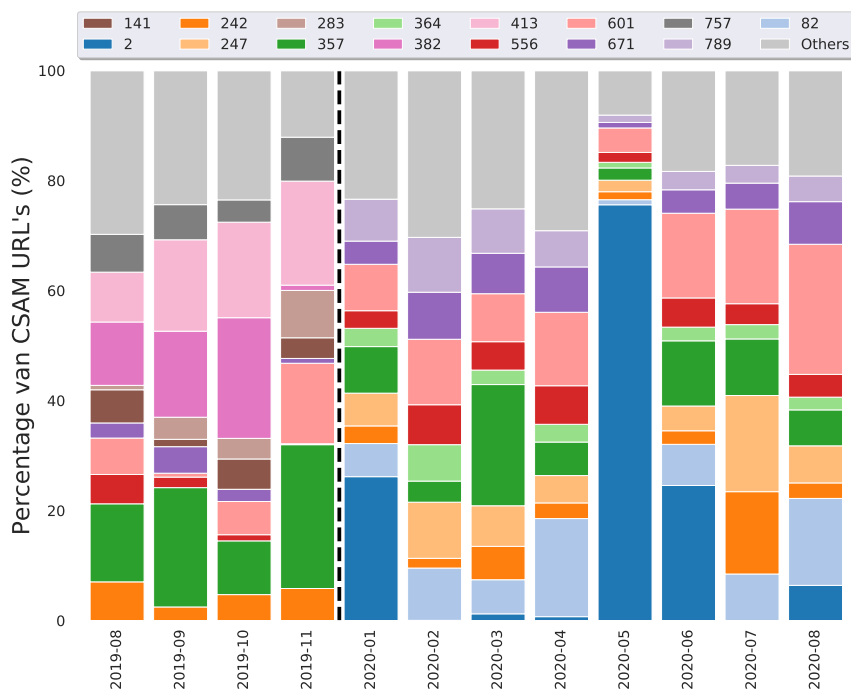
Om de concentratie van CSAM bij hostingbedrijven beter te begrijpen, bekijken we de domeinen waar het meeste materiaal terechtkomt. Uit het jaarverslag van INHOPE en EOKM weten we dat de domeinen met de meeste CSAM image-hosting bedrijven zijn.² Dit patroon geldt zowel wereldwijd als in de Nederland.

¹Zie ook: <https://badpackets.net/a-conversation-with-ripe-ncc-related-quasi-networks-ltd/>

²Zie pagina 29 van het jaarverslag van INHOPE (https://www.inhope.org/media/pages/the-facts/download-our-whitepapers/009c452175-1595854476/annualreport_inhope_2019.pdf) en pagina 10 van het jaarverslag van EOKM (<https://www.eokm.nl/wp-content/uploads/2020/03/Jaarverslag-2019.pdf>).

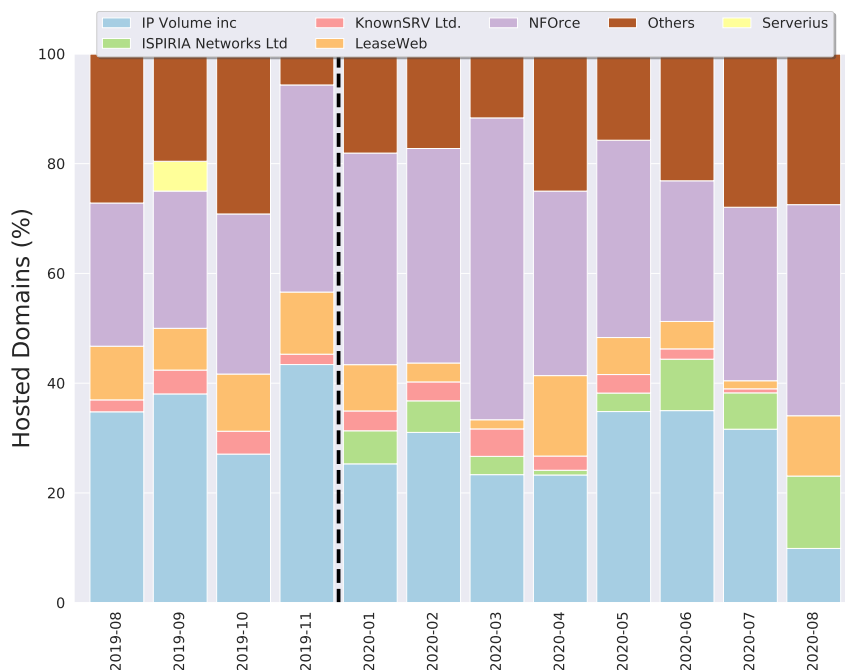


Figuur 3.2: Aantal URL's per hostingbedrijf per maand



Figuur 3.3: Percentage van URL's per domein

Hoe zijn de URL's verdeeld over de verschillende domeinen? Zien we daar ook een patroon van concentratie? Het korte antwoord is: ja. Slechts 10 domeinen zijn goed voor 79,85 % van alle URL's in 2020 en 76,55 % in 2019 – zie figuur 3.3. Het iets minder korte antwoord is echter: de set van domeinen die top 10 vormen, fluctueert nogal van maand tot maand. Deze domeinen bevinden zich natuurlijk meestal binnen de netwerken van de top 4 hostingbedrijven. Zoals te zien is in figuur 3.4, hosten de top 4 providers in 2019 en 2020 meer dan 70 % van alle domeinen die NTD's voor CSAM hebben ontvangen. En slechts twee providers hosten al meer dan 50%.



Figuur 3.4: Percentage dat elk hostingbedrijf heeft van alle domein met CSAM

Ook al is de CSAM geconcentreerd in een klein aantal domeinen, het patroon hier is veel dynamischer dan bij de hostingbedrijven. Allereerst kan het aandeel CSAM per domein behoorlijk veranderen, zelfs in een kort tijdsbestek als een maand. In paragraaf 3.1 zagen we twee pieken in het volume van gerapporteerde CSAM in januari en mei. Als we kijken naar de verdeling van URL's over de domeinen in figuur 3.3, kunnen we zien dat deze pieken werden veroorzaakt door URL's die op slechts één domein (domein #2) werden gedetecteerd. In andere maanden was dat domein nagenoeg afwezig.

Ten tweede is de set domeinen waar de CSAM wordt gehost veranderlijk in de tijd. In tabel 3.3 nemen we de top 10 domeinen in 2020 en kijken naar hun positie in 2019 en 2018. De meeste domeinen van de top 10 van 2020 staan niet in de top 10 in 2019. En geen enkel domein stond in de buurt van de top 10 in 2018. De top 10 van 2020 was verantwoordelijk voor bijna 80% van de URL's in dat jaar, terwijl die zelfde domeinen minder dan 1% voor hun rekening namen in 2018.

In tabel 3.4 maken we dezelfde vergelijking voor de top 10 van 2019. Slechts de helft daarvan stond nog in de top 10 in 2020 en maar vier stonden in de top 10 het jaar ervoor, 2018. Voor de volledigheid presenteren we ook deze vergelijking voor de top 10 uit 2018 (tabel 3.5). Slechts 5 domeinen van die top 10 waren nog aanwezig in 2019 en geen enkele blijft in de top 10 staan in 2020. Een andere manier om deze verandering te kwantificeren is door te zeggen dat van de URL's die in 2020 werden gedetecteerd, slechts 10,28% werd geproduceerd door de domeinen uit de top 10 van 2018.

Kortom, we houden een ietwat raadselachtig beeld over. De domeinen die de URL's herbergen veranderen de afgelopen twee jaar behoorlijk snel, maar de hostingbedrijven met de meeste CSAM blijven min of meer hetzelfde. Met andere woorden, hoewel de domeinen van het ene jaar op het andere veranderen, komen ze op de een of andere manier toch overwegend in het netwerk van dezelfde vier hostingbedrijven terecht.

We vroegen ons af of dit te wijten was aan de grootte van deze providers: hoe groter een provider is, hoe groter de kans dat hij image-hosting-diensten en daarmee CSAM aantrekt. Met andere woorden, domeinen zouden bij deze providers terecht komen omdat zij de dominante spelers op de hostingmarkt zijn. In figuur 3.5 hebben we alle Nederlandse providers in kaart gebracht in termen van grootte, gemeten aan de hand van twee indicatoren: het aantal domeinen dat ze hosten (Y-as) en het aantal IP-adressen dat ze aankondigen (X-as). Dus hoger en meer naar rechts in de plot betekent groter. We kunnen zien dat de vier providers die de meeste CSAM-domeinen en URL's hosten qua grootte over de markt verspreid zijn. Ze zijn niet de grootste providers. Slechts één provider, Leaseweb, zit qua omvang aan de

Domein	2020			2019			2018		
	Rang	Totaal URL's	%	Rang	Totaal URL's	%	Rang	Totaal URL's	%
2	1	49,030	26.27	58	16	0.059	75	2	0.017
601	2	20220	10.83	4	1680	6.27	-	-	-
357	3	14343	7.68	1	4635	17.29	-	-	-
82	4	13986	7.49	-	-	-	127	1	0.008
247	5	12011	6.43	16	298	1.11	-	-	-
789	6	9735	5.21	12	507	1.89	42	8	0.07
671	7	8972	4.80	9.0	766	2.85	36	12	0.10
556	8	8115	4.34	10	635	2.37	19	54	0.47
242	9	7163	3.83	6	1337	4.9	-	-	-
364	10	5459	2.92	23	201	0.75	34	14	0.12
Totaal		14,9034	79.85		10,075	37.60		91	0.79

Tabel 3.3: De positie van de top 10 domeinen uit 2020 in andere jaren

Domein	2020			2019			2018		
	Rang	Totaal URL's	%	Rang	Totaal URL's	%	Rang	Totaal URL's	%
357	3	14343	7.68	1	4635	17.29	-	-	-
413	11	4982	2.66	2	4049	15.11	5	985	8.64
382	21	1185	0.63	3	3614	13.48	60	3	0.02
601	2	20220	10.83	4	1680	6.27			
757	14	3803	2.03	5	1665	6.21	1	1626	14.27
242	9	7163	3.83	6	1337	4.99		-	-
141	12	3994	2.13	7	1123	4.19	6	788	6.91
283	13	3900	2.08	8	1035	3.86	2	1582	13.88
671	7	8972	4.80	9	766	2.85	36	12	0.1
556	8	8115	4.34	10	635	2.37	19	54	0.47
Totaal		76,677	41.08		20,539	76.65		5050	44.33

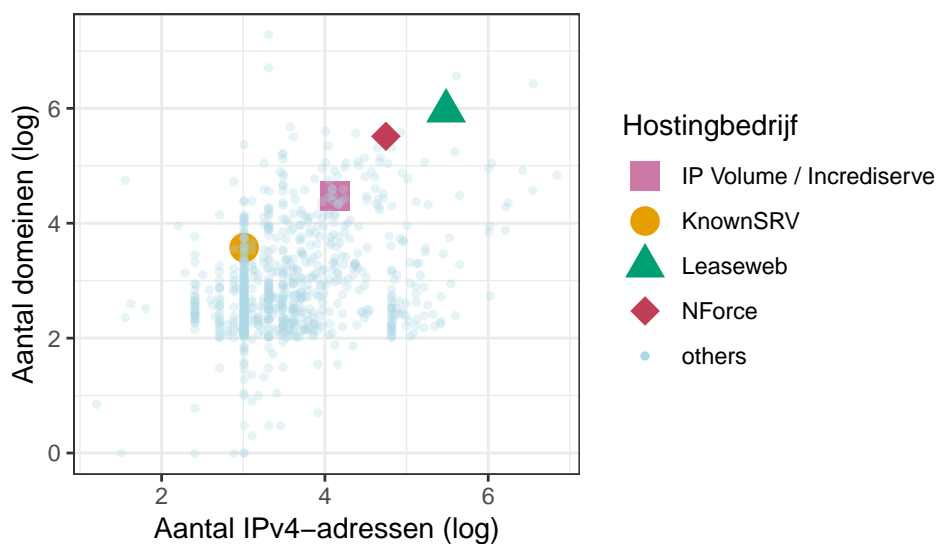
Tabel 3.4: De positie van de top 10 domeinen uit 2019 in andere jaren

bovenkant en juist deze provider heeft eigenlijk een veel kleiner aandeel CSAM-domeinen en URL's dan NForce. En een andere aanbieder van de top 4, KnownSRV, is eigenlijk weer relatief klein, vergeleken met de gemiddelde omvang in de markt. Met andere woorden, grootte is niet de verklaring voor het concentratiepatroon van CSAM dat we zien.

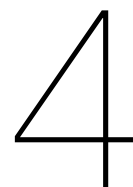
Een andere verklaring zou kunnen zijn dat verschillende image-hostingdomeinen in feite van dezelfde eigenaar zijn. In sommige gevallen kunnen we dit zien aan de WHOIS-records en abuse-contacten voor de domeinen: verschillende domeinen hebben dezelfde contactgegevens. Als een eigenaar zijn sites al bij een specifieke provider heeft opgezet, is het logisch dat daar ook nieuwe sites van dezelfde eigenaar worden opgezet. Kortom, als de nieuwe domeinen eigendom zijn van dezelfde bedrijven die de oudere domeinen exploiteren, dan zou het logisch zijn dat de wijzigingen in de lijst met domeinen met CSAM nog steeds bij dezelfde top 4 providers terechtkomen. Deze verklaringen spelen wellicht een rol maar ze bieden nog geen afdoende antwoord.

Domein	2020			2019			2018		
	Rang	Totaal URL's	%	Rang	Totaal URL's	%	Rang	Totaal URL's	%
757	14	3803	2.03	5	1665	6.21	1	1626	14.27
283	13	3900	2.08	8	1035	3.86	2	1582	13.88
474	-	-	-	-	-	-	3	1128	9.90
373	-	-	-	-	-	-	4	1073	9.41
413	11	4982	2.66	2	4049	15.11	5	985	8.64
141	12	3994	2.13	7	1123	4.19	6	788	6.91
214	27	447	0.23	20	241	0.89	7	566	4.96
218	-	-	-	-	-	-	8	543	4.76
365	-	-	-	-	-	-	9	483	4.24
235	15	2070	1.10	11	591	2.20	10	456	4
Totaal		19,196	10.28		8704	32.48		9230	81

Tabel 3.5: De positie van de top 10 domeinen uit 2018 in andere jaren



Figuur 3.5: Grootte van hostingbedrijven uit de top 4 vergeleken met de andere hostingbedrijven in Nederland, geplott in termen van het aantal IP-adressen dat ze adverteren en het aantal domeinen dat ze hosten, beide in log-schaal



Snelheid van CSAM verwijdering

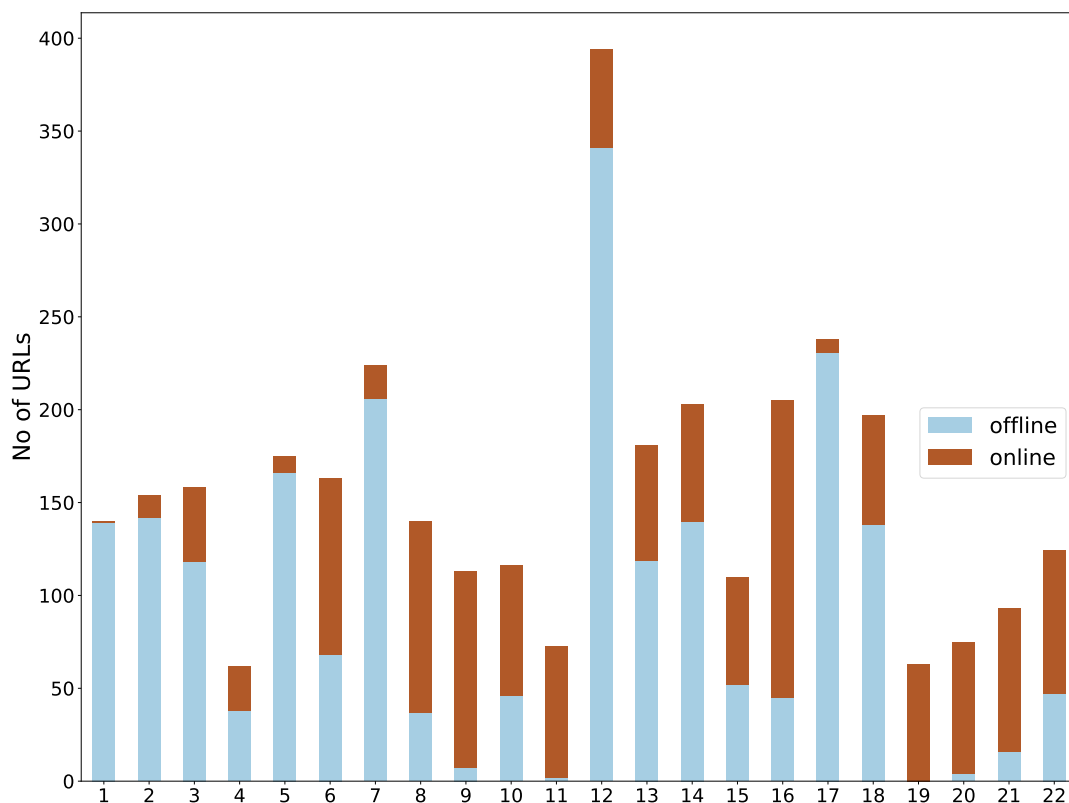
In de publiek-private werkgroep die CSAM-hosting wil bestrijden, is een norm afgesproken: de hostingbranche moet CSAM binnen 24 uur na het indienen van een NTD-verzoek verwijderd hebben. In augustus 2020 heeft EOKM handmatig gecontroleerd hoe snel providers of domeineigenaren CSAM verwijderden na ontvangst van een NTD. Deze meting werd uitgevoerd op basis van een steekproef van URL's opgesteld door de TU Delft. In dit hoofdstuk bespreken we de resultaten. In welke mate halen providers en domeineigenaren de 24-uursvenster voor het verwijderen van CSAM?

4.1. Meten van snelheid

Hoe lang duurt het voordat een domeineigenaar of hostingbedrijf het CSAM-materiaal heeft verwijderd? In augustus 2020 voerden EOKM-medewerkers handmatige controles uit bij een steekproef van nieuwe en bevestigde CSAM URL's. Vanwege het lagere aantal URL's dat die maand via ICCAM binnenkwam, konden op sommige dagen alle URL's worden gecontroleerd. Op andere dagen moesten we een selectie nemen van de nieuwe URL's, om het EOKM-personeel niet te overbelasten. We hebben een gestratificeerde steekproef per provider gedaan, zodat zelfs providers met een klein aantal URL's in de steekproef zouden worden opgenomen. Een niet-gestratificeerde willekeurige steekproef zou in feite hebben betekend dat we vooral URL's hadden gecontroleerd die bij NForce worden gehost en dat we geen betrouwbare meting konden krijgen voor de andere bedrijven.

Elke URL in de steekproef werd een dag later, minimaal 24 uur na de NTD, opnieuw bezocht. Als het materiaal nog steeds online stond, werd er een tweede bezoek gepland voor de volgende dag, na minimaal 48 uur. We hebben de verwijderingssnelheid gemeten in drie categorieën: (i) inhoud verwijderd binnen 24 uur, dus binnen de industrie-norm; (ii) inhoud verwijderd tussen 24-48 uur; en tot slot (iii) inhoud die 48 uur en langer online bleef. We hebben ook enkele URL's gecontroleerd die naar de politie zijn gestuurd en die buiten het reguliere NTD-proces werden afgehandeld.

Figuur 4.1 toont het aantal URL's dat door de TU Delft is geselecteerd en dagelijks handmatig is gecontroleerd door EOKM. De meting werd uitgevoerd in augustus 2020. Het aantal URL's verschilt per dag omdat er voor bepaalde dagen meer URL's zijn geverifieerd als CSAM. We hebben ook eerdere gecontroleerde URL's aan ten minste één extra controle onderworpen voordat we deze voor verdere controles hebben verwijderd. In totaal zijn 2114 URL's een of meerdere malen gecontroleerd.



Figuur 4.1: Aantal URL's dat per dag werd gecontroleerd in augustus 2020

De NTD wordt naar het hostingbedrijf gestuurd en, indien EOKM ook een abuse-contactadres voor het domein heeft, ook naar de domeineigenaar. Hoewel we van buitenaf niet kunnen zien wie de verwijderingsactie heeft ondernomen, de hoster of de domeineigenaar, is de standaardpraktijk dat de hoster de NTD doorstuurt naar zijn klant, de domeineigenaar, en dat die laatste de verwijdering uitvoert. In dit scenario is de verwijderingssnelheid van een hostingbedrijf afhankelijk van de verwijderingssnelheid van de domeineigenaar.

Om hiermee rekening te houden, laten we zien hoeveel domeinen geassocieerd waren met de URL's in Tabel 4.1. (Er zijn in totaal 70 domeinen, maar twee van deze domeinen werden parallel gehost door twee providers. Dit brengt het totale aantal in de tabel op 72.) Zoals gezegd heeft EOKM voor sommige domeinen geen abuse-contactinformatie voor het domein. In die gevallen krijgt alleen het hostingbedrijf een NTD toegestuurd. Hoewel EOKM in de steekproef van augustus slechts abuse-contactinformatie had voor van 35 van de 70 domeineigenaren, omvatte deze set van 35 ongeveer 94% van de URL's. Dus voor de meeste URL's werd contact opgenomen met zowel het hostingbedrijf als de domeineigenaar.

4.2. Verwijderingssnelheid per hostingbedrijf

We meten de uptime voor een steekproef van 2114 URL's die naar 15 hostingbedrijven en 70 domeinen zijn verzonden. Dit was een steekproef van alle inkomende URL's in augustus 2020. De steekproef was gestratificeerd per hostingbedrijf ook bedrijven met een klein aantal URL's mee te kunnen nemen in de meting. Het is belangrijk op te merken dat vanwege deze opzet van de steekproef tabellen met resultaten niet moet worden gelezen als een indicatie van het totale aantal URL's, noch als de relatieve hoeveelheid CSAM per hostingbedrijf of domein.

Voor één provider uit de top 4, IP Volume, konden we de inkomende URL's niet meenemen in de steekproef vanwege een speciale overeenkomst tussen EOKM en de politie. In de overeenkomst staat dat alle URL's van IP Volume zouden worden gedeeld met de politie, in plaats van opgenomen in het normale NTD-proces. Er werden later wel NTD's verzonden, maar niet via de reguliere procedure en

Hostingbedrijf	Aantal domeinen	Domeinen met abuse-contact
NFOrce Entertainment B.V.	33	27 (81.88%)
INTERNET IT COMPANY INC	9	0 (0%)
LeaseWeb Netherlands B.V.	9	6 (66.66%)
ISPIRIA Networks Ltd	7	0 (0%)
Host Sailor Ltd.	3	0 (0%)
ON-LINE DATA LTD.	2	0 (0%)
Abelohost BV	1	0 (0%)
Contel OOO	1	0 (0%)
DataWeb Global Group B.V.	1	0 (0%)
EuroByte LLC	1	0(0%)
FXOBO Private Enterprise	1	0 (0%)
IP Volume inc	1	1 (100%)
ServerStack, Inc.	1	0(0%)
Webzilla B.V.	1	0(0%)
xTom GmbH	1	1 (100%)

Tabel 4.1: Verdeling van de domeinen uit de steekproef over de hostingbedrijven

het was niet mogelijk om ze mee te nemen in de meting. We konden slechts 3 URL's opnemen die wel het normale proces hebben doorlopen. Om deze reden zijn de URL's voor IP Volume sterk ondervertegenwoordigd in de meting. Dit geeft niet het werkelijke aantal URL's weer dat bij die provider wordt gehost.

Wat de resultaten betreft, hebben we vastgesteld dat 84% van alle CSAM binnen 24 uur wordt verwijderd. Nog eens 12% wordt tussen 24-48 uur verwijderd en 4% blijft 48 uur of langer online. Sommige URL's in de laatste categorie zijn meerdere keren opnieuw bezocht. Het patroon was zeer gevarieerd. Sommige gingen een dag later offline, na 72 uur, andere bleven online gedurende de hele periode dat EOKM controles uitvoerde, tot een maximum van 21 dagen.

In tabel 4.2 laten we de verwijderingssnelheid per hostingprovider zien. Alleen providers met een zeer klein aantal URL's slagen erin om perfect aan de 24-uurs norm te voldoen. De top 5 aanbieders in augustus 2020 verschillen sterk in het naleven van deze norm. In het ergste geval werd 0% van alle CSAM binnen 24 uur verwijderd uit het netwerk van de provider. In het beste geval werd 89% binnen 24 uur verwijderd. Na 48 uur was het grootste deel van alle URL's verwijderd, behalve "ISPIRIA Networks Ltd". Bij zes providers blijft een deel van het materiaal meer dan 48 uur online.

4.3. Verwijderingssnelheid per domein

Omdat hostingproviders doorgaans afhankelijk zijn van domeineigenaren voor het verwijderen van de CSAM, analyseren we de gegevens ook op het niveau van domeinen. In ons voorbeeld hebben we 2114 URL's verdeeld over 70 domeineigenaren. In Tabel 4.3 presenteren we de resultaten voor de 20 domeinen met de meeste URL's. Alle andere domeinen worden gecombineerd in de laatste rij.

Onze resultaten laten zien dat 11 van de 20 domeinen 90% of meer van het materiaal binnen 24 uur verwijderen. De overige 9 domeinen verwijderen een aanzienlijk deel tussen 24 en 48 uur. Meer dan 90 % van alle CSAM wordt binnen 48 uur verwijderd. Een belangrijk punt om op te merken is dat 19 van de 20 domeinen een abuse-contactpunt aangeven. In sommige gevallen wordt hetzelfde e-mailadres voor abuse gebruikt voor meerdere domeineigenaren.

Hostingbedrijf	Totaal Url's	Offline binnen 24u	Offline tussen 24-48h	Online meer dan 48u
NForce	1969	1696 (86.13%)	220 (11.17%)	53 (2.69%)
INTERNET IT	35	31 (88.57%)	1 (2.85%)	3 (8.57%)
Host Sailor Ltd.	34	1 (2.94%)	32 (94.11%)	1 (2.94%)
Contel OOO	32	0 (0%)	32 (100%)	0 (0%)
ISPIRIA Networks Ltd	24	4 (16.6%)	4 (16.66%)	16 (66.66%)
LeaseWeb	22	19 (86.36%)	0 (0%)	3 (13.63%)
xTom GmbH	13	13 (100%)	0 (0%)	0 (0%)
ON-LINE DATA LTD.	7	6 (85.71%)	1 (14.28%)	0 (0%)
IP Volume inc	3	0 (0%)	3 (100%)	0 (0%)
DataWeb Global	2	1 (50%)	1 (50%)	0(0%)
ServerStack	1	0 (0%)	0 (0%)	1 (100%)
Abelohost BV	1	1 (100%)	0 (0%)	0 (0%)
EuroByte LLC	1	1 (100%)	0 (0%)	0 (0%)
FXOBO Private Enterprise	1	1 (100%)	0 (0%)	0 (0%)
Webzilla B.V.	1	1 (100%)	0 (0%)	0 (0%)

Tabel 4.2: Meting van verwijderingssnelheid CSAM per hostingbedrijf op basis van een gestratificeerde steekproef

4.4. Verwijderingssnelheid van de URL's voor politie

Voor URL's die naar de politie worden gestuurd, wordt de NTD later naar de hostingprovider en domeineigenaar gestuurd. Zowel de politie als het EOKM sturen een NTD. We hebben 38 URL's van deze URL's gecontroleerd. Geen van de URL's ging offline binnen 48 uur nadat ze naar de politie waren gestuurd. Gezien de vertraagde NTD is dit te verwachten. Uiteindelijk waren er 21 (55 %) URL's die op een later tijdstip offline waren bij de controle van het EOKM.

4.5. Tot slot

Samenvattend stellen we vast dat er veel verschil is in hoe snel providers en domeineigenaren CSAM-materiaal verwijderen. Sommigen slagen erin om meer dan 90% verwijderd te krijgen binnen de 24-uurs norm. Anderen hebben er 48 nodig om het meeste materiaal te verwijderen. Slechts een klein deel van CSAM blijft na 48 uur online. We weten niet waardoor deze URL's blijven bestaan. Deze providers en domeineigenaren hebben soms meer dan 90 % van de URL's binnen 24 uur verwijderd, dus ze negeren de NTD's niet categorisch. Toch blijft een kleine fractie van de URL's langer online.

Domein-nummer	Totaal URL's	Offline binnen 24u	Offline tussen 24-48u	Online meer dan 48u
82	493	485 (98.37%)	8 (1.62%)	0 (0%)
601	316	288 (91.13%)	26 (8.22%)	2 (0.63%)
247	212	166 (78.30%)	45 (21.2%)	1 (0.47%)
357	184	167 (90.7%)	7 (3.80%)	10 (5.43%)
556	111	106 (95.49%)	5 (4.50%)	0 (0%)
671	106	53 (50%)	51 (48.11%)	2 (1.88%)
789	94	40 (42.55%)	54 (57.44%)	0 (0%)
355	71	71 (100%)	0 (0%)	0 (0%)
364	53	34 (64.15%)	18 (33.96%)	1 (1.88%)
242	52	49 (94.23%)	0 (0%)	3 (5.76 %)
413	51	51 (100%)	0 (0%)	0 (0%)
283	50	35 (70%)	6 (12%)	9 (18.0%)
715	32	0 (0%)	32 (100%)	0 (0%)
235	28	19 (67.85%)	0 (0%)	9 (32.14%)
551	25	25 (100%)	0 (0%)	0 (0%)
141	24	17 (70.83%)	0 (0%)	7 (29.16%)
46	21	21 (100%)	0 (0%)	0 (0%)
757	21	17 (80.95%)	0 (0%)	4 (19.04%)
7	17	16 (94.11%)	0 (0%)	1 (5.88%)
2	14	14 (100%)	0 (0%)	0 (0%)
Others	139	101 (72.66%)	10 (7.19%)	28 (20.14%)

Tabel 4.3: Meting van verwijderingssnelheid CSAM per domein op basis van een gestratificeerde steekproef

5

Conclusie

De CSAM Hosting Monitor is een instrument in een bredere set van publiek-private acties om de hosting van CSAM in Nederland tegen te gaan. Het helpt de belanghebbenden door transparant te maken welke hostingproviders het meeste materiaal in hun netwerken hebben, hoe de locatie van het materiaal in de loop van de tijd verandert (of niet) en hoe snel de bedrijven het materiaal verwijderen.

In principe zou de monitor ook de voortgang kunnen tonen van de algehele aanpak om CSAM-hosting te bestrijden. Simpel gezegd: daalt de hoeveelheid CSAM die in Nederland wordt gehost? En zo ja, welke providers en domeinen verbeteren het meest en welke niet? Dit zijn cruciale vragen om te bepalen wat de vervolgacties moeten zijn in de publiek-private samenwerking. In de huidige fase van de monitor is het echter te vroeg om een dergelijke evaluatie te maken. Zoals we op verschillende punten in dit rapport hebben besproken, zijn er veel op elkaar inwerkende factoren die samen het volume van CSAM in de Nederlandse hostingnetwerken bepalen. Deze factoren kunnen niet worden ontward. Dus ja, er is een duidelijke neerwaartse trend in het volume van CSAM sinds de Minister van Justitie en Veiligheid in juni 2020 aankondigde dat hij “een fictieve zandloper [had] omgedraaid” – oftewel: dat de klok tikte voor Nederlandse hostingbedrijven en hun klantendomeinen om actie te ondernemen tegen CSAM.¹ De neerwaartse trend is echter moeilijk te beoordelen, omdat we niet weten welke veranderingen er zijn opgetreden in de wijze waarop INHOPE-leden de gegevens verzamelen. Het is op dit moment nog onduidelijk welk deel van de daling wordt bewerkstelligd door de publiek-private acties.

Een andere belangrijke tegenmaatregel was de lancering door EOKM en Web-IQ van de zogenaamde hashcheckservice. Sommige hostingproviders dringen er bij hun klanten die worstelen met CSAM op hun domein aan om deze hashcheckservice te gebruiken.² De lijst met domeinen die de hashcheckservice gebruiken is vertrouwelijk, evenals het aantal checks dat domeinen bij de service uitvoeren. We hebben EOKM voorzien van een lijst met 20 domeinen met de meeste CSAM-URL's in de afgelopen zes maanden (maart – augustus 2020). Van die lijst hebben er 4 de dienst overgenomen. Ze staan in de top 20 op positie 4, 10, 12 en 17 en hosten ongeveer 13 % van alle URL's uit die periode. (Als we FQDN's tellen, zoals we in de rest van het rapport hebben gedaan, dan is het aantal 6 domeinen uit de top 20 en worden ze gerangschikt op 4, 10, 16, 17, 18 en 19.) De domeinen die zich hebben aangemeld voor de hashcheckservice bevinden zich binnen het netwerk van twee hostingproviders: drie bij NForce en één bij Leaseweb.

¹Zie <https://www.rijksoverheid.nl/documenten/kamerstukken/2020/07/07/tk-voortgangsbrief-aanpak-online-seksueel-kindermisbruik-en-kindersekstoerisme>.

²Zie voor meer informatie de recente brief aan de Tweede Kamer van de Minister van Justitie en Veiligheid: <https://www.rijksoverheid.nl/documenten/kamerstukken/2020/07/07/tk-voortgangsbrief-aanpak-online-seksueel-kindermisbruik-en-kindersekstoerisme>. Zie ook: <https://www.meldpunt-kinderporno.nl/over-ons/hash-database/>

Ook hier zijn de effecten van deze preventieve maatregelen nog niet zichtbaar, vanwege de vele factoren die van invloed zijn op de hoeveelheid CSAM die in Nederland wordt gedetecteerd. We hoorden ook, anekdotisch, dat een provider bepaalde klant domeinen die worstelen met CSAM-uploads had gevraagd om te vertrekken. Het beeld van hoe CSAM over domeinen is verdeeld, is echter behoorlijk dynamisch van de ene maand op de andere. In figuur 3.3 zien we veel variatie en geen duidelijke verandering in de set domeinen met de meeste CSAM. We kunnen dus (nog) niet zien dat de klanten die de aanbieder verlaten een impact hebben gehad. Om een lang verhaal kort te maken: gezien deze natuurlijke variabiliteit kunnen we nog geen duidelijke resultaten zien van de meer proactieve maatregelen van hostingproviders.

Voor de monitor is de volgende stap om CSAM-hosting over een langere periode te observeren, zodat we hopelijk de effecten van de interventies beter kunnen identificeren. In ieder geval zal het ons in staat stellen om transparantie te scheppen rondom welke bedrijven betrokken zijn bij het probleem – en dus ook bij de oplossing. Voor de verdere ontwikkeling van de monitor zelf werken we samen met EOKM en de ontwikkelaars van SCART om de functionaliteit voor het meten van de verwijderingssnelheid te verbeteren. De prioriteit is om deze meting meer geautomatiseerd te maken. Dat gezegd hebbende zullen er bepaalde technische beperkingen volledig geautomatiseerde tracking onmogelijk maken. Een aanzienlijk deel van de pagina's waar afbeeldingen gehost worden bevat bijvoorbeeld CAPTCHA's om de automatische verzameling van pagina-inhoud te voorkomen (zie paragraaf 2.5). Om deze reden voorzien we dat de monitor zal moeten streven naar een mix van automatisering en handmatige controle door EOKM-medewerkers. Op die manier zouden regelmatig metingen kunnen worden gedaan om te controleren of aan de 24-uurs norm wordt voldaan.