# CSAM Hosting Monitor

## Report March 2022

Qasim Lone, Carlos H. Gañán, Michel van Eeten


TU Delft

# Contents


Executive Summary . . . . . 2

**1 Introduction** **3**

**2 Methodology** **5**

2.1 CSAM Data . . . . . 5

2.2 CSAM Hosting Monitor . . . . . 6

2.3 Tracking CSAM Over Time . . . . . 7

2.4 NTD Takedown Speed . . . . . 8

2.5 Corroborating NTD data with providers . . . . . 9

2.6 Limitations . . . . . 10

**3 CSAM Takedown** **13**

3.1 Measuring Takedown Speed . . . . . 13

3.2 Takedown Speed per Hosting Provider . . . . . 14

3.3 Takedown Speed per Domain . . . . . 15

3.4 In Sum . . . . . 16

**4 CSAM Landscape in the Netherlands** **19**

4.1 CSAM Volume . . . . . 19

4.2 Distribution of CSAM Across Hosting Providers . . . . . 20

4.3 Distribution of CSAM Across Domains . . . . . 21

**5 Conclusion** **27**

# Executive Summary

Since October 2018, TU Delft has been developing and operating a monitor to analyze where Child Sexual Abuse Material (CSAM) is being hosted in the Netherlands. This was done in close collaboration with EOKM. Looking back at 2021, the monitor's main findings are:

1. In the public-private roundtable to fight CSAM hosting, started in 2018, a norm was agreed upon: the hosting industry should remove CSAM within 24 hours after the NTD. In December 2021 – January 2022, EOKM conducted manual checks how fast providers or domain owners removed CSAM material after receiving an NTD, based on a sample of URLs prepared by TU Delft. We checked 6268 URLs located at 33 hosting providers and 70 domains. We found that 87% of all CSAM is removed in 24 hours. A further 3% is removed between 24-48 hours, and 10% remains online for 48 hours or longer – in some cases more than a week.

2. Some providers have achieved an impressive responsiveness in removing CSAM after receiving an NTD. While NForce got the most URLs reported to it, it managed to remove 100% within 24 hours. The number 2 provider with most CSAM, KnowSRV, managed to remove 84% in 24 hours, but 15% remained online for longer than 48 hours. One of the new providers in the top 5, Telegram Messenger Inc (no relation to the instant messaging service), did extremely poorly: only 6% was removed and a staggering 94% of all URLs stayed online for 48 hours and longer.

3. In addition to takedown actions, some providers are taking further pro-active measures in preventing CSAM being hosted on their network. NForce had by far the largest share of CSAM chose to sever relationships with legitimate customers whose services were being abused to share CSAM. While these customers were very diligent in removing the material rapidly, it was hard, if not impossible, for them to prevent the material from being uploaded in the first place. These customers have left NForce and migrated their domains to outside the Netherlands.

4. The amount of CSAM detected and notified about in the Netherlands has peaked in 2021, yet there is no clear trend in terms of growth or decline. Notwithstanding the fluctuations in volume, the material remains concentrated with a small number of hosting providers: Four providers have the overwhelming majority of all URLs: 97.37% in 2019, 98.49% in 2020 and 95,97% in 2021. The composition of which four providers is partially stable over these two years. NForce and KnownSRV are consistently present as the largest contributors hosted CSAM. Other providers, like Leaseweb NL and IP Volume have dropped out of the top 5.

5. We did observe a dramatic change in the CSAM volume at NForce during the end of our measurement period. This provider has had over 90% of all CSAM URLs in 2020 and 2021. Yet, in the last four months of 2021, we observed a steady reduction of CSAM, going to zero reported URLs in the course of January 2022. The domains that hosted the CSAM have all left NForce. The company reported to us that it had asked these customers to move elsewhere.

6. Compared to providers, there is much more change at the level of domains within these provider networks host the CSAM. Of the top 10 domains with the most CSAM in 2021, only 4 domains were in the top 10 in 2020 and only 1 domain in 2019. Even in the course of a year, the presence of CSAM on a domain fluctuates wildly, with some domains receiving huge spikes in volume, only to disappear not much later.

7. Some customer domains have been encouraged by their hosting provider to adopt the hash-check service offered by EOKM. This is a preventative measure that would help to reduce the uploading of known CSAM material to the websites, but it cannot fully prevent it. We cannot evaluate the

effect of the hash-check service, because it is kept confidential which domains have signed up to the service.

8. One way to summarize the results of the past two years of public-private action against CSAM hosting in the Netherlands is to say that the providers that have been the focus of the public-private initiative to combat CSAM hosting, have been responsive. They already took action, but many increased their efforts. Takedown speeds are (very) high and in some cases providers have parted ways with legitimate customers who were being abused by people sharing CSAM. It is unlikely that these effects would have happened without the public-private initiative. The providers who were not in focus of the policy, because they only recently became more visible, are much less responsive or, in the case of Telegram Messenger Inc, not responsive at all. NForce dominated the Dutch CSAM hosting landscape in the past two years. If it manages to sustain its drastic reduction of CSAM, this success will mean that other providers will come into focus. The success of the policy also has ambiguous effects. The domains that migrate to outside the Netherlands might be less reachable for NTDs and less active in removal than during their presence in the Netherlands.

# 1

# Introduction

For years, the Netherlands has been identified as a prime location for the hosting of child sexual abuse material (CSAM). Data on the hosting of CSAM is collected via INHOPE, the global network of national hotlines to combat CSAM. In their annual report over 2020, INHOPE listed the Netherlands as the second country worldwide and the leading hosting location in Europe, harboring 90% of all CSAM in Europe.[1]

As we noted in our previous report, these problems are not caused directly by the hosting companies. Their customers might operating legitimate image-hosting websites where unknown users upload and share CSAM. It is increasingly recognized, however, that providers do have to take part of the responsibility to combat the hosting of CSAM and make their networks more safe and secure. Providers are at least expected to respond swiftly to Notice & Takedown (NTD) requests, which are issued by INHOPE hotlines and which ask them to remove the CSAM material from their network. The response from providers varies widely, ranging from vigilant to slow to negligent. In rare and extreme cases, providers even offer 'bulletproof' services for criminals, knowingly facilitating the hosting of CSAM.

In response to the prominent presence of CSAM in Dutch hosting networks, a 2018 roundtable (March 27, 2018) brought together a broad coalition of representatives from the Dutch government, industry and academia.[2] The goal of this public-private initiative was to identify more effective ways to combat the hosting of CSAM in the Netherlands. This led to the articulation of several shared "ambitions". A key ambition with wide industry support is to create more transparency regarding which industry providers are involved in hosting CSAM and how swiftly they remove CSAM from their network once they are notified of its presence via an NTD request. As a result of the roundtable process and the government initiatives, the first public TU Delft report that monitored the hosting of CSAM in the Netherlands was sent to parliament in October 2020 by the minister of Justice and Security.[3] The current report is a follow-up of that initial report. Some of the text on the methodology is included again for the reader's convenience. Both reports are based on data from Expertisebureau Online Kindermisbruik (EOKM), the Dutch national hotline and member of INHOPE. We report on the patterns in volume, location, providers and domains. Overall, we will answer the following research questions:

1. Which Dutch providers host CSAM materials in their network? The first objective is to identify where CSAM materials are hosted and by whom. We will combine the data from EOKM with IP

[1] INHOPE Annual Report 2020, available online at: https://inhope.org/media/pages/the-facts/download-our-whitepapers/annual-report/bb4dd3cdc3-1628156678/inhope-annual-report-2020.pdf

[2] https://zoek.officielebekendmakingen.nl/kst-31015-150.pdf

[3] https://https://www.rijksoverheid.nl/documenten/rapporten/2020/10/08/csam-hosting-monitor---rapport-september-2020

addresses, AS numbers and WHOIS registration data to identify the hosting providers operating the networks where the CSAM is located.

2. How does the location of CSAM material in the Dutch market change over time? We will track the concentrations of CSAM overtime via several snapshots taken during the development of the monitor in the period October 2018 - August 2020.

3. How can the distribution of CSAM material over providers and domains be understood? We will take the output of questions 1 and 2 and interpret these findings in the context of the overall Dutch hosting landscape.

4. How fast do the hosting providers and their customers respond to an NTD request to remove CSAM? We report on what portion of CSAM is removed within 24 hours of the NTD request and how different providers perform in terms of meeting the 24-hour removal norm. For this, manual checks were conducted by EOKM in August 2020, in collaboration with TU Delft.

The remainder of this report consists of the following parts. We first explain the methodology of the CSAM hosting monitor (Chapter 2). Then, we present the results on the location of CSAM in the Dutch hosting landscape and identify how this pattern has changed over time (Chapter 4). We then turn to the results on the measurement of the NTD takedown speed in August 2020 (Chapter 3). We end with a few concluding notes on the next steps for the monitor.

# 2

# Methodology

The main goal of this project is to monitor the location of CSAM material across the Dutch hosting market and to track the speed of responding to notice and takedown requests (NTD). We first outline what data we use on the detection of CSAM and how to attribute CSAM material to the relevant hosting provider and domain.

## 2.1. CSAM Data

CSAM material is located at URLs (Uniform Resource Locators) – simply put, a link to a webpage or picture. The URL can point to a single picture or it can link to a page that contains more CSAM. These URLs are processed and checked by EOKM. Once checked, an NTD is sent to the hosting provider and domain owner asking them to remove the material at the specified URL. The monitor tracks the number of NTDs sent to each provider and domain owner, as well as the speed with which the CSAM material was taken offline. The latter is based on manual checks done on a sample of NTDs/URLs – as will be discussed in Chapter 4.

The core data on the online location of CSAM is provided by the EOKM. EOKM receives reports of suspected CSAM from volunteers via its “meldpunt” (Dutch hotline) and from other INHOPE member hotlines through a system called ICCAM. A few of the hotlines contribute most of the reports, because they have the mandate and tools to conduct scans, thus discovering more CSAM than other hotlines. The Canadian hotline has been one of the main contributors.

In March 2020, the Canadian hotline left the INHOPE network. This means that EOKM is no longer receiving those CSAM reports and thus sends no notifications for them. The Canadian hotline sends its own NTDs to providers. This development has two implications: first, Dutch providers now get some NTDs from EOKM and some from the Canadian hotline; second, our report is undercounting known CSAM for Dutch providers, since we only have access to the EOKM data. Recently, the Canadian hotline has begun to share the NTD messages with EOKM. For a future report, we might be able to integrate this data into the monitor.

EOKM staff checks the reported URLs and verifies that it does in fact contain CSAM material. Next, it checks that it is hosted in the Netherlands. If not, the URL is sent via ICCAM to the relevant hotline in another country. If the reported material is confirmed as being CSAM and as being hosted in the Netherlands, then EOKM then decides on whether to send the link to the police (law enforcement agencies, LEA) or to issue an NTD request to the hosting provider, domain owner or, in some cases, the registrar for the domain name. When EOKM sends a link to the police, it will also issue an NTD, but with a bit of a delay, typically one day later. There are several reasons to send a link to the police. It could be previously unknown material (rather than known CSAM), it could be material with a possible Dutch

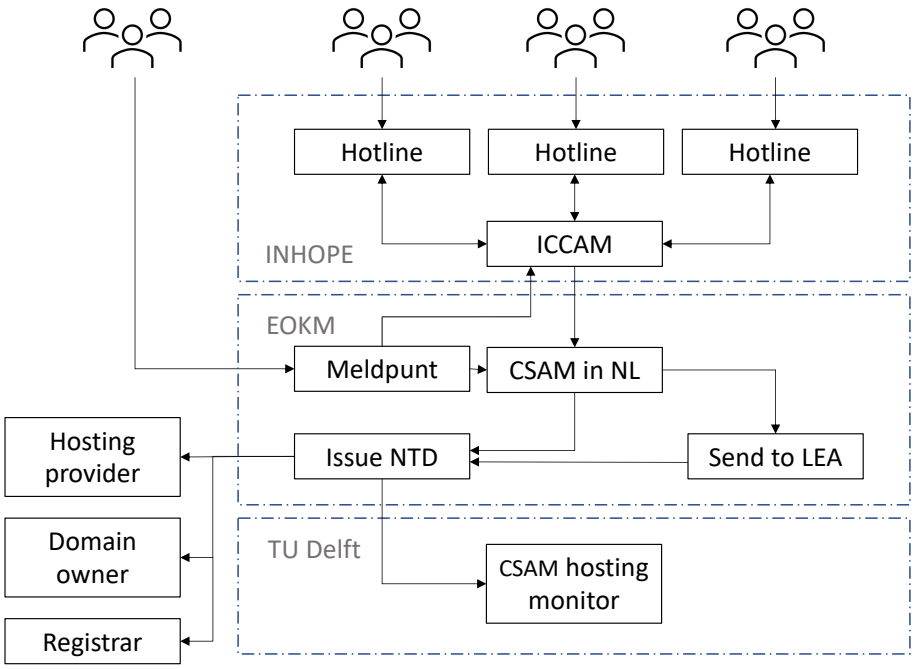


Figure 2.1: Overview of CSAM NTD process and TUD Monitor

connection or the link might otherwise be relevant to ongoing investigations.

From the data at EOKM, TU Delft then extracts the relevant properties to feed into the CSAM hosting monitor. Until January 2020, the processes at EOKM contained many manual actions, like manually crafting the NTD emails every day. In January 2020, an automated system called SCART (based on AbuseIO[1]) has been taken into operation. It automates certain steps after the manual verification of the suspected CSAM.

For the monitor, we had to develop different techniques and tools to process the data from EOKM, depending on whether the NTD process was conducted manually or via the new system, SCART. With support and feedback from EOKM and the SCART developers, we wrote software scripts to consistently enrich the reported CSAM URLs and count the number of URLs per hosting provider and per domain. This software runs on EOKM infrastructure, so that no CSAM URLs or other CSAM-exposing data needs to leave the network. Only the final aggregated outputs per hosting provider and per anonymized domain are stored at TU Delft.

## 2.2. CSAM Hosting Monitor

The data going into the scripts for the monitor contains the IP addresses where the CSAM was located, the domain, the time stamp, the approximate country location for that IP address (based on GeoIP), the Autonomous System (AS) in which the IP address resides, and the organization to which the corresponding IP address belongs, according to WHOIS data.

WHOIS data is derived from databases that contain information about the registered users or assignees of an Internet resource, such as domain names, IP addresses, and Autonomous System Numbers (ASN), which are used for routing traffic across networks. WHOIS data for domain names is maintained by the registry for the top-level domain under which the domain name is located. For .nl domains, for example, SIDN is the registry and its WHOIS system provides information about the domain owner. WHOIS for IP addresses and Autonomous System Numbers are operated by Regional Internet Reg-

[1] https://abuse.io/

istries (RIR). In the case of Europe, the RIR is RIPE NCC (Réseaux IP Européens Network Coordination Centre).

WHOIS data on IP addresses identifies what entity administers the network in which the IP address is located and specifically to which hosting provider the IP address has been assigned by the RIR. The hosting provider may further assign the IP addresses to one of its customers, which might itself be a company. In terms of responsibility, there are two scenarios. If the hosting provider agrees with the customer that the customer takes responsibility for handling potential abuse issues, then the provider updates the WHOIS data to reflect this. The WHOIS record will then show the customer name and contact information, including an email address where abuse can be reported. This is called a 'sub-allocation'. In that case, the hosting provider has to update the WHOIS data at the RIR to reflect the contact information for the customer. If no such sub-allocation is made in WHOIS, the provider retains the responsibility for dealing with abuse of resources at those IP addresses.

In Figure 2.2, we summarize our process to map CSAM URLs to hosting providers. We begin by selecting only the URLs that are verified to contain CSAM and for which NTDs were sent to the hosting providers and/or domain owners. (In a fraction of cases, no NTD was sent to the provider, because the CSAM had already been taken down before the NTD could be sent. Also, before the automation via the SCART systems, during peak time some URLs did not lead to an NTD because the limited availability of EOKM staff to manually send the NTD.)

Next, we perform IP WHOIS queries at the RIPE NCC database.[2] This provides us with the organization name of the entity to which the IP address has been assigned. It also provides us with the abuse contact address. We compare the abuse contact addresses from WHOIS with the contact address used for the NTD request. If these are not the same, then we do not include the URLs in our count of URLs for that provider, because we cannot verify that the provider actually received the NTD at the address that the provider has specified in WHOIS.

In a very small fraction of cases, we did not find an organization record in RIPE NCC WHOIS. We then mapped the missing organization using the MaxMind IP2ORG database[3] and manually validated this mapping by comparing the organization name to the domain in the abuse contract email address and the Autonomous System owner name. If these were consistent, we attributed the URL to that provider.

In addition to identifying the relevant hosting provider for the URLs, we also identify which domains contain CSAM. We extract the Fully Qualified Domain (FQDN) from the URL as a domain name. A URL is composed of sub-domain and the parent domain. For instance, a sub-domain `abc` and parent domain `example.com` has the fully qualified domain name `abc.example.com`. We used FQDNs to identify domains since, in some cases, sub-domains are hosted in different hosting providers. We assign a unique randomly-generated number to each of the FQDN to obfuscate the domain names.

## 2.3. Tracking CSAM Over Time

Ideally, the monitor would track CSAM hosting in the exact same way over time. This would then give us a continuous timeline of CSAM hosting in the Netherlands. This is now possible, using the SCART system at EOKM. However, from the period before SCART was fully operational, so before January 2020, we have only partial snapshots.

First, it took many procedural steps and safeguards to get access to the CSAM hosting data within the conditions specified in the law. This delayed access to the relevant data. Once access was obtained, some of the older data could not longer be reliably reconstructed for analysis. Second, because of the manual nature of the EOKM NTD process before January 2020, data was stored in various hand-made formats that were difficult to process automatically. For these reasons, we could only reconstruct the

[2] `https://apps.db.ripe.net/db-web-ui/query`

[3] `https://dev.maxmind.com/geoip/legacy/csv/#GeoIP_Organization_Edition_CSV_Database_Fields`

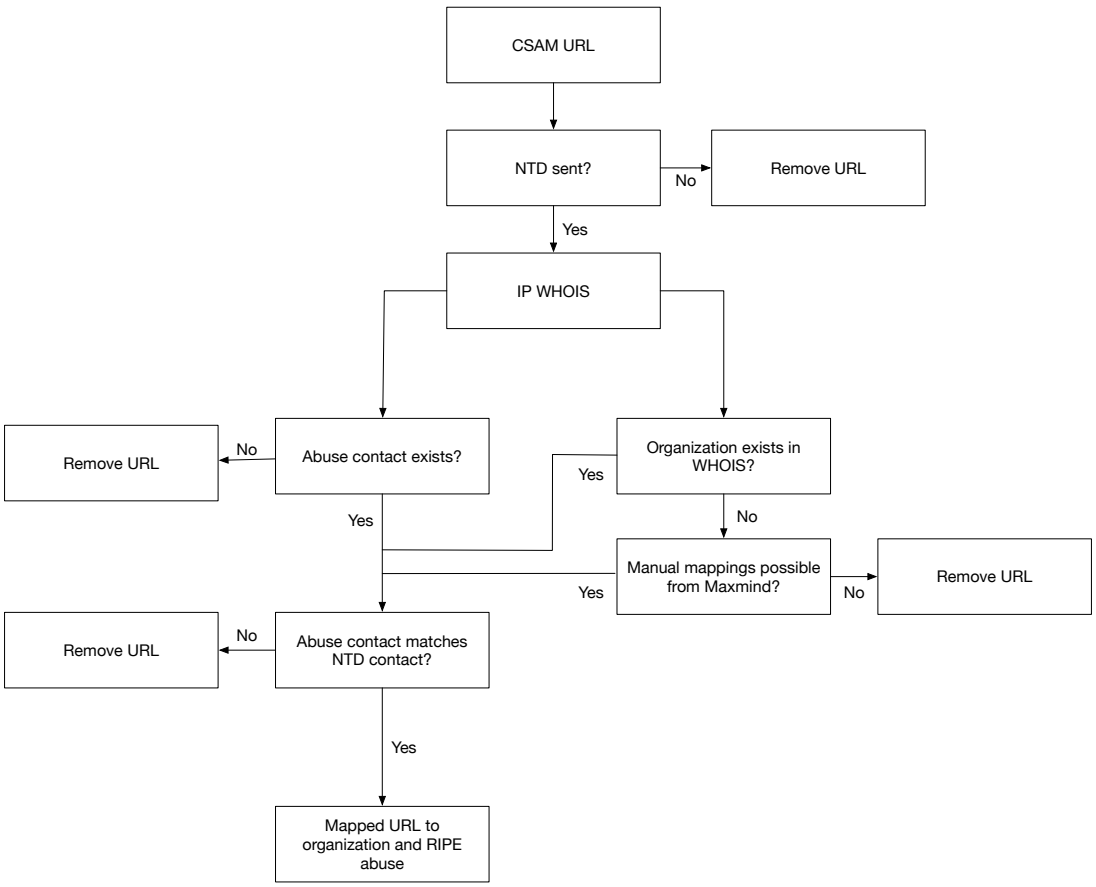


Figure 2.2: Mapping of CSAM URLs to hosting providers

CSAM hosting data for specific snapshots: one in October-November of 2018 and a longer snapshot for August-November of 2019. We included both snapshots in our previous report. In the current report, we drop the 2018 snapshot, to keep the tables and figure more consistent and because that snapshort only contained a two-month measurement window.

Another change that occurred is that there were difference in the data flowing into ICCAM. Since most of the EOKM CSAM data comes from ICCAM, changes in how INHOPE data sharing works have a serious effect on what is observed in the Netherlands. The data in ICCAM comes from other INHOPE hotlines. The hotlines in UK and Canada proactively search for CSAM with webcrawlers. When their crawlers visit more domains hosted in the Netherlands, then more CSAM would be found and thus more URLs would flow from ICCAM to EOKM and our monitor. The opposite change also occurs. As we mentioned before, in March of 2020, the Canadian hotline left INHOPE en thus also stopped sharing the CSAM it discovered in the Netherlands with EOKM. Instead, the Canadian hotline decided to send out its own NTDs.

In light of these complexities, we include the CSAM monitor data in two parts: a snapshot of 4 months in 2019 and a continuous measurement from 2020 onwards. Since these periods are based on different data generation processes, we do not put them into the same timeline. Instead, we make several tables to look at the main similarities and difference in CSAM hosting in 2019, 2020 and 2021. For 2021, we then present a more in-depth picture, including trends over time.

## 2.4. NTD Takedown Speed

Together with EOKM, we have also conducted a measurement of NTD responsiveness. The industry norm agreed upon in the public-private working group was that CSAM should be removed by providers within 24 hours after the NTD. We present the measurement and underlying methodology in more detail

in Chapter 3.

Because the uptime measurements are critical to evaluating the responsiveness of the hosting providers to CSAM NTDs, this process relies completely on manual checks by EOKM staff. We do not use automated measurements. While these are partially implemented in SCART, they are not yet reliable enough to use as evidence on provider compliance with the industry norm.

## 2.5. Corroborating NTD data with providers

Over the course of 2020 and 2021, the monitor had been working with data that was exported from SCART and that was assumed to contain all NTDs to providers. In the Fall of 2021, IP Volume, one of the providers that was included in the monitor, asked us to receive the NTD data that the monitor had used for its network.[4] That data was shared. IP Volume then ran a comparison of the shared data against its own internal logs of the NTDs it had received. The comparison found that most of the data was consistent, but it also unearthed three sources of discrepancies.

First, for some URLs, the provider logged different time stamps for receiving the NTD than the SCART export data logged for sending the NTD. We investigated this and there was indeed a discrepancy. However, the exact moment of the NTD is only important for the uptime measurement, not for the count of how many URLs were hosted by the provider. The discrepancy was corrected and we received time stamps that were consistent with those reported by the provider. We then recalculated the results for the uptime measurement we conducted in December 2021 – January 2022 with the new, corrected data.

Second, some NTDs were in our SCART export data, but not received by the provider. This turned out to be caused by the following scenario: the URL was first sent to the police. Normally, the URL would then be sent in an NTD the next day. Before this NTD was sent, EOKM staff did another check to see whether the material was still online. They found it had already been taken down, so there was no need anymore to send the NTD. These URLs did, however, appear on the provider network and thus should be included in the count of URLs for the provider, even though the provider never received an NTD for them.[5] This issue did not impact the uptime measurements. We double checked and found that for all URLs included in the uptime measurement, NTDs had been sent. We used the corrected NTD data and time to calculate the removal speed.

Third, the provider found NTDs that it had received, but that were not in our SCART export data. Further investigation made clear that these NTDs were incorrectly missing from the export data. This affected all providers. It means we undercounted the number of URLs for the providers. This omission has been corrected and we now receive a more complete NTD export data from SCART. This is the basis for the current report. We write “more complete”, because a small discrepancy still exists after the corrections. For a very small number of URLs, the SCART data does not have a record of an NTD directed at the abuse email of the provider, but the provider did state that they received an NTD. The volume of these exceptions is too small to impact the findings in this report.

We re-calculated the numbers for 2020-2022 from the corrected data and found that the overall findings from our 2020 CSAM hosting monitor report were unchanged. The top 5 remained the same in each year and the percentage of URLs hosted by each provider only changed marginally and did not impact the conclusions.

Because this check with IP Volume was very valuable at improving the quality of the monitor, we

[4] We should note that there are two providers called IP Volume: IP Volume Inc and IP Volume LTD. The former is registered in the Seychelles, while the latter is registered in the United Kingdom. When we refer to IP Volume in this report, we mean IP Volume Inc.

[5] Before this discrepancy was discovered, we assumed that for all URLs we included an NTD was sent. Indeed, we stated this assumption in our 2020 report. We now know that this was incorrect and that we should have stated that for a fraction of URLs no NTD was sent.

repeated this process with two other providers. These checks confirmed that the corrected export data was consistent with the provider findings. For NForce, we also corrected five URLs in the uptime measurement, since it was demonstrated that the domain left the NForce network during the measurement. Thus, the uptime that was observed was no longer associated with the NFore network.

In sum, regarding the discrepancies, we found that (1) there were inconsistencies in the time stamps of the NTDs for the uptime measurement, which we were able to correct; (2) providers sometimes do not receive an NTD, yet we are correctly counting them in the monitor, since the URL was hosted by them; and (3) there were NTDs missing in the monitor data, thus undercounting the total number of URLs hosted by the provider. This undercounting did not impact the findings in our prior report, nor in the current report.

The latter problem is now fixed going forward in time, but it did leave us with the problem for the historical data. As explained in Section 2.2, we independently verify the hosting location for all URLs, double checking the attribution done by SCART. For this purpose, we run our own scripts to enrich the data on the same day as the NTDs are sent. This enrichment process does an independent lookup as to which provider is hosting the URL. We cannot do this backward in time, since we need WHOIS and BGP data from the moment the NTD was sent. We were able to approximate this independent lookup for the 2021 data. For all the URLs that we missed, we checked whether we did a lookup for a different URL on the same domain in the same month as the previously-missed NTD was sent. For about 95% of all previously-missed NTDs, we found such a URL. This allowed us to reliably map these missed URLs to the respective provider. In this way, we were able to recover most of the missing NTD data for 2021. We could not conduct this process for the 2020 data, so in this report, we are still using the original numbers as included in the report from October 2020.

It was important to discover and fix these data quality issues and we are grateful for the efforts made by the providers, IP Volume in particular. While CSAM data faces grave restrictions in terms of sharing, we advise to conduct these cross checks with providers for subsequent analyses of the SCART data, e.g., by the new public authority that is currently being instituted.

## 2.6. Limitations

In this section, we discuss some of the measurement challenges that the monitor encounters. We take these into account when interpreting the results. First, the monitor data is based on the data received by EOKM. The core of the CSAM reports are based on ICCAM data. Two INHOPE contributors, Canada and U.K., proactively search for CSAM. This has several implications. First, it is unclear how their crawlers and manual searches detect the new content. There is a possibility the crawlers focus their search for content on hosting providers and domains where material had been detected before, rather than elsewhere. This might result in a biased picture of the problem, since material at those hosting providers is more likely to be discovered than elsewhere.

Second, there are fluctuations driven by increases and decreases in intake of data from the INHOPE network. Since March, the Canadian hotline is no longer a member of INHOPE. This means that IWF, the U.K. INHOPE member, is now a major source of data. It is not transparent how this data is collected. The number of URLs received from IWF went down significantly in the past three months. It is not clear what is causing these changes in CSAM volume to occur. We cannot tell whether this reflects changes in the presence of CSAM in Dutch networks or changes in the detection of CSAM. In all likelihood, it is a combination of these effects.

Another issue that we observed is that some domains respond with multiple IP addresses when they are queried via DNS. In other words, a domain name has identical copies of the same website residing on multiple servers, which in some cases maps to different hosting providers. The domain owners often use a well-known methodology, “round-robin DNS,” in which a different IP responds to every new query. It provides domain owners with the option of load balancing the across different providers and improves the fault tolerance of their website in case of downtime at one provider. This setup means the same

domain, and thus the same material, is mirrored across different providers. For the monitor, we have attributed the URL based on the DNS lookup conducted at the time of the URL being prepared for NTD.

EOKM maintains a 'green list' of cooperative domains. They are quick in removing the content, and most of the reported content for these domains is usually illegal. EOKM sends them an NTD right away, without performing additional checks. For the other domains, there is more investigation for each URL. In peak times, this might mean that it might take a while before EOKM has time to conduct this investigation. In that period, the URL might have gone offline. Once the investigation starts, the URL is no longer valid and thus no NTD is sent. This would mean that this domain owner and provider get slightly fewer NTDs than the 'green list' domains and providers get, even if they originally had the same amount of material. Given that we count on the basis of NTDs, it means that the 'green list' domains and providers get a slightly higher count, compared to those not on the green list.

A final and important challenge is to conduct automated uptime measurements of the URLs received by EOKM. While automated measurements would be ideal, as it would give us fine-grained data on how long a URL stays online after an NTD, a variety of technical challenges prevent this. One key challenge is the use of CAPTCHAs by the domain owners. CAPTCHAs are tests embedded on a web page where a visitor has to conduct a simple task, to determine whether or not the user is human. The current tracking scripts cannot automatically solve the CAPTCHAs and without this, the page content cannot be retrieved. For this reason, our report includes a manually conducted uptime measurement, conducted by EOKM in close collaboration with TU Delft. We describe this in more detail in Chapter 4. On one side, the manual process helped us by-pass CAPTCHAs. On the other side, the timing is less precise. The check after 24 hour actually takes place a bit later, say after 26 hours. It was never conducted earlier than 24 hours. So the error in the measurement is always to the benefit of the provider. In other words, we overestimate takedown speed. A related issue is the NTDs that were sent out on Friday. Since EOKM did not conduct checks during the weekend, these URLs were revisited on Monday. If they were removed at that time, then we counted them as removed within 24 hours. Again, we erred on the side of caution.

In summary, the monitor encounter certain challenges in measuring and monitoring the domains with CSAM data. While these issues have an effect on the specific amounts of CSAM that are calculated, these issues are not large enough to change the longitudinal picture and the concentrations in CSAM material that the monitor reports.

# 3

# CSAM Takedown

In the public-private roundtable to fight CSAM hosting, started in 2018, a norm was agreed upon: the hosting industry should remove CSAM within 24 hours after an NTD request was received. In our previous report, we included a manual check on the uptime of CSAM material based on a sample of URLs prepared by TU Delft. EOKM staff manually verified whether the CSAM at a URL was still online 24 and 48 hours after the NTD was sent, to measure how fast material was removed. In this chapter, we describe the results of the most recent manual checking of takedown speed, conducted in December 2021 – January 2021. We analyze how many providers and domain owners adhere to a 24h window for removal of CSAM.

## 3.1. Measuring Takedown Speed

How long does it take before a domain owner or hosting provider has taken down the CSAM material? Between between December 10, 2021 and January 20, 2022, we daily selected a sample of URLs to be checked manually by EOKM staff. These URLs were included in NTDs sent the day before (or the Friday before, when sampling on a Monday).

Before selecting a URL for the sample, we first checked whether it was still open in the system. In some cases, we found that EOKM staff had already done a manual check that same day and found that the material had been removed. We therefore did not include those URLs in our sample, but we did log that these URLs were removed within 24 hours.

On some days, the volume of URLs coming in through ICCAM was relatively small. In that case, all URLs could be checked. On other days, we had to sample the incoming new URLs so as to not overburden EOKM staff. We did stratified sampling per provider so that even providers with a small number of URLs would be included in the sample. A non-stratified random sample would have basically meant the sample would be consist mostly of the top 4 providers and we not have been able to get a reliable measurement for the other providers. After selecting the sample and sending it to EOKM staff, all URLs were visited after 14:00h, which was more than 24 hours after the NTD had been sent. If the material was still online, it was not compliant with the industry norm. Then a second visit would be scheduled for the next day, at least 48 hours after the NTD.

To measure takedown speed, we combine two datasets: (1) the URLs that were already taken down before we could select it for our sample (these URLS are all removed within 24 hours); and (2) the URLs from our sample, where EOKM conducted additional manual checks to see if the material remained online. We represent the takedown time in three categories: (i) content removed within 24h, so within the industry norm; (ii) content removed within 48h; and, finally, content that remained online for 48 hours and more. Some URLs in the last category were revisited several times more. The pattern was highly

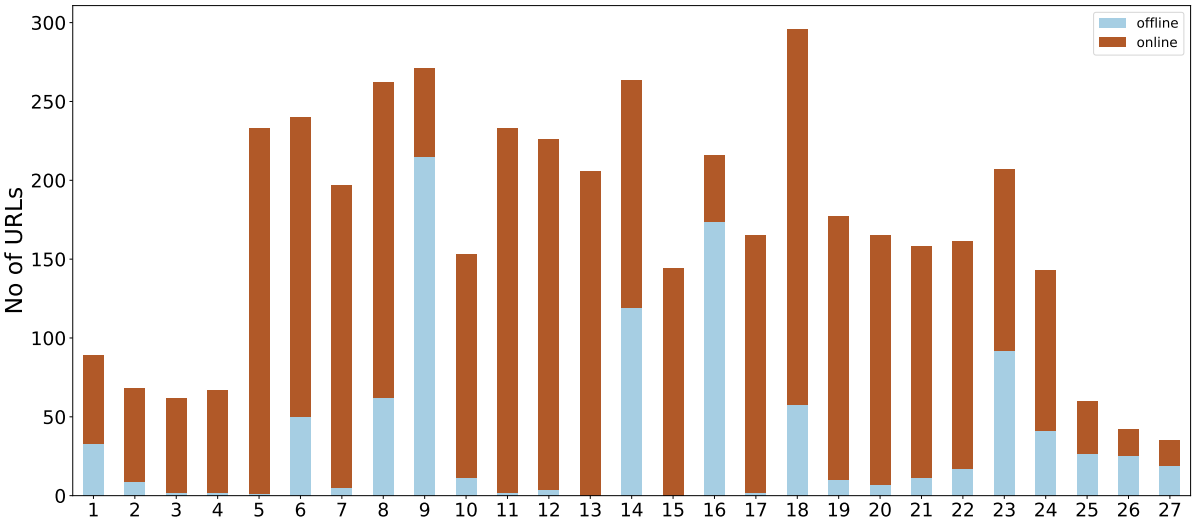


Figure 3.1: Number of URLs visited per day in (Dec 2021- Jan 2022) for checking whether the CSAM was still online

varied. Some went offline one day later, after 72 hours, some stayed online for the whole period where EOKM did checks, in an extreme case going up to 21 days.

Figure 3.1 shows the number of URLs that were selected by TU Delft and manually checked by EOKM every day during the measurement in December 2021 – January 2022. The number varies per day because, for certain days, there were more URLs that were verified as CSAM. We also monitored previously online URLs for at least one more check before dropping it from further checks. The color indicates how many URLs were seen on that day as online versus offline.

We cannot determine which actor actually took down the CSAM. The NTD is sent to the hosting provider and, if EOKM also has an abuse contact address for the domain, also to the domain owner. While we cannot observe from the outside who actually took the action to remove the CSAM material, the standard practice is that the provider forwards the NTD to its customer, the domain owner. If EOKM has an abuse contact address for the domain, then the domain owner itself will also receive the NTD directly, at the same time as the hosting provider. The domain owner is typically the entity that actually takes the CSAM offline. In some cases, the hosting provider might intervene itself, if it feels the domain owner is not acting responsibly. Under these industry practices, the takedown speed of a hosting provider is mostly dependent on the takedown speed of its clients, the domain owners, though it can also intervene directly or put pressure on the clients. To take this joint responsibility into consideration, we will also show how many domains were associated with each provider for the URLs in the uptime measurement, in the next section.

## 3.2. Takedown Speed per Hosting Provider

In Table 3.1, we present the results of the manual uptime measurement, combining both datasets (as explained above). In total, we have data for 6268 URLs located at 33 hosting providers and 98 domains. It is important to note that because of the sampling approach discussed in the previous section, the number of URLs in the sample should *not* be read as indicating the total volume of URLs, nor off the relative amount of URLs for each provider.

In terms of the results, we found that 87% of all CSAM is removed in 24 hours. A further 3% is removed between 24-48 hours, and 10% remains online for 48 hours or longer – in some extreme case it was still online when we stopped the measurements.

As in our previous report, we find a high variance in performance. NForce, the provider with largest portion of CSAM in 2021, manages to get 100% of the URLs removed within 24 hours. Given the large volume of URLs, this a truly impressive result. It is even higher than the already high performance

measured in 2020. WIBO Baltic, Digital Ocean, Leaseweb NL and several other providers also manage to get to a perfect record of 100%, which demonstrates the effectiveness of their efforts to remove CSAM. All in all, 19 of 33 providers in the uptime measurement (58%) manage to get more than 90% of all URLs in our measurement removed within 24 hours. KnowSRV, the number two with the most CSAM, has been removing over 84% of the CSAM material within 24 hours. About 15% was still online after 48 hours.

Then there are number of providers who have a weaker performance. Take Hostslim. It removes only about 38% within 24 hours, then an additional 58% between 24-48 hours and leaves 4% online for even longer. Intersect-Host LTD removed 67% within 24 hours, but left 33% online beyond 48 hours. Host Sailor Ltd operates in the same range. About 50% of the CSAM URLs is still online after 48 hours.

And then there are the providers who seem to take very little action, if any at all. Telegram Messenger Inc, not to be confused with the smartphone communications service, is a hosting provider which removed only 6% in time. Over 94% of the URLs were still online after 48 hours. Amaratu Technology, SpectraIP and SKB Enterprise also fail to meet the norm, leaving over 90% of the CSAM online longer than 48 hours. In total, 10 out of 33 providers manage to remove less than 50% within 24 hours. Seven providers have more than half of their CSAM online for longer than 48 hours. If we focus on the providers that have dominated the top 5 of CSAM hosting in the past two years (see Table 4.2), we can see that NForce has been very responsive and actively working to combat the problem. It has been consistently good at removing the material within 24 hours. KnownSRV does remove the bulk of the material on time, but is less effective than NForce. A new entrant into the top 5, Telegram Messenger Inc, did extremely poorly. Another new entrant, Hostslim B.V., did better, but took 48 hours before most material was removed. (MSK.HOST Hosting was not in the uptime measurement, because it did not host any URLs during the measurement period.) An earlier top 5 provider, Leaseweb NL, now has only a modest volume of URLs left and manages to remove 100% within the norm. No URLs were reported for IP Volume during the uptime measurement, which in itself is a good thing. It also meant we could not observe their removal speed.

Within our uptime measurement, the number of domains with CSAM per provider varies substantially. In some providers all URLs are located on just a single domain. At the other extreme, we find KnownSRV, which is hosting in total 21 domains with CSAM. More domains means the provider is dependent on more clients (domain owners) to get the material removed, unless the provider is willing to take unilateral action.

## 3.3. Takedown Speed per Domain

Since hosting providers are typically dependent on domain owners for the removal of the CSAM, we also analyze the data at the level of domains. For this purpose, we only use the data from our manual sample. In our sample, we have 1493 URLs across 69 domain owners. In Table 3.2, we present the results for the 20 domains with the most URLS. All other domains are combined in the last row.

We see a pattern where three distinct groups are visible. Some domain owners are highly responsive, such as the owner of domain 1681, which removed a large set of URLs all within 24 hours. Then there are domains which are responsive, but slow. Domain 1581, for example, removed everything within 48 hours, but only 39% within 24 hours. Finally, there is a group of eight unresponsive domain owners, who left most material, often more than 90%, online beyond 48 hours. For domain 321, this is even 99%.

In short: our results show that 8 of the top 20 domains remove 90% or more of the material within 24 hours. A further 4 domains remove more than 90% in 48 hours. And 8 domains are very unresponsive. Most of their material remains online beyond 48 hours. A single unresponsive domain owner with a substantial amount of CSAM can significantly impact the performance of a hosting provider. The provider then has to pressure the domain owner into responsiveness or drop them as a client. Some providers took the latter action.

| Hosting Provider | Total Urls | Domains | Offline within 24h | Offline between 24-48h | Online for 48h or more |
|---|---|---|---|---|---|
| NFOrce Entertainment | 4091 | 8 | 4091 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| KnownSRV Ltd. | 1215 | 21 | 1013 (83.37%) | 24 (1.98%) | 178 (14.65%) |
| Telegram Messenger Inc | 335 | 1 | 19 (5.67%) | 1 (0.3%) | 315 (94.03%) |
| HostSlim B.V. | 229 | 2 | 88 (38.43%) | 132 (57.64%) | 9 (3.93%) |
| WIBO Baltic UAB | 70 | 4 | 70 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Digital Ocean | 60 | 1 | 60 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Leaseweb NL | 48 | 5 | 48 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Amarutu Technology Ltd | 44 | 1 | 4 (9.09%) | 0 (0.0%) | 40 (90.91%) |
| Intersect-Host LTD | 36 | 5 | 24 (66.67%) | 0 (0.0%) | 12 (33.33%) |
| Host Sailor Ltd | 22 | 5 | 3 (13.64%) | 8 (36.36%) | 11 (50.0%) |
| SKB Enterprise B.V. | 17 | 6 | 0 (0.0%) | 1 (5.88%) | 16 (94.12%) |
| SpectraIP B.V. | 17 | 3 | 0 (0.0%) | 1 (5.88%) | 16 (94.12%) |
| WorldStream B.V. | 10 | 1 | 9 (90.0%) | 0 (0.0%) | 1 (10.0%) |
| IROKO Networks Corp. | 9 | 4 | 9 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| ABC Consultancy | 9 | 8 | 9 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| ONLINE SAS | 9 | 2 | 1 (11.11%) | 0 (0.0%) | 8 (88.89%) |
| ITL LLC | 8 | 1 | 0 (0.0%) | 8 (100.0%) | 0 (0.0%) |
| The Infrastructure Group | 7 | 1 | 7 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| SkyLink Data Center BV | 7 | 1 | 7 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Zomro B.V. | 5 | 3 | 4 (80.0%) | 0 (0.0%) | 1 (20.0%) |
| Hostiserver | 3 | 3 | 1 (33.33%) | 1 (33.33%) | 1 (33.33%) |
| HostUS | 3 | 1 | 2 (66.67%) | 0 (0.0%) | 1 (33.33%) |
| DataWeb Global Group | 2 | 2 | 2 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| BlazingFast | 2 | 1 | 2 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Global Layer B.V. | 2 | 1 | 2 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Hostkey B.v. | 2 | 1 | 2 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| King Servers B.V. | 1 | 1 | 0 (0.0%) | 0 (0.0%) | 1 (100.0%) |
| ServerStack, Inc. | 1 | 1 | 1 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| Hivelocity Inc | 1 | 1 | 1 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| FiberXpress BV | 1 | 1 | 1 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| UA-HOSTING SIA | 1 | 1 | 1 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| UAB Cherry Servers | 1 | 1 | 1 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| WEB_GroupInternet INC | 1 | 1 | 1 (100.0%) | 0 (0.0%) | 0 (0.0%) |

Table 3.1: Uptime measurement for hosting providers (Dec 2021 – Jan 2022)

## 3.4. In Sum

In summary, we observe that there is a lot of variance in how fast providers and domain owners get CSAM material removed. Some manage to get more than 90% removed within the 24-hour norm. Others need 48 hours to get most material removed. Overall, about 10% of CSAM remain online after 48 hours and this is located on specific domains that are unresponsive and where the hosting provider is also not intervening.

| Domain Identifier | Total Urls | Offline within 24h | Offline between 24-48h | Online for 48h or more |
|---|---|---|---|---|
| 321 | 318 | 2 (0.63%) | 1 (0.31%) | 315 (99.06%) |
| 1681 | 226 | 226 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| 1581 | 212 | 82 (38.68%) | 130 (61.32%) | 0 (0.0%) |
| 1620 | 189 | 16 (8.47%) | 0 (0.0%) | 173 (91.53%) |
| 963 | 65 | 65 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| 224 | 55 | 46 (83.64%) | 9 (16.36%) | 0 (0.0%) |
| 1395 | 44 | 4 (9.09%) | 0 (0.0%) | 40 (90.91%) |
| 1696 | 44 | 44 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| 307 | 27 | 27 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| 1651 | 25 | 13 (52.0%) | 12 (48.0%) | 0 (0.0%) |
| 1632 | 21 | 21 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| 1255 | 19 | 19 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| 1285 | 15 | 14 (93.33%) | 1 (6.67%) | 0 (0.0%) |
| 1699 | 15 | 0 (0.0%) | 1 (6.67%) | 14 (93.33%) |
| 1410 | 15 | 4 (26.67%) | 2 (13.33%) | 9 (60.0%) |
| 1387 | 12 | 10 (83.33%) | 2 (16.67%) | 0 (0.0%) |
| 937 | 12 | 0 (0.0%) | 0 (0.0%) | 12 (100.0%) |
| 1580 | 11 | 0 (0.0%) | 0 (0.0%) | 11 (100.0%) |
| 1193 | 10 | 1 (10.0%) | 0 (0.0%) | 9 (90.0%) |
| 1776 | 10 | 9 (90.0%) | 0 (0.0%) | 1 (10.0%) |
| Others | 149 | 100 (67.11%) | 18 (12.08%) | 30 (20.13%) |

Table 3.2: Uptime measurement for domains (Dec 2021 – Jan 2022)

# 4

# CSAM Landscape in the Netherlands

In this chapter, we explore which providers in the Netherlands host CSAM. The main goal of this chapter is to understand the current state of CSAM and how it has evolved.

As explained in Chapter 2, we have two datasets that cover different time periods, from different stages in the development of the monitor. From August till December 2019, we have extracted data from the manual NTD process logs. From January 2020 onwards, we have longitudinal data exported from the SCART system. The automation offered by SCART has greatly increased the number of URLs that EOKM could send NTDs for. It also coincided with a larger influx of URLs from ICCAM. See Chapter 2 for more details. We did uncover discrepancies in the exported data in early 2022, which we tried to correct for backwards in time for 2021 (see Section 2.5. The aim of this chapter is the following questions:

- How has the volume of CSAM changed over time?
- How is the CSAM content distributed over hosting providers? How has this changed?
- How is the CSAM content distributed over domains? How has this changed?

We first discuss patterns in the volume of CSAM in the Netherlands and how this material is distributed over the hosting landscape. Next, we look at the domains where this material has been hosted.

## 4.1. CSAM Volume

In Figure 4.1, we show the number of URLs reported to hosting providers per month. It can be seen that the number of URLs remained around 5,000 to 7,000 URLs per month at the end of 2019. In December 2019, the automated system, SCART, was tested and EOKM personnel were trained in using it. That transition lowered the number of URLs that were sent to providers for that month. From January 2020 onward, the new SCART system was fully operational. Figure 4.1 shows that the number of URLs for which EOKM has sent NTDs increased significantly since then.

The amount of CSAM detected and notified about in the Netherlands has fluctuated a lot over the past two years. In 2021, we saw many months with more than 20,000 URLs. This is similar to 2020. There is no consistent trends one way or the other. Over the last six months of 2021, the volume has gone down, but this effect could be temporary.

The fluctuations are the result of several interacting factors that cannot be disentangled. There might be more material available online at certain times, but there is also better detection going on, so more material is discovered. There are also fluctuations in the intake of data from the INHOPE network. Since

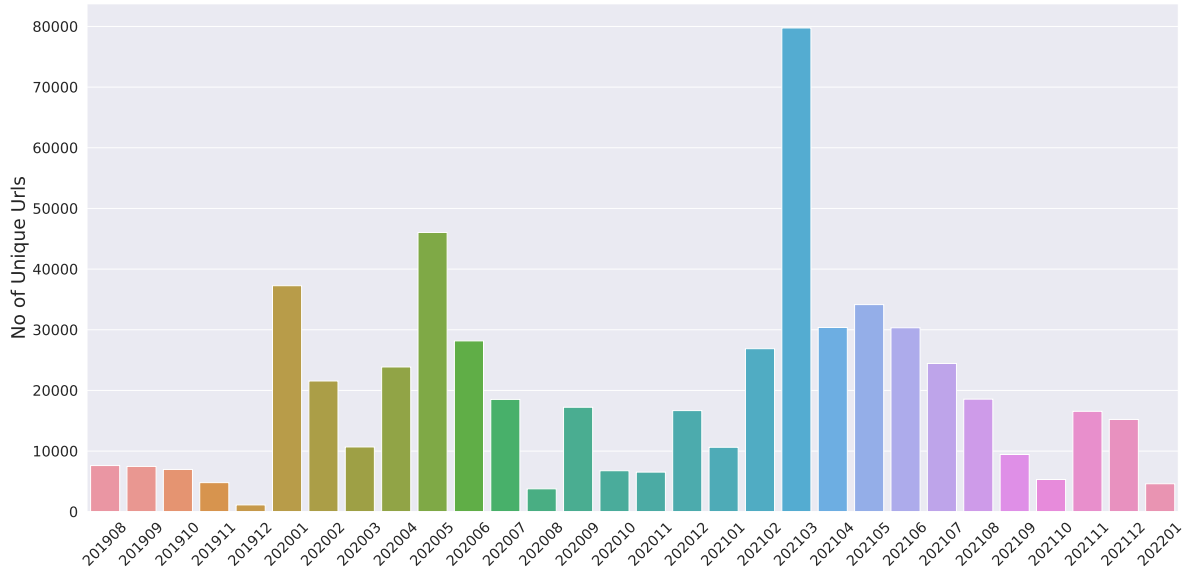


Figure 4.1: Number of URLs notified per month after SCART is deployed

| Aug(2019) - Dec(2019) | | Jan(2020) - Dec(2020) | | Jan(2021) - Dec(2021) | |
|---|---|---|---|---|---|
| **Provider** | **No of URLs** | **Provider** | **No of URLs** | **Provider** | **No of URLs** |
| NFOrce | 20,447 (73.31%) | NFOrce | 216,183 (92.76%) | NFOrce | 269,423 (91.33%) |
| IP Volume | 5135 (18.41%) | KnownSRV | 5938 (2.54%) | KnownSRV | 8570 (2.90%) |
| Leaseweb NL | 1006 (3.60%) | IP Volume | 5558 (2.38%) | HostSlim | 2602 (0.88%) |
| KnownSRV | 574 (2.05%) | Leaseweb NL | 1907 (0.81%) | Telegram Messenger | 2552 (0.86%) |
| Serverius Holding | 269 (0.96%) | HOSTKEY | 1255 (0.53%) | MSK.HOST Hosting | 1629 (0.55%) |
| Others | 460 (1.64%) | Others | 2200 (0.94%) | Others | 10,199 (3.45%) |

Table 4.1: Number of CSAM URLs hosted in Hosting Providers

Canadian hotline is no longer a member of INHOPE, IWF, the U.K. INHOPE member, is now a major source of data for ICCAM. The detection process at IWF is not public and we cannot evaluate how it impacts the discovery of CSAM in the Netherlands.

## 4.2. Distribution of CSAM Across Hosting Providers

In Table 4.1, we show distribution of CSAM across hosting providers in the period January 2020 – January 2022. Notwithstanding the fluctuations in the data, the distribution of URLs across providers reflects some stable patterns. The material is concentrated in a small number of providers. Four providers have the overwhelming majority of all URLs: 97.37% in 2019, 98.49% in 2020 and 95,97% in 2021. Moreover, over the past two year, most of the content was hosted by domains located with one hosting provider: NFOrce – though we should add that this pattern has been changing dramatically since the end of 2021, as we will discuss later.

There are some changes in which providers form the top 5 (see Table 4.2. While NForce and KnownSRV are in there consistently, other companies shift positions over time. In 2021, three providers move into the top 5: HostSlim, Telegram Messenger, MSK.HOST Hosting. These providers were not completely new. They had been hosting small quantities of CSAM in earlier periods. Though they have a slightly larger share of material than before, it still represents only a small fraction of the overall volume. Leaseweb NL and IP Volume dropped out of the top 5. Beyond the top 5 providers, there is a long and fluctuating list of providers that have a small number of reported URLs. In total, 28 hosting providers received at least one CSAM NTD in 2019, 52 providers received an NTD in 2020 and 87 did so in 2021. Moreover, in January of 2022, 24 hosting providers received an NTD.

Some of the domains seems to move across providers. We are not sure whether these domains

| Hosting Provider | 2021 | 2020 | 2019 |
|---|---|---|---|
| NFOrce | 1 | 1 | 1 |
| KnownSRV | 2 | 2 | 4 |
| HostSlim B.V. | 3 | | |
| Telegram Messenger Inc | 4 | | |
| MSK.HOST Hosting | 5 | | |
| IP Volume | | 3 | 2 |
| Leaseweb NL | | 4 | 3 |
| HOSTKEY B.V. | | 5 | |
| Serverius Holding | | | 5 |

Table 4.2: Top 5 hosting providers across three time periods

| Hosting Provider | URLs |
|---|---|
| Telegram Messenger | 5139 (50.00%) |
| KnownSRV | 1167 (22.70%) |
| NFOrce | 941 (18.31%) |
| Amarutu Technology | 181 (3.52%) |
| Intersect-Host | 75 (1.45%) |
| Others | 205 (3.98%) |

Table 4.3: Top 5 hosting providers January (2022)

moved here or whether we are seeing some form of multi-casting, where several providers all host the same domain. Some of the domains appeared to return to KnownSrv after a period where the domain name was being resolved to a different provider. The process of locating CSAM content at EOKM and TU Delft does not have a reliable way of handling multi-casting – that is, to observe whether the same domain is hosted by multiple providers at the same time. This would also mean that after removal of the domain or specific material by one provider, the domain or material might still be served from another provider.

Figure 4.2 shows that the picture over the past two years has been mostly stable month to month, because NForce had the bulk of all material. In many months, their share consisted of over 90% of the URLs. The remaining fraction was being hosted by some recurring providers, like KnownSRV, and a long list of fluctuating providers.

This pattern changed dramatically during the last quarter of 2021. The portion of CSAM hosted by NForce has been diminishing consistently in dramatically decreased. As Table 4.3 shows, they ended up with less than 20% in January 2022. The drop appears to continue further in the course of that month. Figure 4.3 shows the number of URLs hosted at Force per week, from August 2021 to January 2022. In the course of January, the number of URLS detected at NForce goes to zero. As we will see in the next section, this trend break is directly associated with a number of domains with CSAM leaving NForce. The role of domains is further explored in the next section.

## 4.3. Distribution of CSAM Across Domains

To better understand the concentration of CSAM at hosting providers, we are taking a closer look at the domains that host most of the domains. We know from the annual report of INHOPE and EOKM that websites, including image-hosting websites, make up the bulk of the discovered CSAM material.[1] This pattern holds true globally, as well as in the Netherlands.

[1]See p. 26 of the annual report of INHOPE for 2020 (`https://inhope.org/media/pages/the-facts/download-our-whitepapers/annual-report/bb4dd3cdc3-1628156678/inhope-annual-report-2020.pdf`)

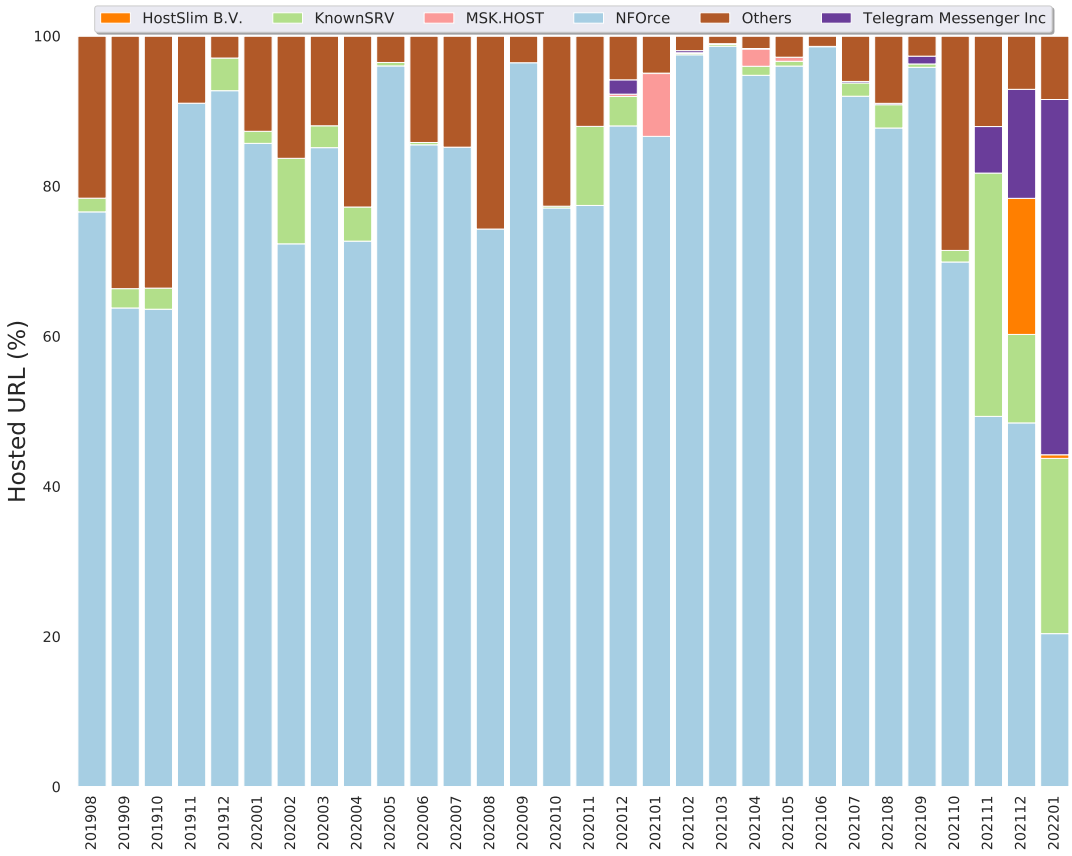


Figure 4.2: Number of URLs notified per hosting provider for each month, for the top 5 of 2021

How are the CSAM URLs distributed across different domains? Do we see a pattern of concentration there as well? The short answer is: yes. We identified which 15 domains contained the most CSAM URLs in the period 2019-2021. Figure 4.5 plots what portion of all URLs were located on these 15 domains: 57.13% of all URLs in 2020 and 83.87% in 2021. While this fluctuates from one month to the next, in most months in 2021, the top 15 domains contain over 70% of the URLs

The overall set of domains is much larger than 15, of course. The variance across these domains is enormous. Some of these domains contain thousands of URLs with CSAM, while others contain only a handful. The networks of the top 5 hosting providers have a disproportionately large portion of all these domains. In Figure 4.4, we plotted the portion of domains located with the top 5 providers from 2021, going back to 2020 and 2019. The portion hosted by the top 5 roughly fluctuates around 40%. While that signals some concentration, it also shows that the number of domains outside the top 5 is larger. As said, most of these domains contain only small amounts of CSAM.

Even though the CSAM is concentrated in a small number of domains, the pattern here is much more dynamic than with the providers that host the CSAM, as can be seen in Figure 4.5. First of all, the proportion of CSAM per domain can change quite a lot, even in a short time frame like a month. In one month, a domain dominates the total volume of CSAM, while in the next month it is negligible or disappears altogether. In Section 4.1, we saw a spike in the volume of reported CSAM in March. When we look at the distribution of URLs over the domains in Figure 4.1, we can see that this spike was caused by URLs detected on just one domain (domain 2).

Second, the set of domains where the CSAM is hosted is dynamic over time. It changes quite a lot. In Table 4.4, we take the top 10 domains in 2021 (left columns) and then look at their position for 2020 and 2019 (middle and right columns). As can be seen, most of the top 10 of 2021 was not in the top 10 in 2020. And in 2019, none of these domains were anywhere near the top 10, except domain 2. Most had a tiny fraction of URLs and six of these domains had no CSAM URLs at all in that year.

In Table 4.5, we make the same comparison for the top 10 from 2020. Only half of them were already

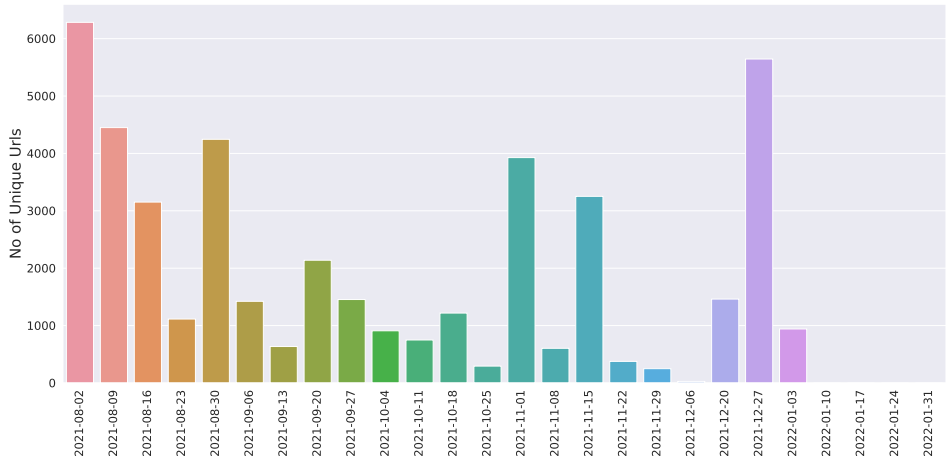


Figure 4.3: Distribution of NFOrce URLs per week

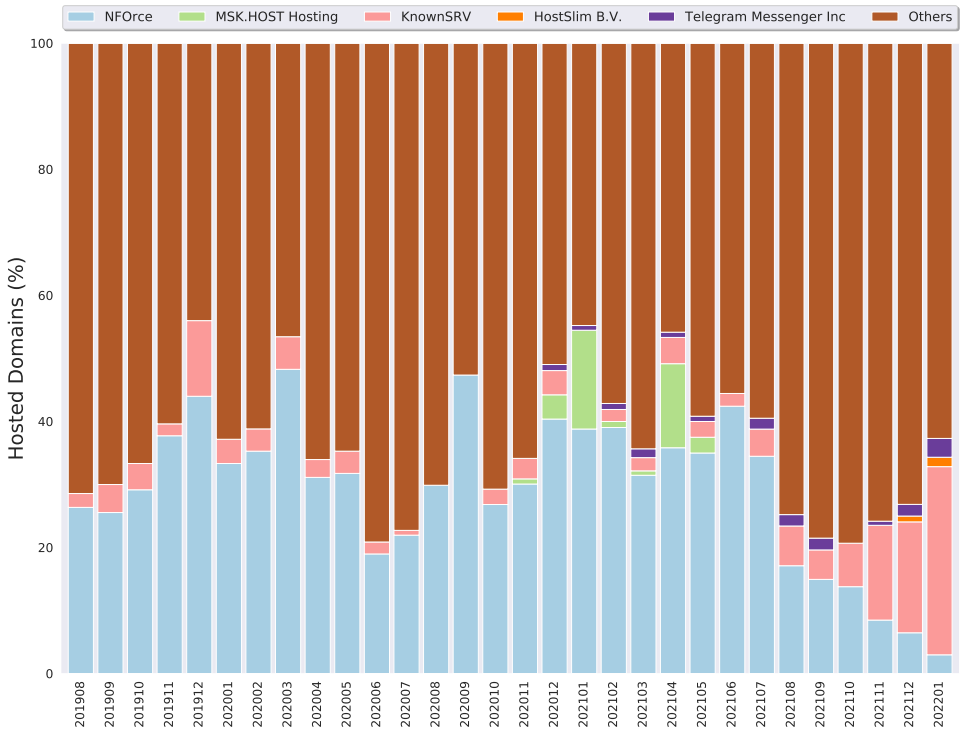


Figure 4.4: Proportion of all domains with CSAM per provider (Top 5 Hosting provider between Aug 2019- Jan 2022)

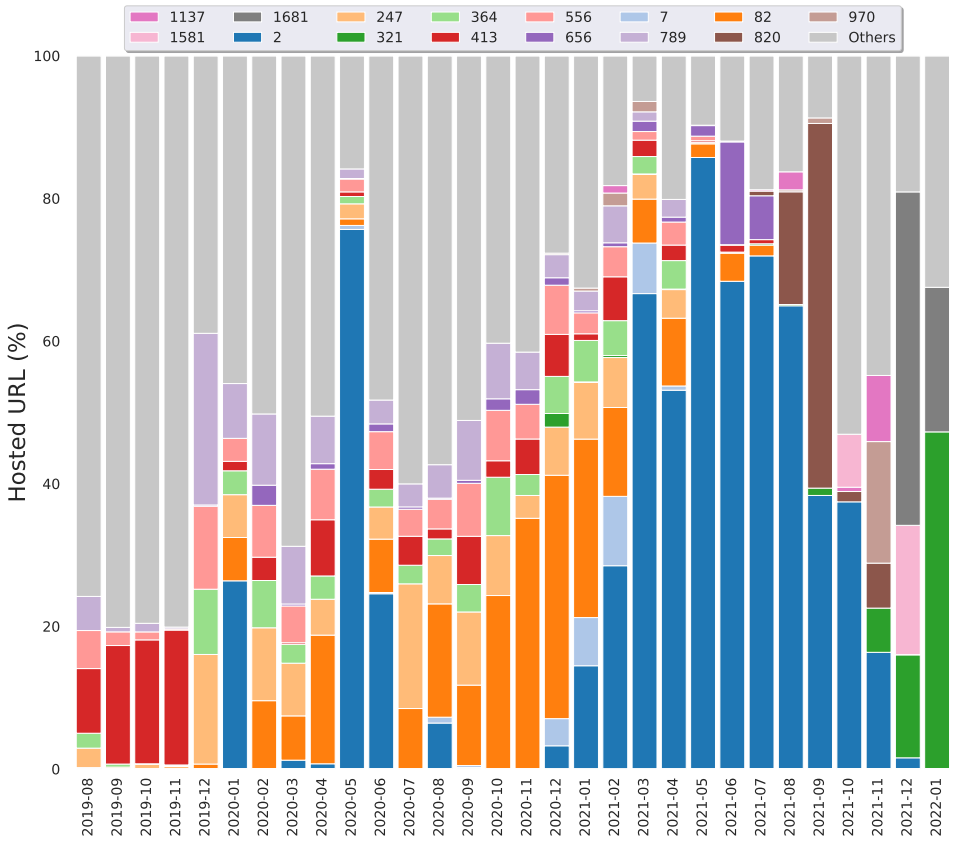


Figure 4.5: Percentage of URLs hosted per domain (Top 15 domains from Oct 2019-Jan 2022)

in the top 10 in 2019 and only four remain in the top 10 in 2021. For completeness sake, we also present the comparison for the top 10 from 2019 (Table 4.6). Five domains were still present in 2020 and only one remains in the top 10 in 2020. Another way to quantify this change is to say in 2021, the top 10 domains hosted 79.41% of all detected CSAM URLs (see the bottom row of Table 4.4). This same set of domains hosted only 17.02% in 2019.

In sum, we are left with a somewhat puzzling picture. The domains that contain the URLs has some stability, but it does show large fluctuations over time. Yet the hosting providers with most CSAM form a more stable set. Even though the domains change from one year to the next, they somehow still predominantly end up in the network of the same hosting providers.

We wondered whether this was explained by the size of these providers: the larger a provider is, the higher the probability that it attracts domains with CSAM. In other words, domains would end up with these providers because they are the dominant players in the hosting market. In Figure 4.6, we mapped four providers in terms of size, as measured by two indicators: the number of domains they host (Y-axis) and the number of IP addresses they announce (X-axis). (The data for this plot comes from 2019, but the pattern has not changed since.) So higher and more to the right means larger. We can see that these four providers are not the biggest providers. Only one provider, Leaseweb NL, is at the high end in terms of size, and this provider actually has a much smaller share of CSAM domains and URLs than NFOrce. And another provider, KnownSRV, is actually relatively small, compared to the median size in the market, and yet it has had a significant portion of all CSAM in the past two years. In other words, size is not the explanation for the pattern of concentration of CSAM that we are seeing.

Another explanation could be that various image-hosting domains belong, in fact, to the same owner. In some cases, we can this from the WHOIS records and abuse contact information for the domains: different domains have the same contact information. If one owner already has its sites set up with a specific provider, then it would make sense that new sites of the same owner would also get set up there. In short, if the new domains are owned by the same firms that operate the older domains, then it would make sense that the changes in the list of domains with CSAM still end up with the same top 4 providers.

| Domain | 2021 | | | 2020 | | | 2019 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | Total Urls | % | Rank | Total Urls | % | Rank | Total Urls | % |
| 2 | 1 | 165,300 | 56.26 | 1 | 49613 | 21.28 | 59 | 16 | 0.06 |
| 82 | 2 | 15642 | 5.32 | 2 | 25559 | 10.96 | | | |
| 7 | 3 | 9021 | 3.07 | 25 | 1022 | 0.44 | | | |
| 820 | 4 | 8168 | 2.78 | | | | | | |
| 656 | 5 | 7928 | 2.70 | 20 | 1721 | 0.74 | | | |
| 247 | 6 | 6796 | 2.31 | 5 | 15681 | 6.73 | 16 | 298 | 1.11 |
| 1681 | 7 | 6003 | 2.04 | | | | | | |
| 364 | 8 | 5095 | 1.73 | 9 | 7734 | 3.32 | 23 | 201 | 0.75 |
| 413 | 9 | 4765 | 1.62 | 11 | 7609 | 3.26 | 2 | 4049 | 15.11 |
| 970 | 10 | 4633 | 1.58 | 79 | 35 | 0.02 | | | |
| **Total** | | **233,351** | **79.41** | | **108,974** | **46.75** | | **4564** | **17.02** |

Table 4.4: Comparison of top 10 domains relative to 2021

| Domain | 2021 | | | 2020 | | | 2019 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | Total Urls | % | Rank | Total Urls | % | Rank | Total Urls | % |
| 2 | 1 | 165,300 | 56.26 | 1 | 49613 | 21.28 | 59 | 16 | 0.06 |
| 82 | 2 | 15642 | 5.32 | 2 | 25559 | 10.96 | | | |
| 601 | 25 | 9021 | 0.46 | 3 | 24940 | 0.44 | 4 | 1680 | 6.27 |
| 357 | 41 | 8168 | 0.17 | 4 | 16617 | | 1 | 4635 | 17.30 |
| 247 | 6 | 7928 | 2.31 | 5 | 15681 | 0.74 | 16 | 298 | 1.11 |
| 789 | 12 | 6796 | 1.20 | 6 | 12592 | 6.73 | 12 | 507 | 1.89 |
| 556 | 11 | 6003 | 1.22 | 7 | 11341 | | 10 | 635 | 2.37 |
| 671 | 21 | 5095 | 0.72 | 8 | 11152 | 3.32 | 9 | 766 | 2.86 |
| 364 | 8 | 4765 | 1.73 | 9 | 7734 | 3.26 | 23 | 201 | 0.75 |
| 242 | 29 | 4633 | 0.24 | 10 | 7675 | 0.02 | 6 | 1337 | 4.99 |
| **Total** | | **204,628** | **69.63** | | **182,904** | **78.45** | | **10,075** | **37.6** |

Table 4.5: Comparison of top 10 domains relative to 2020

We cannot independently verify that this explanation is actually correct.

| Domain | 2021 | | | 2020 | | | 2019 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | Total Urls | % | Rank | Total Urls | % | Rank | Total Urls | % |
| 357 | 41 | 485 | 56.26 | 4 | 16617 | 7.13 | 1 | 4635 | 17.30 |
| 413 | 9 | 4765 | 5.32 | 11 | 7609 | 3.26 | 2 | 4049 | 15.11 |
| 382 | 38 | 505 | 0.46 | 22 | 1519 | 0.65 | 3 | 3614 | 13.49 |
| 601 | 25 | 1360 | 0.17 | 3 | 24940 | 10.70 | 4 | 1680 | 6.27 |
| 757 | 36 | 527 | 2.31 | 14 | 4104 | 1.76 | 5 | 1665 | 6.21 |
| 242 | 29 | 714 | 1.20 | 10 | 7675 | 3.29 | 6 | 1337 | 4.99 |
| 141 | 34 | 543 | 1.22 | 12 | 4477 | 1.92 | 7 | 1123 | 4.19 |
| 283 | 42 | 441 | 0.72 | 13 | 4255 | 1.82 | 8 | 1035 | 3.86 |
| 671 | 21 | 2124 | 1.73 | 8 | 11152 | 4.78 | 9 | 766 | 2.86 |
| 556 | 11 | 3579 | 0.24 | 7 | 11341 | 4.86 | 10 | 635 | 2.37 |
| **Total** | | **15043** | **5.1** | | **93689** | **40.16** | | **20,539** | **76.65** |

Table 4.6: Comparison of top 10 domains relative to 2019

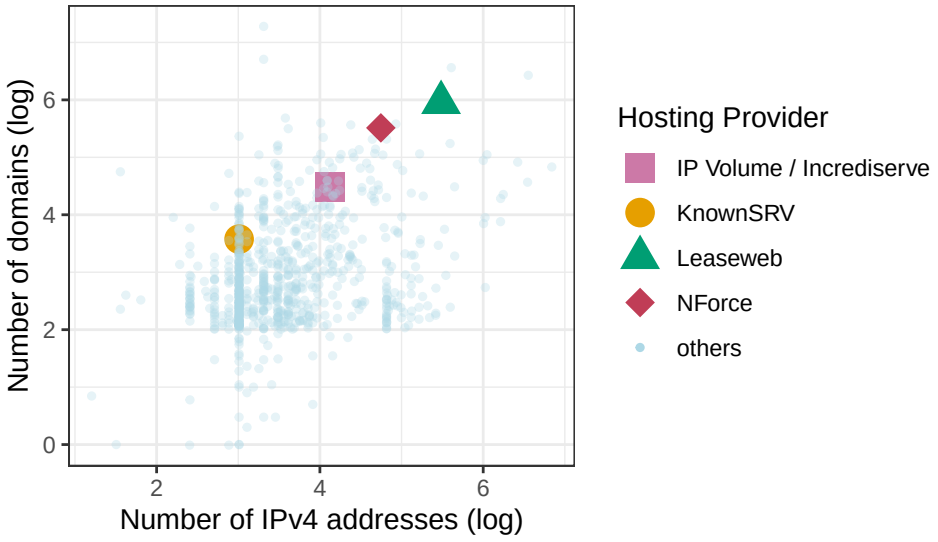


Figure 4.6: Dutch hosting providers mapped by size in terms of the number of IP addresses and the number of domains they host (based on Farsight DNSDB data from 2019)

# 5

# Conclusion

The CSAM Hosting Monitor is one instrument in a broader set of public-private actions to combat the hosting of CSAM in the Netherlands. It helps the stakeholders by making transparent which hosting providers have most of the material in their networks, how fast they are in removing the material once they receive an NTD, and how the location of the material changes over time.

In principle, the monitor would also track the progress of the overall approach to combat CSAM hosting. Simply put: is the amount of CSAM hosted in the Netherlands going down? And if so, which providers and domains are improving the most and which ones are not improving? These are critical questions to guide the future actions of the public-private partnership to combat CSAM hosting. In reality, however, the volume of Dutch CSAM is highly volatile. There are a lot of interacting factors that together drive the volume of CSAM in the Dutch hosting networks. Better detection leads to more known CSAM. This is driven, among other things, by the crawlers operated by IWF). These crawlers are a black box. One month they detect a lot, the next month very little. A second factor is the behavior of the criminals – that is, the people uploading the CSAM. They chose certain services and they move their material elsewhere, when needed. Providers might suddenly be struck by a spike in CSAM, even though they did not change anything on their end. A third factor is the actions of the providers and the domain owners, like the adoption of the hash-check service.

These factors cannot be disentangled. So yes, sometimes there appears to a clear downward trend in the volume of CSAM. For example, when the minister announced in June 2020 that he had “turned over the hour glass” and that the clock was ticking for providers and their customer domains to act, it did seem like the volume of CSAM subsequently went down.[1] A major part of the downward trend, however, was driven by changes in the data supplied to EOKM. It is unclear that it has anything to do with the public-private actions. The reduction was also short-lived. In the course of 2021, the volume went up again, even though providers were doing more rather than less than before.

An important countermeasure has been the launch by EOKM of the so-called hash-check service. Some hosting providers have been urging their customers who struggle with CSAM on their domain to adopt this hash-check service.[2]. The list of domains that are using the hash-check service is confidential, as is the volume of queries that domains run against the service. So it is not possible for us to connect the trends in volume for certain domains against the list of users of the service.

[1] See https://www.rijksoverheid.nl/documenten/kamerstukken/2020/07/07/tk-voortgangsbrief-aanpak-online-seksueel-kindermisbruik-en-kindersekstoerisme.

[2] See for more information the letter to parliament by the minister of Justice and Security on the progress made in the fight against CSAM, available at: https://www.rijksoverheid.nl/documenten/kamerstukken/2020/07/07/tk-voortgangsbrief-aanpak-online-seksueel-kindermisbruik-en-kindersekstoerisme. See also: https://www.meldpunt-kinderporno.nl/over-ons/hash-database/

Of the current top 5 providers with most CSAM, we see a clear picture emerge. NForce has been exceptionally effective in removing CSAM. A perfect 100% of all URLs are taken down within 24 hours. Also in terms of volume, NForce has made a dramatic reduction. NForce informed us that they had asked certain customer domains to leave — domains that were struggling with preventing CSAM uploads. And indeed, we have seen a dramatic reduction in the volume of CSAM in the NForce network, reducing to zero URLs in the second half of January. This can be seen as a direct result of the public-private policies to reduce CSAM in the Netherlands. The number 2, KnownSRV, has been consistently in the top 5 over the last two years. Its volume is not going down. It is also less responsive than NForce in removal. It does manage to remove over 80% in 24 hours, but about 15% remained online beyond 48 hours.

Then we have new entrants HostSlim B.V., Telegram Messenger Inc and MSK.HOST Hosting. Their performance in terms of removal is worse than KnownServe, ranging from weak to to extremely poor. Telegram Messenger Inc has left more than 90% of the CSAM online for more than 48 hours. Interestingly enough, Telegram Messenger Inc did receive a letter from the ministry that they in the top 15 providers hosting CSAM and asking them to take action. In absolute amounts, the newcomers do not have large numbers of URLs, because the total volume of CSAM has gone down in the second half of 2021. They do have a large share of the volume of CSAM, though, mostly because the share of NForce has been dramatically reduced.

Some of the earlier top 5 providers, like Leaseweb NL and IP Volume, have been successful in reducing the relative amount of CSAM on their networks. Leaseweb NL also demonstrated a perfect score for removal within 24 hours. For IP Volume no CSAM URLs were reported during our uptime measurement, which is in itself a positive sign. It also meant that we could not measure their takedown speed in the most recent measurement.

One way of summarizing these results is to say that the providers that have been the focus of the public-private initiative to combat CSAM hosting have been responsive to the policy. They were already actively removing the material, but many of them increased their efforts. In response, the problem has shifted to different providers.

For the monitor, the next steps are to observe these changes over a longer time frame, and to develop better automation for the measurements of takedown speed, so we might be able to better identify the effects of the interventions.