Rethinking the Detection of Child Sexual Abuse Imagery on the Internet

Elie Bursztein[◇] Travis Bright[‡] Michelle DeLaune[†] David M. Eliff[†] Nick Hsu[†]
Lindsey Olson[†] John Shehan[†] Madhukar Thakur[◇] Kurt Thomas[◇]

Google[◇] National Center for Missing and Exploited Children (NCMEC)[†] Thorn[‡]

ABSTRACT

Over the last decade, the illegal distribution of child sexual abuse imagery (CSAI) has transformed alongside the rise of online sharing platforms. In this paper, we present the first longitudinal measurement study of CSAI distribution online and the threat it poses to society's ability to combat child sexual abuse. Our results illustrate that CSAI has grown exponentially—to nearly 1 million detected events per month—exceeding the capabilities of independent clearinghouses and law enforcement to take action. In order to scale CSAI protections moving forward, we discuss techniques for automating detection and response by using recent advancements in machine learning.

KEYWORDS

child sexual abuse; child exploitation; online abuse

ACM Reference Format:
Elie Bursztein, Travis Bright, Michelle DeLaune, David M. Elifff, Nick Hsu, Lindsey Olson, John Shehan, Madhukar Thakur, Kurt Thomas. 2019. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308558.3313482

1 INTRODUCTION

Child sexual abuse is a horrific crime affecting an estimated 9–19.7% of girls and 3–7.9% of boys [2, 31, 38]. These children endure indecent exposure, forced intercourse, and sex trafficking [33]. Even after escaping abusers, victims face immediate and lasting consequences such as a heightened risk of depression, substance abuse, and suicide [5, 23].

Compounding the enormity of this situation, in the last decade, child sexual abuse imagery (CSAI) has evolved to leverage online sharing platforms as a tool for distributing abusive videos and images. The children depicted are re-victimized every time the content is accessed. According to the Internet Watch Foundation, which reviews reports of CSAI in the United Kingdom, 53% of CSAI images depict the abuse of children under the age of ten and 28% of CSAI images involve rape and sexual torture [20].

Platform operators including Google, Microsoft, Facebook, and Twitter have responded by proactively blocking attempts to access, share, or distribute CSAI [1, 9, 28, 32]. This involves scanning user-generated content using fingerprints of known abusive images [29] and reporting all incidents, as required by law, to independent clearinghouses that then manually process and action reports. Likewise, law enforcement actively disrupts criminal communities and web forums that foster child sexual abuse [16]. For example, law enforcement agencies dismantled the W0nderland Club, an online network of 300 participants who shared over 750,000 photos of child sexual abuse taken of 1,236 children [4].

Despite intense scrutiny into CSAI on the part of researchers and law enforcement, the broad scale of the problem and the effectiveness of current solutions for protecting online sharing platforms remains unknown. We address this gap by conducting the first longitudinal measurement study of CSAI distribution across the Internet. In collaboration with the National Center for Missing and Exploited Children (NCMEC)—the United States clearinghouse for all CSAI content detected by the public and online sharing platforms in the region—we examine metadata associated with 23.4M incidents of CSAI from 1998–2017. Our study captures the exponential growth of detected CSAI content, the expanding international scope of abuse, and the evolution of the technologies and mediums used to create and distribute CSAI content online.

Overall, our analysis reveals that online sharing platforms have accelerated the pace of CSAI content creation and distribution to a breaking point where NCMEC's manual review capabilities and law enforcement investigations no longer scale. To cope with the current volume of nearly one million CSAI reports per month, we argue there is a pressing need for these institutions and processes to leverage technological advancements to automate CSAI detection and response. We outline promising directions, such as extensive use of deep-learning, that address pain points highlighted in our measurement findings.

In summary, the key insights of this study are:

CSAI reports are growing exponentially. Of 23.4M reports of CSAI, 9.6M (40%) occurred in 2017 alone. The ability of clearing houses to manually investigate each incident cannot scale to this volume of reports absent automated prioritization, labeling, and clustering.

Protecting against CSAI requires coordinated, global action. Ten years ago, 70% of CSAI reports reflected abuse in the Americas. Today, 68% of reports relate to abuse in Asia, 19% the Americas, 6% Europe, and 7% Africa.

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313482

New CSAI content is constantly emerging. 84% of detected CSAI images and 91% of videos are reported only once. The specter of potential false negatives due to unknown abusive imagery requires moving from the existing, decades old blacklist-based approaches to algorithms that recognize the nature of CSAI content.

Criminals readily adopt new technologies. Over the years, distribution vectors have changed from email and FTP, to peer-to-peer, and now Tor and resurgent chat and messaging apps. Protections must keep pace with advancements in both distribution channels and digital mediums

2 CHILD SAFETY LAWS & REPORTING

Before diving into our study, we briefly outline the laws in the United States surrounding child sexual abuse imagery (CSAI) and the process that the public and companies use to report abuse to the National Center for Missing and Exploited Children (NCMEC) [30]. These reports form the basis of our study.

2.1 Reporting requirements

CSAI refers to any visual depiction of sexually explicit content involving a minor under the age of 18. United States law prohibits the distribution, production, receipt, and possession of any such material (see generally 18 U.S. Code §2251, 2252). Furthermore, any United States-based electronic service provider (ESP) must report any apparent CSAI to the CyberTipline operated by NCMEC [30].

We note that similar laws exist in 82 countries around the world such as the 1978 Protection of Children Act [13] and the 1988 Criminal Justice Act [14] in UK. Many countries also have CSAI clearing houses—including Canada with the Canadian Centre for Child Protection [12] and United Kingdom with the Internet Watch Foundation [18].

Our study focuses on NCMEC reports as NCMEC is by far the largest clearing house in the world: In 2016 NCMEC received 8.2 million CSAI reports, compared to 39,714 for the Centre for Child Protection [6] and roughly 105,000 for the Internet Watch Foundation during the same period [20]. The large volume gap between NCMEC and the other entities is due to the fact that many large US companies report all material detected on their platforms to NCMEC, with NCMEC routing reports to the correct law enforcement agency.

2.2 Life cycle of a NCMEC report

NCMEC's CyberTipline was first launched in 1998 and has since served as a clearing house that coordinates between the public, ESPs, and law enforcement. We provide a rough breakdown of the life cycle of a NCMEC report in Figure 1. Reports first stem from a public entity observing an abusive actor accessing or distributing CSAI material, or by an ESP detecting such activities on their service. Automated detection tools currently include, among others, Microsoft's PhotoDNA for images [29] and Google's CSAI Match for videos [42]. Both tool leverage a variation of the local-sensitive hashing algorithm [10] to not only detect known bad content but also slightly modified versions that abusers produce in an attempt to avoid detection. These CSAI detection technologies are deployed by many companies including Google, Microsoft, Facebook, and Twitter [1, 9, 28, 32].

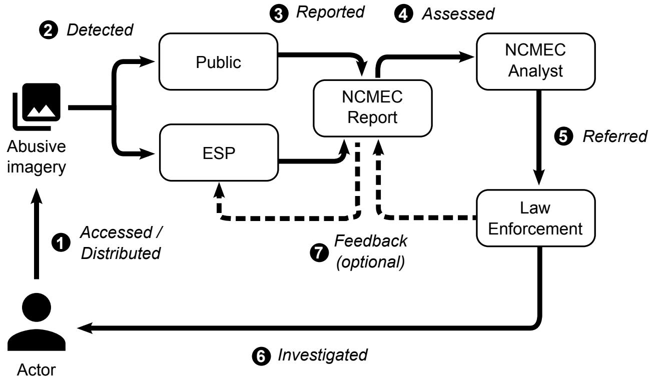

Figure 1: Life cycle of a CSAI report. Upon detection of an actor distributing or accessing abusive imagery, the public or an ESP report the activity to NCMEC. A NCMEC analyst reviews the report and refers actionable leads to law enforcement.

Once NCMEC receives a report of suspected CSAI, an analyst reviews, augments, and deconflicts the report. As part of this, the analyst manually characterizes the type of incident and the distribution channel involved. The analyst also infers the likely location of the abusive actor and victim, using among multiple factors, the IP address embedded in each abuse report. Ultimately, the analyst assesses the priority of the report for law enforcement referral.

NCMEC refers all US cases to United States state or federal law enforcement. For those reports related to abusive actors or victims outside the United States, NCMEC works with Interpol, Europol, and national police forces. These agencies then investigate reports, take action when appropriate, and may provide optional feedback to NCMEC (and through NCMEC potentially to ESPs) on the outcome. During our study we found out that fewer than 3% of all NCMEC reports have such optional feedback. This precludes any detailed analysis on the actionability of reports.

3 DATASET

Our study builds on a snapshot of anonymized metadata associated with the 23,494,983 NCMEC reports related to suspected CSAI that NCMEC received from March 1, 1998 when the CyberTipline was created till September 30, 2017. As mentioned earlier these reports come from the US public and many US ESPs. In this section, we provide an overview of the data used through the study and discuss inherent limitations that might bias our study.

3.1 Metadata per Report

Every report consists of at least a timestamp for when NCMEC received the report, the content reported as well as optional metadata such as the IP address of the uploader. As reports have evolved since 1998, more recent reports contain more details. Here are the key metadata fields we use throughout this study:

Source. Reports originate from one of two sources: the public, or electronic service providers (ESPs). For privacy reason, our metadata is limited to these coarse groupings; researchers involved in the study never obtained access to detailed information on reporting entities.

File hashes. Starting in 2013, report metadata contains an MD5 hash of every image, video, and other file attached to the reports. We note that more recent reports include a PhotoDNA hash (discussed in Section 2) and other local sensitive hash, but we did not use these due to the lack of historical coverage.

File format. Starting in 2013, NCMEC's staff began tracking and annotating CSAI content present in each report into one of three file formats: images, videos, and other (e.g., PDFs, HTML). In total, reports reference 20,568,240 abusive images, 2,335,536 abusive videos, and 337,575 other files.

Distribution vector. When possible, NCMEC's staff will annotate each report with a distribution vector. A majority of CSAI distribution vectors are either URLs (90.5%) or unidentifiable due to insufficient report details (8.5%). Other vectors in the remaining 1% of reports include IRC channels, email messages, chat applications, P2P websites, and Tor. We discuss these in further detail in Section 4.

Country. When possible, NCMEC's staff will annotate each report with a country representing either the victim or the abusive actor accessing or distributing content. We note this location information is overloaded: it can reflect the server where CSAI is hosted, the location of the person accessing the CSAI, or even the suspected location of a victim. As such, we are cautious when drawing conclusions about global CSAI trends.

3.2 Limitations

Our dataset is skewed towards US ESPs that proactively scan content using automated tools. As such, our findings likely underestimate the volume of CSAI content that actors access or distribute online. Furthermore, our data is skewed towards the market share of reporting ESPs; for example, China is entirely absent from our dataset. For confidentiality reasons, we are unable to share the identities of any of the ESPs involved or the relative volumes of CSAI reports they contribute. Despite these limitations, we believe that our large-scale dataset provides the first portrait of the evolution of CSAI online and the challenges faced by clearing houses and law enforcement. We discuss these challenges later in Section 5.

3.3 Ethics

Working with CSAI reports presents significant ethical hurdles. Due to the sensitivity and legal protections surrounding reports, the researchers involved in this work never had direct access to reports, or even the metadata of reports. We constructed all measurements based of a report's schema, which were then executed by NCMEC's staff on their infrastructure after vetting by NCMEC's leadership. Additionally, to avoid re-victimizing the children involved in cases reported to NCMEC, and to avoid adding additional burdens to NCMEC's staff, we avoided any reprocessing of old reports. This also precludes analyzing images to understand the nature of abuse, the age of victims, or the uniqueness of the content beyond the hashes present in a report's metadata.

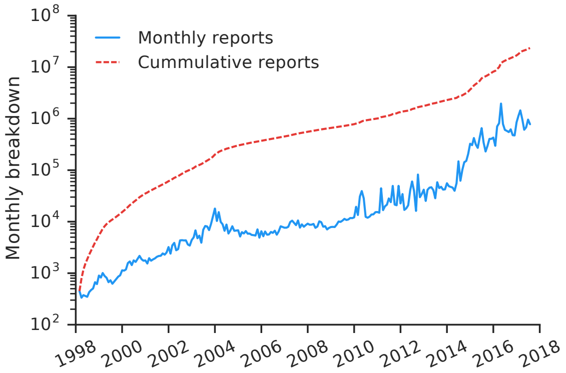

Figure 2: Monthly volume of CSAI reports (log scale) received by NCMEC since the creation of the CyberTipline in 1998. An exponential increase in report volume has lead to 9.6M reports in the last year alone.

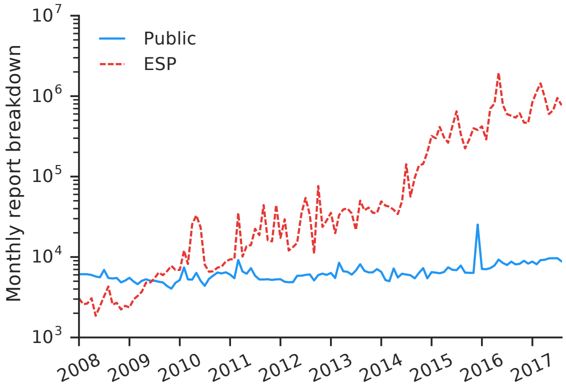

Figure 3: Breakdown of CSAI reports by the public and ESPs. Since 2009, an exponential growth in proactive, automated reporting has overtaken the public as the primary source of reports.

4 ANALYSIS

Our longitudinal measurement study surfaces how reports of CSAI have grown exponentially in parallel with online sharing platforms. In particular, we highlight the global nature of abuse, the adaptation of abusers to new technologies and mediums, and the frequent appearance of new content. Our measurements underscore the challenges of scaling manual review and law enforcement as a response to this threat.

4.1 Exponential Report Growth

Since the founding of the CyberTipline, NCMEC's report volume has grown a median of 51% year-over-year as shown in Figure 2. In aggregate, the public and ESPs have reported over 23.4M suspected incidents of CSAI. Over 9.6M reports (40%) occurred in the year 2017 alone—nearly 1 million per month. For contrast, NCMEC received only 565,000 reports (2.4%) in its first ten years of operation.

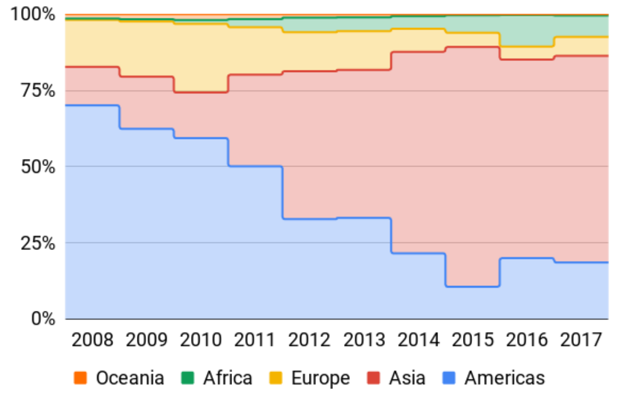

Figure 4: Annual breakdown of CSAI reports by geographic region. NCMEC has transformed from a regional reporting body to a global clearing house of CSAI reports.

Country	Reports	Frac. reports	Normalized per 1K capita
India	3,880,235	24.1%	11.9
Indonesia	1,743,362	10.8%	31.0
Thailand	1,706,055	10.6%	63.8
Mexico	1,243,028	7.7%	17.8
Bangladesh	985,347	6.1%	40.5
United States	792,957	4.9%	3.3
Brazil	719,704	4.5%	6.0
Vietnam	701,315	4.3%	14.1
Algeria	631,190	3.9%	41.8
Pakistan	567,850	3.5%	15.8

Table 1: Top 10 countries flagged in reports. To account for population, we also consider the volume of reports per 1000 estimated Internet users for each country.

A major contributor to the observed exponential growth is the rise of proactive, automated detection efforts by ESPs, as shown in Figure 3. In July, 2009, for the first time the number of reports generated by ESPs exceeded that of the number of public reports. This coincides with the release of Microsoft's PhotoDNA tool (discussed in Section 2). Since then, reporting by ESPs increased an average of 101% year-over-year, likely due to increasing user bases and an influx of user-generated content. While automated detection solutions help ESPs scale their protections, law enforcement and NCMEC analysts currently contend with the deluge of reports in a non-automated fashion as they are required to manually reviews the reports.

4.2 Global Threat

Over the last ten years, NCMEC has transformed from a regional reporting body—with 70% of reports related to victims or abusive entities in the Americas and 54% from the United States—to a worldwide clearing house with 68% of its reports relating to Asia, 19% to Americas, 6% to Europe, 7% to Africa, and 0.04% Oceania (Figure 4). Consequently, combating CSAI increasingly requires broad international cooperation with law enforcement.

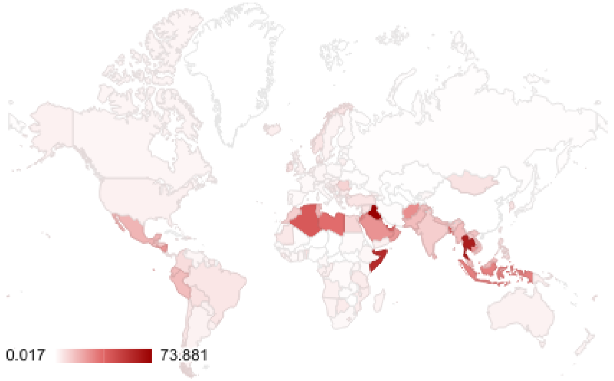

Figure 5: Estimated volume of CSAI reports per one thousand Internet users. South-East Asia, the Middle East, and North Africa have the highest incident rates.

We provide a breakdown of reports by the top 10 locations of victims and abusive entities (when discernible by analysts) in Table 1. As this information is often based on the IP address for abusive entities, we caution it may include VPNs or proxies that skew geographic distributions. We find that India, Indonesia, and Thailand account for roughly 37% of all reports. Both Indonesia and Thailand lack a legal definition for "child pornography" and do not require ISP reporting [18], potentially contributing to CSAI in the region.

As the populations of the top 10 countries differ by an order of magnitude, we also estimate the volume of reports per 1,000 Internet users (based on the International Telecommunication Union Internet penetration estimates for 2015 [19]). After this normalization, the top three countries involved in CSAI include Iraq, Thailand, and Somalia. Both Iraq and Somalia lack any laws related to child sexual exploitation [18]. We visualize these normalized rates across the globe in Figure 5, excluding countries with fewer than 500 reports. We observe higher incident rates of CSAI involvement in South-East Asia, the Middle East, and North Africa. A multitude of factors likely play a role in this, including the lack of criminalization [18], the presence of sexual exploitation in conflict zones [8], and entrenched commercial production [22].

4.3 Proliferation of Contexts

As technologies for communicating and sharing over the Internet have shifted, CSAI distribution has evolved in kind (Figure 6). While over 91% of reports since 1998 reference a URL, the minority of other reports provide a lens into how distribution channels fall in and out of favor over time. For example, prior to 2004, abusive entities relied on FTP and email attachments before a rise in peer-to-peer technology. In recent years, we find some sophisticated abusers have adopted Tor (likely to avoid law enforcement authorities), while less sophisticated abusers have favored the convenience of chatroom and messaging apps. One other notable trend is the rise of gaming communities in the last 2–3 years (where children are likely to be participants), both as an enticement platform and as a distribution

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Sample size
FTP	9%	10%	22%	25%	9%	5%	3%	6%	1%	1%	1%	1%	0%	1%	1%	0%	1%	0%	2%	0%	274
Email	0%	1%	1%	2%	9%	16%	18%	8%	3%	4%	6%	4%	3%	4%	3%	4%	5%	4%	3%	2%	86,601
P2P	0%	0%	0%	2%	9%	10%	7%	9%	11%	11%	6%	6%	5%	2%	1%	1%	2%	7%	1%	9%	8,900
Chatroom/IRC	1%	1%	1%	1%	2%	2%	3%	3%	7%	10%	6%	3%	2%	1%	2%	2%	2%	5%	22%	23%	36,086
Instant Messanger	1%	1%	1%	2%	3%	3%	3%	3%	10%	17%	7%	4%	2%	2%	4%	2%	3%	1%	13%	19%	13,733
Forum	2%	4%	3%	3%	4%	4%	4%	5%	5%	7%	3%	6%	7%	24%	2%	2%	2%	2%	7%	4%	30,743
Gaming	0%	0%	0%	0%	0%	0%	0%	0%	0%	12%	18%	5%	3%	3%	7%	8%	9%	9%	15%	10%	3,838
SMS	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	13%	12%	10%	18%	12%	13%	8%	4%	7%	1%	511
Cell phone	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	2%	1%	4%	10%	11%	17%	27%	25%	38,711
Tor	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	3%	5%	26%	42%	22%	4,427
URL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	2%	2%	5%	20%	39%	26%	21,431,212

Figure 6: Evolution of technologies involved in the distribution of CSAI. For example, 25% of all CSAI found on FTP servers was reported in 2001. Conversely, 42% of CSAI content distributed via Tor was reported in 2016. Our findings illustrate how technologies other than URL hosting fall in and out of favor over time.

mechanism. While the overall volume of reports in these mediums is small (possibly due to a lack of ESP monitoring), our data illustrates the broad cooperation necessary among Internet services to tackle CSAI and the adaptability of abusers to new platforms.

4.4 Rise of Videos

In conjunction with video-capable smartphones and the availability of higher bandwidth, we find that CSAI video sharing has dramatically increased from under 1,000 reports per month in 2013 to over 2 million reports per month (Figure 7). In the last year alone, NCMEC observed a 379% year-over-year growth in reported CSAI videos, compared to an 18% increase in CSAI images. In contrast, other reports, such as HTML or PDFs, have remained largely stable over the years at 6,000 reports per month.

This exponential growth of video sharing highlights the necessity for online sharing platforms to develop robust and scalable detection. In particular there is an urgent need for platforms to agree on a standard for video fingerprinting, with Microsoft's PhotoDNA being designed for individual images. At present, ESPs use in-house technologies to flag CSAI videos. Additionally, the size and volume of videos poses a challenge for clearing houses to store and index content. With the advent of virtual and augmented reality, scaling will only increase as a future challenge.

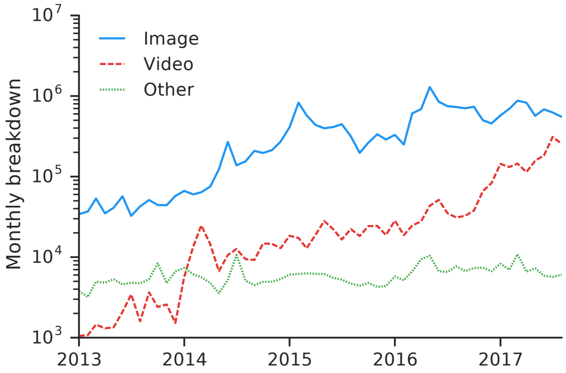

Figure 7: Volume of images, videos, and other file types reported each month since annotations appeared in our dataset in 2013. Simplified video sharing and recording has lead to a 379% increase in CSAI videos in 2017 alone.

4.5 Uniqueness & Lifetime of Imagery

CSAI content is constantly being transformed and created. Using each reported content's (non-PhotoDNA) hash as an indicator of uniqueness, we find that the public and ESPs report 84% of images and 91% of videos only a single time. Outliers exist, with reports flagging a single image that resurfaced 491,200 times and a single video that resurfaced 5,876 times. Our results illustrate the challenge of identifying new CSAI purely from approximate histograms of previously flagged images; algorithms instead need a deeper understanding of the abusive nature of CSAI content.

Re-distributed content offers an estimate of the lifetime of CSAI. Here, we calculate the elapsed time between the first and last report of CSAI for all hashes reported more than once (Figure 8). We caution this is strictly an underestimate. We find the median lifetime of an image was 257 days, while 10% of images resurfaced over the course of three years. For videos, which just recently emerged as a popular CSAI medium, the median lifetime was 210 days while 10% of videos resurfaced over the course of 1.5 years. Other media is skewed towards long-lived file bundles, where 70% of files resurfaced over the course of 3 years. These findings suggest that once CSAI content is uploaded to the Internet, it continues to resurface, with multiple months elapsing between attempts at distribution.

5 DISCUSSION

Our measurements highlight how the rapid adoption of the latest advancements in Internet platforms and digital mediums allows abusers to create and distribute CSAI across the Internet at an exponential rate. This exponential growth currently threatens our society's ability to combat online child abuse as the rate at which abusive content emerges vastly outpaces the processing capabilities of clearing houses and law enforcement agencies.

At its core this imbalance between abusers and defenders stems from the fact that the development of detection and processing technologies substantially lags behind the development of content creation and sharing capabilities. This section discuss the key technological directions that need to be pursued to reverse this trend.

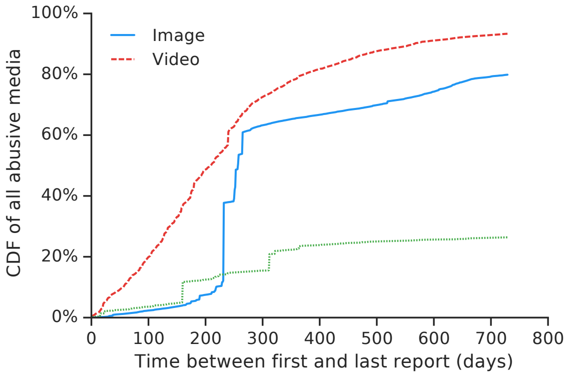

Figure 8: Time elapsed between the first and last report of a unique CSAI hash. We find that a median image has an observed lifetime of 257 days, while videos have a lifetime of 210 days.

Scaling the review process. For NCMEC and other clearing houses, a significant amount of time and human labor is spent processing, aggregating, and de-duplicating reports. Recent advancements in deep learning offer the opportunity to scale the review process by automatically grouping reports that have content in common. In particular, two types of clustering could help:

- **Scene clustering:** Content retrieval techniques [11, 27, 41] can be used to group reports coming form various ESP by looking for reports that have similar images/videos that are potentially from the same scene.
- **Facial clustering:** Another possibility involves leveraging facial recognition [7, 15, 37] to find reports that contains the same child or perpetrator. A key challenge here is that traditionally models, including the latest ones based on triplet-loss, have struggled with children's faces because due to less defined facial traits. This represents a challenge both for facial recognition and identifying key attributes such as gender or approximate age.

Scaling the investigation process. A key challenge for law enforcement is prioritizing the reports which are most likely to lead to an arrest. Deep learning techniques can potentially help with prioritization by automatically detecting which content is the most "actionable" from an investigation standpoint. In particular:

- **Facial detection:** Recent progress in facial detection [21, 26] and gender or age prediction [25, 34] lend themselves to automatically reporting the number of victims or perpetrators found in an image. Additionally, it may be possible to extend these technologies to also report if an image contains blurred faces or if identifying features such as hair and eye color are present.
- **Scene information:** Image classifiers can automatically report objects that are present in a scene [39, 40], an image's environment [35, 45], and potentially identify landmarks that will help locate the region where an abusive image originated [43].

Detecting new content. Microsoft's PhotoDNA, originally developed in 2009, is a specialized version of minhashing that fingerprints images based on the histogram of grid-segmented images [29]. This technique is resilient to image transformations related to scale, color, or angle, but is incapable of detecting new CSAI images that have yet to be flagged during manual review. Likewise, the technique does not currently scale to videos. A critical direction here is to develop new algorithms that automatically identify CSAI material based on an image's subject matter. Any such algorithm should work seamlessly across digital mediums, including images and videos.

Reducing reviewer burden. Human analysts currently play a critical role in reviewing CSAI images and videos, both to identify victims, but also to confirm the abusive nature of content. Finding ways to reduce the burden and emotional toll of this task—even through non-technical means—is more pressing than ever given the exponential increase in the amount of content undergoing review.

6 RELATED WORK

Prior investigations of the distribution of CSAI online have largely focused on criminal sharing in peer-to-peer networks. Bissias *et al.* studied roughly 300,000 CSAI torrents and found that, despite law enforcement actions against criminals sharing CSAI, torrents had a survival rate of over 60% even after four years [3]. Our metrics also show a long lifetime of CSAI content: images and videos resurfaced over the course of nearly a year. Other research has tried to estimate the number of criminals accessing CSAI based off of unique IP addresses tied to torrent requests [44], though such metrics are susceptible to DHCP churn from residential machines. In terms of detection and disruption, research has focused either on flagging queries for CSAI content based on known keywords [24, 36] or on targeting major seeders to prevent further distribution [17]. Given the frequency of content shared outside of torrents—our dataset of 23.4M reports included only 8,900 torrents—solutions beyond peer-to-peer networks are critical.

7 CONCLUSION

In this paper, we showed how CSAI on the Internet has outpaced the capabilities of independent clearinghouse analysts and law enforcement to respond. In 2017 alone, NCMEC received 9.6M reports of CSAI, compared to roughly 10,000 reports a year when NCMEC first started in 1998. This exponential growth and the frequency of unique images requires re-imagining CSAI defenses away from fingerprint-based detection and manual review. Instead, we argued that researchers need to develop algorithms that automatically detect CSAI content, cluster images and videos of victims, and ultimately surface identifying features to help law enforcement. The cost of human review also requires both technical and non-technical solutions that reduce the emotional toll of examining CSAI content. Absent new protections, society will be unable to adequately protect victims of child sexual abuse.

REFERENCES

[1] Charles Arthur. Twitter to introduce photodna system to block child abuse images. https://www.theguardian.com/technology/2013/jul/22/twitter-photodna-child-abuse, 2013.

[2] Jürgen Barth, Lilian Bermetz, Eva Heim, Sven Trelle, and Thomai Tonia. The current prevalence of child sexual abuse worldwide: a systematic review and meta-analysis. *International journal of public health*, 2013.

[3] George Bissias, Brian Levine, Marc Liberatore, Brian Lynn, Juston Moore, Hanna Wallach, and Janis Wolak. Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks. *Child Abuse & Neglect*, 2016.

[4] Martin Bright and Tracy McVeigh. This club had its own chairman and treasurer. its business was child abuse. https://www.theguardian.com/uk/2001/feb/11/tracymcveigh.martinbright, 2001.

[5] Angela Browne and David Finkelhor. Impact of child sexual abuse: A review of the research. *Psychological bulletin*, 1986.

[6] Canadian Center for Child Protection. Cybertip.ca 15-year anniversary report. https://www.cybertip.ca/pdfs/CTIP_15thAnniversaryReport_en.pdf, 2017.

[7] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: a deep quadruplet network for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2017.

[8] Clar Ni Chonghaile. Millions of iraqi children repeatedly and relentlessly targeted, says un. https://goo.gl/vZ33bg, 2016.

[9] Facebook. Meet the safety team. https://www.facebook.com/notes/facebook-safety/meet-the-safety-team/248332788520844, 2011.

[10] Aristides Gionis, Piotr Indyk, Rajeev Motwani, et al. Similarity search in high dimensions via hashing. In *Vldb*, volume 99, pages 518–529, 1999.

[11] Albert Gordo, Jon Almazan, Jerome Revaud, and Diane Larlus. End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, 124(2):237–254, 2017.

[12] Canada government. Canadian centre for child protection, 1985.

[13] UK government. Protection of children act, 1978.

[14] UK government. Criminal justice act, 1988.

[15] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[16] Thomas J Holt, Kristie R Blevins, and Natasha Burkert. Considering the pedophile subculture online. *Sexual Abuse*, 2010.

[17] Ryan Hurley, Swagatika Prusty, Hamed Soroush, Robert J Walls, Jeannie Albrecht, Emmanuel Cecchet, Brian Neil Levine, Marc Liberatore, Brian Lynn, and Janis Wolak. Measurement and analysis of child pornography trafficking on p2p networks. In *Proceedings of the International Conference on World Wide Web*, 2013.

[18] International Center for Missing and Exploited Children. Child pornography: model legislation and global review. http://www.icmec.org/wp-content/uploads/2016/02/Child-Pornography-Model-Law-8th-Ed-Final-linked.pdf, 2016.

[19] International Telecommunication Union. Percentage of individuals using the Internet. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx, 2017.

[20] Internet Watch Foundation. IWF annual report 2016. https://www.iwf.org.uk/sites/default/files/reports/2017-04/iwf_report_2016.pdf, 2017.

[21] Huaizu Jiang and Erik Learned-Miller. Face detection with the faster r-cnn. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 650–657. IEEE, 2017.

[22] Christine Joffres, Edward Mills, Michel Joffres, Tinku Khanna, Harleen Walia, and Darrin Grund. Sexual slavery without borders: trafficking for commercial sexual exploitation in india. *International Journal for Equity in Health*, 2008.

[23] Kathleen A Kendall-Tackett, Linda M Williams, and David Finkelhor. Impact of sexual abuse on children: a review and synthesis of recent empirical studies. *Psychological bulletin*, 1993.

[24] Matthieu Latapy, Clémence Magnien, and Raphaël Fournier. Quantifying paedophile queries in a large p2p system. In *Proceedings of INFOCOM*, 2011.

[25] Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42, 2015.

[26] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5325–5334, 2015.

[27] Kevin Lin, Huei-Fang Yang, Jen-Hao Hsiao, and Chu-Song Chen. Deep learning of binary hash codes for fast image retrieval. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 27–35, 2015.

[28] Microsoft. Microsoft and National Center for Missing and Exploited Children push for action to fight child pornography. https://goo.gl/7oxEtW, 2009.

[29] Microsoft. PhotoDNA cloud service. https://www.microsoft.com/en-us/photodna, 2018.

[30] NCMEC. National center for missing and exploited children, 1984.

[31] Noemí Pereda, Georgina Guilera, Maria Forns, and Juana Gómez-Benito. The prevalence of child sexual abuse in community and student samples: A meta-analysis. *Clinical psychology review*, 2009.

[32] Sarah Perez. Why the gmail scan that led to a man's arrest for child porn was not a privacy violation. https://tcrn.ch/2NlbBrW, 2014.

[33] Frank W Putnam. Ten-year research update review: Child sexual abuse. *Journal of the American Academy of Child & Adolescent Psychiatry*, 2003.

[34] Pau Rodríguez, Guillem Cucurull, Josep M Gonfaus, F Xavier Roca, and Jordi Gonzalez. Age and gender recognition in the wild with deep attention. *Pattern Recognition*, 72:563–571, 2017.

[35] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[36] Moshe Rutgaizer, Yuval Shavitt, Omer Vertman, and Noa Zilberman. Detecting pedophile activity in bittorrent networks. In *Proceedings of the International Conference on Passive and Active Network Measurement*, 2012.

[37] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[38] Marije Stoltenborgh, Marinus H Van Ijzendoorn, Eveline M Euser, and Marian J Bakermans-Kranenburg. A global perspective on child sexual abuse: meta-analysis of prevalence around the world. *Child maltreatment*, 2011.

[39] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[40] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[41] Ji Wan, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 157–166. ACM, 2014.

[42] Nicholas Watt and Juliette Garside. Google to tackle images of child sexual abuse with search and youtube changes. https://www.theguardian.com/technology/2013/nov/18/uk-us-dark-web-online-child-abuse-internet, 2013.

[43] Tobias Weyand, Ilya Kostrikov, and James Philbin. Planet - photo geolocation with convolutional neural networks. *Lecture Notes in Computer Science*, page 37–55, 2016.

[44] Janis Wolak, Marc Liberatore, and Brian Neil Levine. Measuring a year of child pornography trafficking by us computers on a peer-to-peer network. *Child Abuse & Neglect*, 2014.

[45] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495, 2014.