Tennessee v. Meta
Case No. 23-1364-IV
PX-06423

JUNE 2024

Trends in Financial Sextortion

An investigation of sextortion reports in NCMEC CyberTipline data

Research conducted by Thorn in partnership with the National Center for Missing & Exploited Children

THORN

NATIONAL CENTER FOR MISSING & EXPLOITED CHILDREN

© 2024 Thorn

Table of Contents

4 Introduction

6 Key Findings

7 Methods and Limitations

9 How Sextortion Unfolds and to Whom

15 Tactics of Pressuring Victims and Victim Impacts

20 The Role of Platforms in Sextortion

24 Perpetrator Differences by Country

28 Platform Reporting Landscape

34 Conclusions

© 2024 Thorn

Acknowledgments

Understanding the complex intersection of technology and child sexual abuse empowers us to safeguard kids from the ever-evolving threats they face online. This can only be done through multi-stakeholder collaboration that builds on the insights and talents of all in the ecosystem: nonprofits; policymakers; tech companies; communities; and most importantly, the young people we commit to serve.

This research was conducted by Thorn in partnership with the National Center for Missing & Exploited Children.

The National Center for Missing & Exploited Children (NCMEC) is a private, nonprofit 501(c)(3) corporation whose mission is to help find missing children, reduce child sexual exploitation, and prevent child victimization. For more information about NCMEC, please visit www.ncmec.org.

Thorn is a 501(c)(3) nonprofit organization with a mission to build technology to defend children from sexual abuse. For more information about Thorn, please visit www.thorn.org. For inquiries about this research, please email research@thorn.org.

Research team

Tim O'Gorman, Thorn
Melissa Stroebel, Thorn
Rob Wang, Thorn
Cassie Scyphers, Thorn
Jennifer Penrose, National Center for Missing & Exploited Children
Ashleigh Chambers, National Center for Missing & Exploited Children
Jennifer Newman, National Center for Missing & Exploited Children
Lauren Coffren, National Center for Missing & Exploited Children

Design and publication

Yena Lee, Thorn
Cassie Coccaro, Thorn
Heidi Mihelich, cre8ivenergy

Suggested citation

Thorn and National Center for Missing and Exploited Children (NCMEC). (2024). Trends in Financial Sextortion: An investigation of sextortion reports in NCMEC CyberTipline data.

© 2024 Thorn

Introduction

DEFINITION

Sextortion

Threatening to expose sexual images of someone if they don't yield to demands.

Sextortion — threatening to expose sexual images of someone if they don't yield to demands — has been a source of harm to youth for some time, but it has gained added urgency in recent years. Over time, several studies have examined how this abuse takes shape, its prevalence, and those impacted.[1] Importantly, while sextortion can affect all ages, this report focuses explicitly on the sextortion of minors.

Between 3.5% and 5% of people are believed to have experienced sextortion before reaching adulthood,[2] with girls more likely than boys to be impacted. Historical surveys[3] have found demands most often were sexual or relational in nature, including but not limited to demands for additional intimate imagery, engaging in sexual acts, or returning or staying in a romantic relationship. Research has also found the source of threats is mixed, with roughly half coming from people in a victim's offline community, such as acquaintances or romantic partners/ex-partners, and the other half involving people they met online.[4]

In the last several years, concerns about a unique form of sextortion — financial sextortion — have been on the rise. Distinct from more often observed forms of sextortion, which frequently impacted girls and involved demands that were sexual or relational in nature, financial sextortion appears to more often impact boys and involves demands specifically for money. In addition, financial sextortion marks the emergence of new organized endeavors leveraging the internet to engage in financial sextortion at scale.

In both cases, the impact on children can be devastating, leading to severe trauma and, in extreme cases, suicide due to sextortion. Older surveys of sextortion victims found that 12% reported they "moved to a new neighborhood, community or town" and that 24% reported that they "saw a mental health or medical practitioner as a result of the incident."[5]

The National Center for Missing & Exploited Children (NCMEC) has received more than 144 million reports, as of year-end 2022,[6] of possible online child sexual exploitation, including sextortion, and was among the first organizations to raise alarms about the rise of financial sextortion. This report provides a deep dive into the reports submitted to NCMEC

1 Wolak, Janis, David Finkelhor, Wendy A Walsh and Leah Treitman. "Sextortion of Minors: Characteristics and Dynamics." The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine 62 1 (2018): 72 - 79.; Cross, Cassandra, Karen M. Holt and Roberta Liggett O'Malley. "'If U Don't Pay they will Share the Pics': Exploring Sextortion in the Context of Romance Fraud." Victims & Offenders 18 (2022): 1194 - 1215.
2 Patchin, Justin W. and S. Hinduja. "Sextortion Among Adolescents: Results From a National Survey of U.S. Youth." Sexual Abuse: A Journal of Research and Treatment 32 (2020): 30 - 54.; Finkelhor, David, Heather A. Turner and Deirdre Colburn. "Prevalence of Online Sexual Offenses Against Children in the US." JAMA Network Open 5 (2022)
3 Thorn (2017). Sextortion: Summary findings from a 2017 survey of 2,097 survivors. https://www.thorn.org/wp-content/uploads/2019/12/Sextortion_Wave2Report_121919.pdf
4 Thorn (2017). Sextortion: Summary findings from a 2017 survey of 2,097 survivors. https://www.thorn.org/wp-content/uploads/2019/12/Sextortion_Wave2Report_121919.pdf
5 Wolak, Janis and David Finkelhor (2016) "Sextortion: Findings from a Survey of 1,631 Victims." https://www.thorn.org/wp-content/uploads/2016/08/Sextortion_Report.pdf
6 https://www.ncmec.org/ourwork/impact

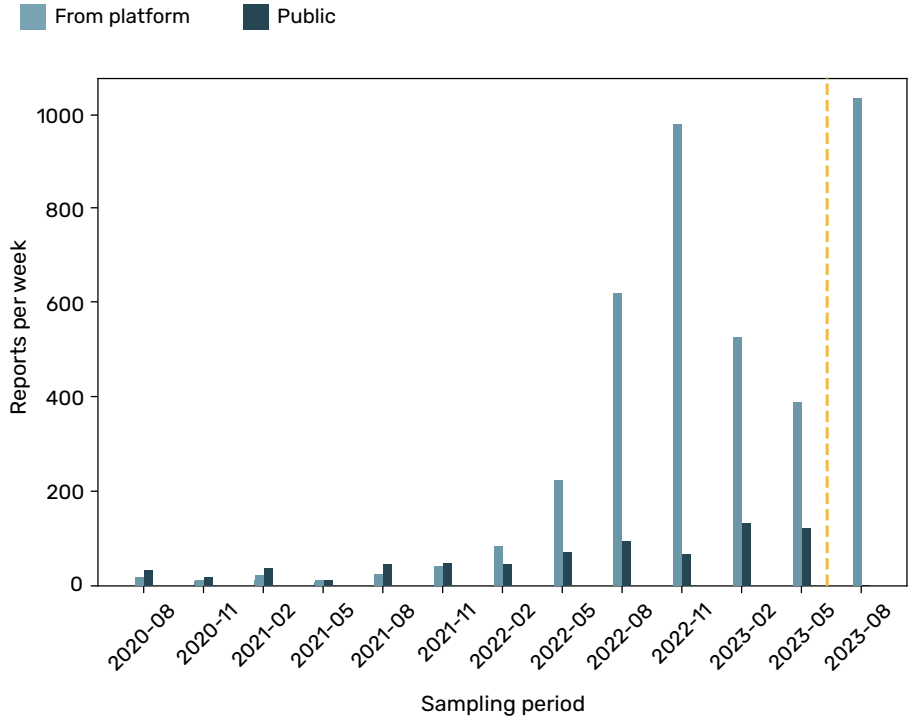
© 2024 Thorn

regarding sextortion, with a focus on the evolving trend of financial sextortion.

The overall trend in NCMEC reports shows a large wave of sextortion cases since the beginning of 2022. Although the numbers do not, on their face, differentiate among types of sextortion, analysis of report details demonstrates this increase is largely driven by reports involving financial sextortion. Figure 1 outlines the rates of all reports made to NCMEC and categorized as sextortion, showing an average of 812 reports of sextortion per week in the last year of data analyzed (from August 2022 to August 2023) and 559 reports/week in the last two years of data (from August 2021 to August 2023), which come from reports submitted to NCMEC by the public,[7] as well as many cases identified by Electronic Service Providers (ESPs) such as social media platforms, listed as "from platform." These rates have many details and limitations, which we measure through sampling and manual coding. Furthermore, there are limitations due to the nature of the reports submitted to NCMEC. For example, although we find that the vast majority of cases submitted in this period are financial in nature, we cannot know how much of this is due to cases, particularly nonfinancial sextortion, being underreported.

These numbers should not be viewed as vague statistics, but rather should be viewed as being many specific cases of children being targeted and extorted by perpetrators seeking to amplify their fears and for them to give in to demands. This report focuses on the chat logs and incident descriptions in these reports because they provide insight into how these incidents unfold and into the situation in which victims find themselves when they experience sextortion.

Fig 1 | **Sextortion Cases Per Week**



7 Note that for logistical reasons, the final period (2023-08) does not include analysis of reports submitted directly to NCMEC via public or hotline data, but that does not mean that no reports were submitted

© 2024 Thorn

Key Findings

Sextortion, and particularly financial sextortion, continues to be a major and ongoing threat, with an average of 812 reports of sextortion per week to NCMEC in the last year of data analyzed, and with reason to expect that the vast majority of those reports are financial sextortion.

(See How Sextortion Unfolds and to Whom)

Perpetrators leverage tactics to intentionally fan a victim's worry about the life-changing impacts of their nudes being shared — often repeating claims that it will "ruin their life."

(See Tactics of Pressuring Victims and Victim Impacts)

While we find that Instagram and Snapchat are the most common platforms used for sextortion, we observe trends regarding the emergence of additional end-to-end encrypted messaging apps to move victims to secondary platforms and the prevalence of Cash App and gift cards for methods of payment.

(See The Role of Platforms in Sextortion)

The two countries from which most sextortion perpetrators seem to be operating, Nigeria and Cote d'Ivoire, make use of slightly different tactics and platforms.

(See Perpetrator Differences by Country)

Reports submitted by Instagram constitute a clear majority of all reports of apparent sextortion submitted to NCMEC. However, there are reasons to worry not only about whether other platforms are underreporting but also about changes in the level of information provided in reports.

(See Platform Reporting Landscape)

© 2024 Thorn

Methods and Limitations

Methods of Analysis

NCMEC received over 32 million reports in 2022. US-based ESPs must report to the NCMEC CyberTipline if they become aware of child sexual abuse material (CSAM) on their platform[7]; these ESPs submitted 99% of NCMEC CyberTipline reports in 2022.[8] In the remaining cases we refer to as "public," a victim or member of the public directly reports through the public form (report.cybertip.org) or the NCMEC hotline.

In addition to reports of potential CSAM, NCMEC also receives reports concerning potential sextortion, grooming, or other forms of online enticement, but online enticement does not currently have the same legal reporting requirements for companies. These reports represent a small but growing percentage of the total reports received by NCMEC each year. NCMEC received 80,524 online enticement reports in 2022, an 82% increase over the 44,155 online enticement reports submitted the previous year.

To focus on a representative but reasonable amount of data, our analysis focused on a subset of reports received between 2020[9] and 2023. We defined four two-week periods each year for the last three years (every three months, starting on the 8th and ending on the 21st of February, May, August, and November) and studied all reports submitted to NCMEC within those periods (the sampled data totals more than 15 million reports).

Our analysis then highlighted all reports appearing to relate to sextortion, building off initial annotations provided by NCMEC analysts as part of the report intake process. For example, of the 4,366 reports identified as sextortion in 2022 in the eight weeks of our sampling windows, 1,938 reports (44%) were already identified as sextortion by that NCMEC analysis.[10] This report started from those annotations and augmented the initial sample of sextortion reports using machine learning algorithms to identify potential sextortion cases for additional annotation. Cases flagged through this process were then manually reviewed to verify they should be included in the research as a likely report of sextortion.

This study was limited to specific fields within CyberTipline reports, as prepared and provided by NCMEC, and did not include attached files such as screenshots of chat logs or other image files.

Across all sextortion cases in those sampling windows, we manually coded those reports to measure specific tactics for sextortion, including the role that platforms mentioned in those reports played in the sextortion. This was done so that we could study not only how often a platform was mentioned, but also more specific questions such as how often that platform was used to first contact the child. This measured how people talked about using platforms, but some of those uses may not necessarily happen, such as with empty threats to post imagery to a platform. When present in financial sextortion reports, we also manually coded monetary quantities with details regarding whether they referred to payments, mere demands, or payments followed by ongoing demands. Contact Thorn for an appendix listing all the labels and definitions used in the data coding and providing further details regarding the annotation methodology.

---

7 This is a requirement defined in federal statute 18 USC 2258A; see https://www.ncmec.org/cybertiplinedata for more details.
8 According to https://www.ncmec.org/cybertiplinedata
9 While sextortion predates 2020 (see prior NCMEC analysis, https://www.missingkids.org/content/dam/missingkids/pdfs/ncmec-analysis/sextortionfactsheet.pdf), this report is focused on the current wave of financial sextortion that started in earnest in 2022; data was studied from before 2022 in order to test whether the trend started earlier than expected.
10 This rate is lower in 2023, as our analysis of the final sampling period (August of 2023) was done before the rigorous stages of report intake were complete.

© 2024 Thorn

Limitations to This Report

- The most significant limitation of this work is that it only measures phenomena when explicitly stated in the report text. This is particularly important for reports where a platform (or victim) only submits a few sentences of summary text, since such brief summaries may not provide insight into how those cases unfolded.
- We attempted to limit the potential for measurement bias, but it is a complex process: While a large number of cases go through the rigorous review of analysts at NCMEC, additional coverage is gained by data scientists for this report. That additional coverage also used machine learning to surface cases for annotation, which always presents the possibility of model bias — further information on model quality is provided in the Appendix (available upon request from Thorn).
- Sampling limitations: We sampled from specific time windows, which means that these numbers are an estimate rather than a full measurement of all sextortion cases in a year.
- Fundamental reliance on platforms: For ESP reports, platforms can be biased in various ways in how they detect sextortion or how they respond to user reports of sextortion.
- Specific biases for low-resource languages and countries: For non-English reports, we relied upon automatic translation tools, and we expect some ESPs to also need to do so. This injects many possibilities for error but specifically does so for speakers of low-resource languages spoken in India, Southeast Asia, and Africa, for whom automatic translation quality is often far worse.
- This is a quickly moving space, both because platforms, perpetrators, and other actors (such as law enforcement) are constantly changing their tools and tactics, and also because of specific trends that have developed or continued in 2023, such as artificial intelligence [AI]–generated deepfakes for sextortion and shifts of specific platforms towards end-to-end encryption. Because the last time period we fully studied was August of 2023, our visibility into those recent developments is limited. We do reference some checks done on November 2023, but those data points did not receive the full manual analysis used for the rest of our data, and so we treat that data only as a tentative hint regarding whether trends seen in our analysis have persisted.
- While we tracked discussion of many apps and platforms, some methods of communication that are often discussed more generically (such as discussing texting without named apps, phone calls or email) — were kept out of the scope of the report due to the difficulty in discerning which tool was used. However, that should not mean that such tools may not be used for sextortion.
- There are times where multiple CyberTipline reports may be submitted for the same victim — such as if a child submitted a report to all platforms involved in their abuse. Deconflicting whether different reports referred to the same child was out of the scope of the current work and thus the numbers we report do have the risk of counting the same victim being represented multiple times.
- The NCMEC tracking of sextortion in public and hotline data was focused on studying the rising trend of financial sextortion, and evolved as more information became available. As a result of that focus, non-financial sextortion (demands for imagery or relationships) may be undercounted, and data from the earliest years of research (2020 and 2021) may have limited coverage due to the evolving understanding of the issue.
- The text provided in NCMEC reports can be of highly variable formats, and it was necessary to make a variety of assumptions regarding how to identify relevant portions of those texts and how to clean up chat logs for analysis. One may contact Thorn for an appendix with further details on the methods and assumptions used.

© 2024 Thorn

How Sextortion Unfolds and to Whom

The most common form of sextortion found in the studied sample is the financial sextortion of teenage boys. These reports most often include the use of "catfishing" — in this case, a perpetrator impersonating another young person — to manipulate a teenage boy into sharing sexual images or videos of himself. That perpetrator then threatens to share that imagery with family, friends, or followers unless they are paid. Although that is the prototypical scenario, sextortion can have demands other than money, target victims other than teenage boys, and can get imagery in ways other than catfishing.

This section outlines the range of differences seen in the NCMEC reports regarding what is demanded during sextortion, who the victims are, and how blackmail material is obtained or produced. The findings should be interpreted with an awareness that not all sextortion ends up in these reports: for example, many financial sextortion cases target young men rather than boys,[11] and historically, sextortion was often the extortion of girls by people they knew offline with demands for CSAM or sex.[12]

> **Established Information to Know:** The vast majority of financial sextortion cases of children seem to start with some form of "catfishing" targeting teenage males, convincing them to exchange images or get on a video call.

> **Key New Findings:** Of the reports where we have discussion of how the imagery was acquired, roughly 17% of situations involve either hacking or fake/inauthentic imagery. While this is a minority of the whole, one should not assume that every single sextortion starts with the child sharing their sexual imagery with the perpetrator.

The Vast Majority of Demands Are Financial, but Not All Are Explicit

The vast majority of sextortion cases reported to NCMEC in the years we studied are financial sextortion cases, in which the child is told to pay the perpetrator money in order to prevent the sharing of their intimate imagery. However, there are many other cases of sextortion in which a victim is extorted using their sexual imagery either for additional sexual imagery, for other sexual demands, or to stay in a relationship, and such nonfinancial demands were the predominant form of sextortion before the current financial sextortion crisis. Such cases have a very different distribution of victims as well. In surveys of survivors preceding the current financial sextortion trend, 83% of victims were female[13], and the majority of victims reported being extorted by people that they also knew offline (such as ex-partners or schoolmates), often in the context of intimate partner violence or stalking. An older study of sextortion in NCMEC data[14] (spanning 2013-2016) noted similar trends, reporting that 78% of the reports involved female children, and that only 7% of perpetrators in that period demanded money, with 5% of those perpetrators demanding to meet for sex and 78% demanding additional content of the child.

11 C3P (2022). An Analysis of Financial Sextortion Victim Posts Published on r/Sextortion. https://content.c3p.ca/pdfs/C3P_AnalysisOfFinanSextortionPostsReddit_en.pdf
12 Wolak, Janis and David Finkelhor (2016) "Sextortion: Findings from a Survey of 1,631 Victims." https://www.thorn.org/wp-content/uploads/2016/08/Sextortion_Report.pdf
13 Wolak, Janis and David Finkelhor (2016) "Sextortion: Findings from a Survey of 1,631 Victims." https://www.thorn.org/wp-content/uploads/2016/08/Sextortion_Report.pdf
14 https://www.ncmec.org/content/dam/missingkids/pdfs/ncmec-analysis/sextortionfactsheet.pdf

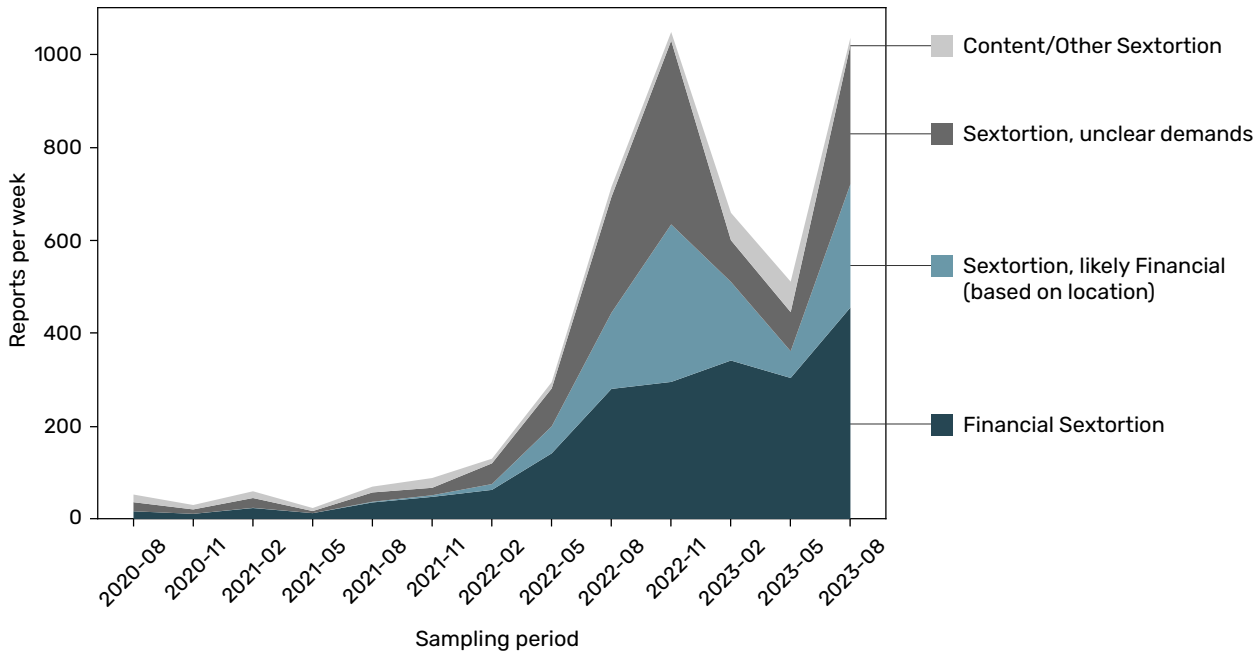
© 2024 Thorn

Figure 2a shows the spread among the current sample of sextortion cases reported to NCMEC, including both clear financial sextortion cases ("Financial sextortion") and some nonfinancial demands ("Content/other sextortion"). That same figure also shows a large middle ground of cases that are clearly sextortion (i.e., there is a threat to expose the child's imagery if they don't give in to demands) but where the data provided to NCMEC does not make it clear that the demands are monetary. These cases often refer to vague demands, requesting that the victim cooperate or accept the deal, but the specific payment details may have been communicated elsewhere (e.g., in the audio of a video chat). We have separated a subset of these — shown as "Sextortion, likely financial (based on their location)" — where we have a clear-cut reason to assume they are financial sextortion because they are reports that have been linked to Nigeria or Cote d'Ivoire, the two countries linked to the financial sextortion.

Such reports often contain threats similar (or even identical) to those seen in reports making explicit financial demands. To contextualize what this means for overall rates of sextortion: over the last 12 months of analysis, we see an average of 812 reports per week of sextortion total, of which 556 reports per week are financial or likely financial, and of which 348 reports per week have explicit monetary demands in the reports themselves. That rate of 556 financial sextortion reports per week would imply at least 28,000 financial sextortion cases per year, although that number is a conservative estimate rather than an official count.

Fig 2a | **Sextortion demands over time**

Reports per week, split by the demands of the perpetrator



812
sextortion reports of any kind per week, over the last year of estimates.

556
financial sextortion reports per week (explicitly or likely financial), over the last year of estimates.

© 2024 Thorn

Victims Being Targeted – Age, Gender

The vast majority of victims of financial sextortion submitted to NCMEC are male teenage victims; of minors in the NCMEC data with both age and gender, 90% were males between 14 and 17.[15]

90%
of victims were males between 14 and 17, of reports submitted to NCMEC public sources that had age and gender data.

For financial sextortion, it's important to remember that the reports submitted to NCMEC are connected to a larger trend of sextortion of young males that includes those 18 and older; the C3P study of financial sextortion discussions in Reddit data[16] also found predominantly male data (98%), but only 38% were under 18. However, the fact that financial sextortion of minors is part of a larger romance fraud trend, including young adults, should not be interpreted to mean that it is age indiscriminate or that adults are the only intended targets of sextortion. While we cannot know the intentions of perpetrators for certain, this report notes specific tactics and platforms that are possibly being used to gain access, or to gain leverage specifically over minors. For example, threats around the victim going to jail because they shared intimate imagery would only be relevant for minor victims. Similarly, use of gift cards may be designed to create a way for minors to get funds without the same access to cash an adult might have.

A small group of reports in the current sample (represented by the "Content/other sextortion" line in Figure 2a) include demands for producing and sharing CSAM or returning to/entering into a romantic relationship. Such nonfinancial sextortion cases were the vast majority of cases in older surveys of victims of sextortion, fielded in years predating the current financial sextortion trend — surveys wherein only 14% reported demands for money.[17] In that study, the victims were 77% female and only 20% male; it's therefore important to emphasize that while financial sextortion is predominantly targeting males, other forms of sextortion exist which most commonly target girls. While such content-based sextortion is rare in the studied NCMEC data, it is also a crime with connections to stalking and intimate partner violence, very often committed by people who know the victim, and thus reports may be submitted to local authorities and be less likely to end up with NCMEC.

Victims Being Targeted – Location

We have two ways of measuring where the victims are located, shown in Figures 2b and 2c below. Some ESP reports analyzed by NCMEC have coding of "victim location," which provides a distribution over a range of countries, although the US and Canada top the list. However, this is a limited subset of the larger set of reports, and thus may not be representative of the larger distribution of victim locations.

Another method of gauging victim demographics is by detecting which language is being used by a victim when a platform reports their chat logs. While the most common language is English, we have a large number of French-speaking victims, followed by Tagalog, Spanish, and German. While the language spoken is only an approximate hint regarding a victim's actual location, this variety of languages spoken highlights how global the sextortion issue is.

36%
of all sextortion reports with chat logs used languages other than English.

15 We measured this over all analyzed sextortion cases studied by NCMEC (not just those in the sampling windows), totaling 3,600 cases.
16 C3P (2022). An Analysis of Financial Sextortion Victim Posts Published on r/Sextortion. https://content.c3p.ca/pdfs/C3P_AnalysisOfFinanSextortionPostsReddit_en.pdf
17 Wolak, Janis and David Finkelhor (2016) "Sextortion: Findings from a Survey of 1,631 Victims." https://www.thorn.org/wp-content/uploads/2016/08/Sextortion_Report.pdf

© 2024 Thorn

Fig 2b | **Country of victim**

Reports with victim location in ESP reports analyzed by NCMEC

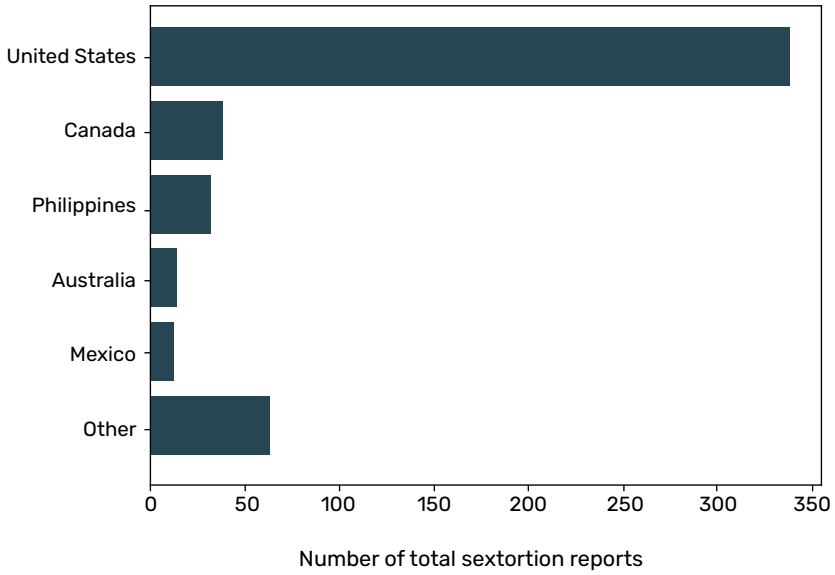
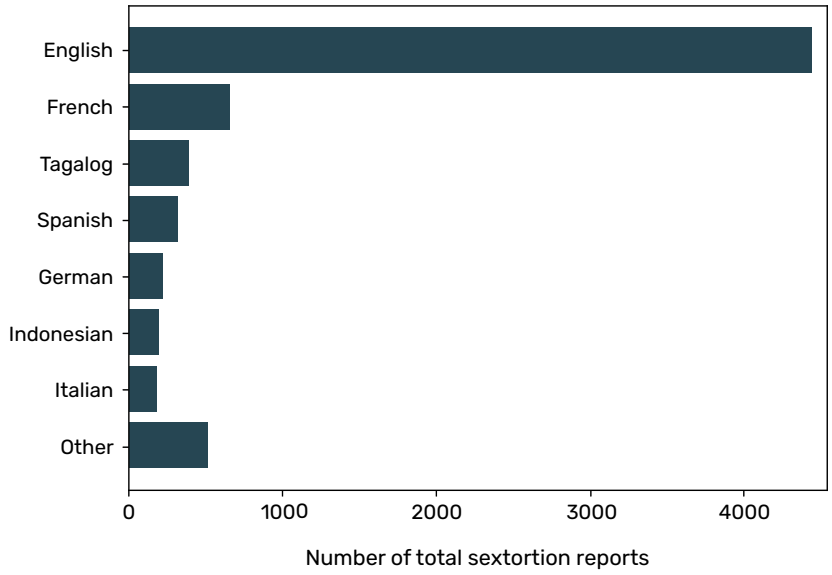


Fig 2c | **Languages in reports**

Language used in reports with chat logs



How Blackmail Imagery Is Acquired

Although children are most commonly sextorted through nudes that they sent to the extorter under false pretenses (often reciprocating after receiving sexual imagery they believe to be from the person they are talking to), additional tactics were apparent in the study. Event details provided in the sampled reports included descriptions of minors' accounts being hacked and their imagery taken without their permission, having images photoshopped, deepfaked, or otherwise being extorted with imagery they did not take, and being offered money or other incentives (modeling contracts, for example) to coerce the minor into sharing their nudes.

Actually measuring the distribution of different methods, however, is difficult. Sometimes, there are clear assertions (either in a victim's report or even in a chat log) describing how blackmail images or videos were acquired, such as a report claiming that they were hacked; however, such reports are rare. Of the studied reports, 69% of reports do not provide any information about how imagery was acquired and 25% are cases where perpetrators offered to exchange nudes with the child (or to go on a mutual video chat to show nudity). We expect that most of the cases involving offers to exchange imagery are "catfishing" – where the perpetrators are impersonating someone else (usually an attractive person of a similar age to the child) to make it easier to get those images or videos; however, currently available data is insufficient to confirm or refute this hypothesis.

© 2024 Thorn

Of the 31% of cases where specific methods are indicated for getting imagery from the child, catfishing is seen in the majority of reports. However, additional tactics, such as threats of creating fake/inauthentic intimate imagery or hacking, also appear. In 11% of reports, the child reports they did not send sexual imagery of themselves but were threatened with images that were in some way fake or inauthentic, such as the child's face being added to explicit images of an adult or another child. Although less prevalent than fake or inauthentic images, children also report they were either threatened with hacking or had intimate imagery stolen following an account hack. We studied other ways that perpetrators might coerce imagery from children (such as with offers of money) but found lower rates. Figure 2d shows the different ways that this blackmail imagery was acquired by perpetrators for the 31% of reports that included information about how the imagery was acquired (note that totals in 2d do not sum to 100 because reports may mention more than one method).

82%

of reports with information about imagery acquisition had signs of reciprocating images in response to the perpetrator sharing imagery.

Fig 2d | **How imagery was acquired**

Method used, out of reports that have acquisition information

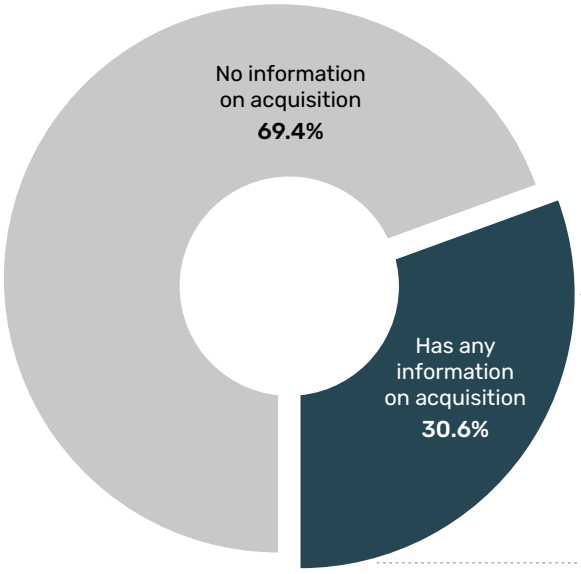
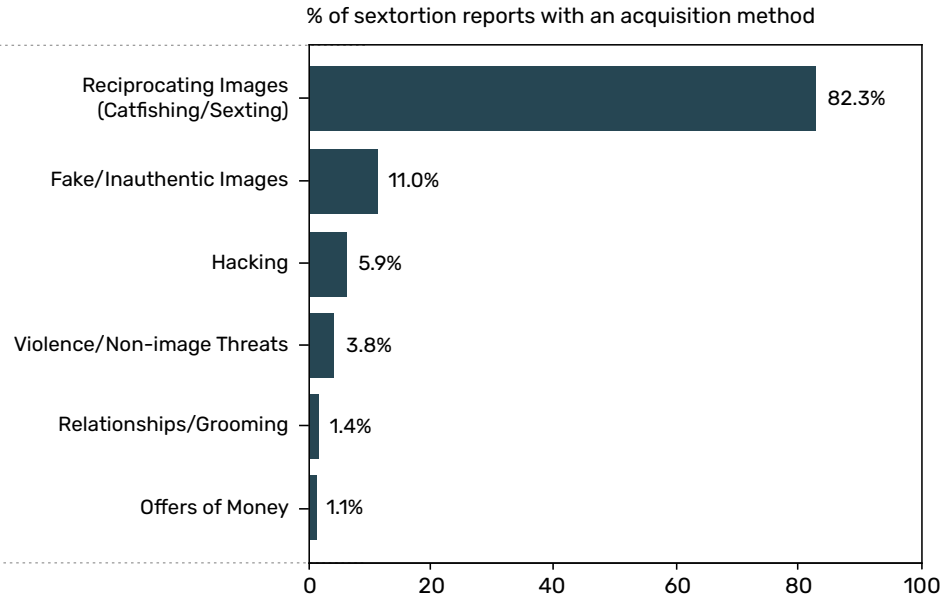


Fig 2e | **Acquisition information**

Rate of reports with information about how imagery was acquired



A report was counted in multiple categories if multiple tactics were used.

© 2024 Thorn

Hacking – What This Looks Like in the Context of Sextortion

Sometimes, children send sexual content within apps or platforms (or even simply save that content on their phone, or use an app image storage feature), and perpetrators can acquire this imagery by hacking. In many reports, the only information about this comes from a statement such as "they believe they were hacked."[18] While mentions of "hacking" appear mostly tied to ways that images or videos were acquired, it also appeared in reference to threats levied by perpetrators to hack a victim's account for the purpose of making it appear they are posting their own imagery directly.

Fake/Inauthentic Imagery in Practice

Threatening children with fake sexual imagery occurs in roughly 11% of the reports in which tactics were apparent. While this category includes photoshopping a child's face onto sexual content or even the use of AI-generated images or "deepfakes," the majority of these were cases in which a perpetrator threatened the child with imagery that was not of that child but was simply an image of someone else. In such cases, perpetrators with a nonsexual image that shows a child's face would threaten to send that image alongside a faceless nude photo that might be the same child, claiming that the child sent them both. While this does not yet quantitatively show increases in sextortion connected to AI-generated image production techniques, the Federal Bureau of Investigation (FBI) has warned of an increase in sextortion using generated imagery,[19] and the prevalence of simple methods for extorting children using fake images highlights just how vulnerable children may be to more serious "deepfakes" methods in sextortion contexts.

Other Coercion Methods

We also measured mentions of three other methods that can be used to get imagery from children: offers of money, threats of violence, and acquiring imagery through relationships (e.g., grooming, in-person relationships). Of these, "offers of money" are the most clear cut, although rare: situations in which children are either propositioned for imagery or in the context of a future monetary reward (e.g., posing as modeling agents), but in which the images or videos are immediately turned around to use in extortion. Other cases of violent or violence-adjacent threats are prototypical situations in which the offender threatens to hurt a child unless they provide imagery. In practice, we did not end up separating out threats of violence or threats of sharing their location for initial imagery vs. other threats of violence raised during sextortion, and so this may be even rarer than the current rate implies.

Finally, this studied sample included a small subset of grooming or relationship instances, such as where a child reports that they gave imagery to a boyfriend or girlfriend who then extorted them. Survivor studies[20] have reported this to be highly common in nonfinancial sextortion. Victims surveyed in 2015 reported that 59% knew their perpetrators offline, and of those that did only know their perpetrator online, 62% shared imagery because they were "in a wanted romantic or sexual relationship." In our sampled data, we only had 40 instances flagged as such, but only 11 involved financial sextortion, with 12 of the rest having "content/other" demands, asking the child to provide CSAM or to be in a relationship.

---

18 We take all accounts by a victim about how their images or videos were acquired at face value for the purpose of this report, with the awareness that there may be cases where victims who were coerced into sharing imagery may believe there will be less judgment stating that their imagery was hacked or deepfaked instead.
19 https://www.ic3.gov/Media/Y2023/PSA230605
20 Wolak, Janis and David Finkelhor (2016) "Sextortion: Findings from a Survey of 1,631 Victims." https://www.thorn.org/wp-content/uploads/2016/08/Sextortion_Report.pdf

© 2024 Thorn

Tactics of Pressuring Victims and Victim Impacts

Many of these financial sextortion threats seem designed to focus children on the risk of others seeing them in intimate imagery — whether actual or manipulated — and the potential for perceived life-ruining impact as a result. This threat is used to pressure children to pay before they have time to either process these threats or seek support. It is therefore important for victims to know that the best course of action is generally not to pay, to report the extortion to both NCMEC and the ESP, and to block the offender — but equally important to understand the work that offenders are doing to convince children that their lives will be ruined. Such threats can also become a roadblock to reporting or seeking help. This is particularly true if a child receives similar messaging from caregivers or others in their community.

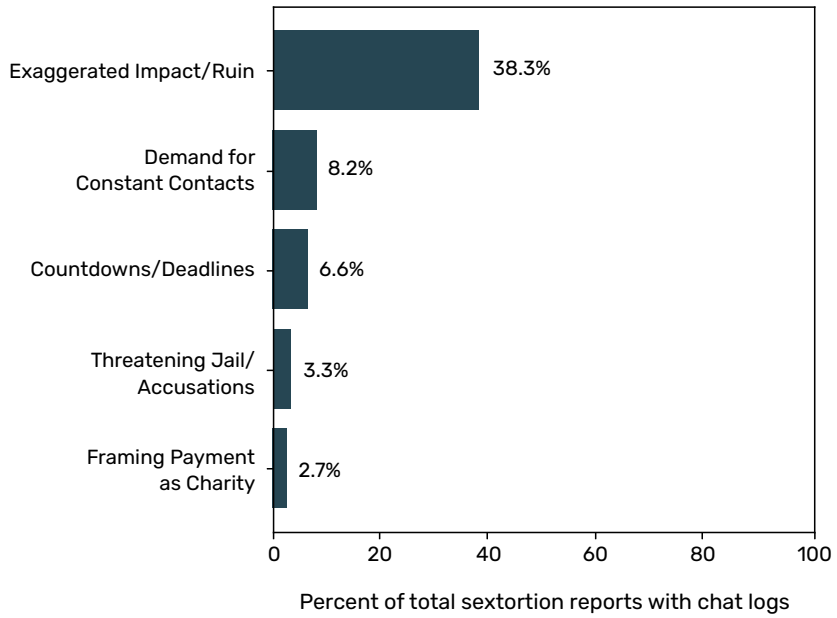
Established Information to Know: Sextortion can be a very stressful situation for children, leading to high stress or even self-harm or suicidal ideation. However, paying the perpetrators can simply lead to more demands.

Key New Findings: There is a range of methods perpetrators use to pressure children and increase the perceived severity of their imagery being exposed.

Figure 3a outlines the overall frequency of the various tactics perpetrators used to pressure victims out of all reports where conversation data could be measured. When these tactics show up in conversations, they are often very formulaic; perpetrators use extremely similar or even identical threats against different victims, as if operating off of a script designed to quickly and efficiently coerce victims to pay.[21] Shared across all cases, in addition to any of these distinct threat types, is the emotional stress associated with threats to have personal intimate imagery released to others.

Fig 3a | **Pressure threats**

Methods used to pressure victims, out of chats with pressure tactics



A report was counted in multiple categories if multiple tactics were used.

21 A new report by Network Contagion Research Institute (NCRI) shows some specific examples of such sextortion scripts. https://networkcontagion.us/wp-content/uploads/Yahoo-Boys_1.2.24.pdf

© 2024 Thorn

Exaggerating Impact

38%
of reports with chat logs included exaggerated impacts and/or threats of ruining the victim's life.

While threats to share content with a victim's family, friends, or followers are inherent to this issue, a large percent of these sextortion cases involved threats of much more dramatic impact — that the extorter would make the content go viral, would make the content get onto the news, would share it in a way that the child would never get a job, or more broadly that it would ruin the child's life. Many such threats are formulaic, repeated nearly identically from victim to victim. Table 2a shows examples of threats that perpetrators have repeated in this exact form four or more times over different reports:

Table 1a | **Example Phrases Used in Exaggerating Impact to Victims**
. . . blocking me won't stop me from posting it viral . . .
. . . or send this to ur school and u know u will be expelled from . . .
. . . you lose a lot of things - your honor - your dignity - your family life . . .
. . . I have ur nudes and everything needed to ruin your life . . .
. . . u will be exempt from universities if u don't cooperate . . .

Listed phrases appeared in at least 4 reports.

Legal Threats and Framing for Harassment/ Rape/Abuse

Some sextortion reports involve more specific threats relating to legal consequences for the minor. Threats can focus on the fact that the child sent nude imagery of themselves (since that imagery is CSAM, offenders threaten children with legal consequences for producing/sending it) or may focus on accusations that the child was abusing another child (as the perpetrator may have assumed a juvenile persona when initially soliciting the images or videos). Instances were observed in the data of perpetrators threatening to frame or publicly accuse the child of soliciting or abusing younger children or of generally being a pedophile (the youngest example of this in our data accused the child of sending content to a 10-year-old). Table 1b outlines example phrases that are commonly used in such threats; like other threats, these accusations are so formulaic that the exact phrasing can occur in many different reports:

Table 1b | **Examples of Accusation Threats Used**
. . . you do not ask my age before masturbate, i just want to say that i am a girl of 15 years . . .
. . . you know very well that this is an act pedophile and that is prohibited by law . . .
. . . you will be arrested by the interpol police . . .
. . . you will be locked up for 5 years in prison . . .

Listed phrases appeared in at least 4 reports.

Countdowns/Deadlines and Demands for Constant Communication

Perpetrators seem to employ a range of methods to attempt to make sure that their victims are required to make quick decisions, attempt to pay quickly, and do not have an opportunity to seek help from their caregivers or other sources of support.

We coded two ways in which perpetrators imposed such urgency: firstly, the use of countdowns and deadlines to make the victim rush to pay,

© 2024 Thorn

and secondly, the demand for constant communication and access. With countdowns and deadlines, perpetrators would give children fixed periods to encourage payment or to extract a promise of a method and amount of payment. For the second method in which perpetrators demanded constant communication from the child, perpetrators would threaten to expose a child's imagery if children simply disconnected from the video chat or did not respond in text chat quickly enough. Such threats might have a range of reasons, including both mental reasons to pressure the child or to prevent them from seeking help as well as enabling verbal instructions (which are more challenging to report).

"Donation to Charity" Tactics

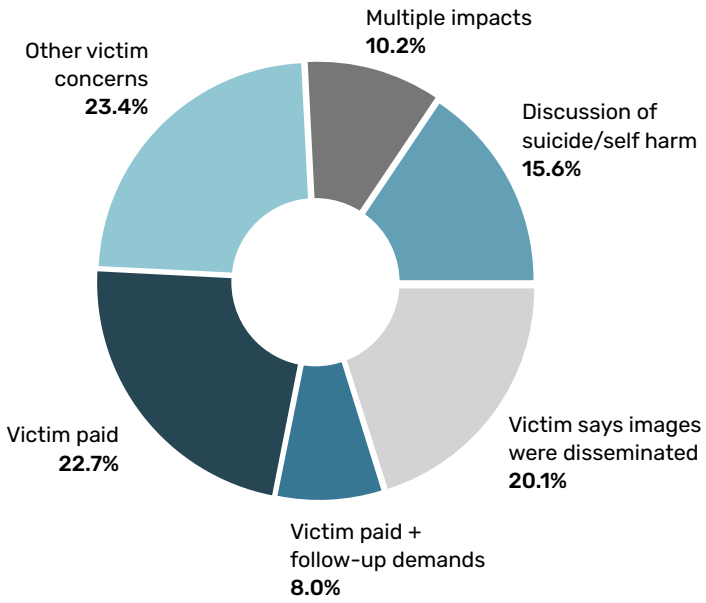
An additional tactic that perpetrators use is the framing of their sextortion as being done for the sake of humanitarian donations — usually a perpetrator claiming that the only reason they are demanding money is for a sick child, sibling, or parent. As with other tactics, this could involve a range of reasons, which might include financial details (e.g., they may want payment details to appear like a charitable donation) as well as to manipulate the children or make them more likely to pay or less likely to report the sextortion.

Victim Impacts – Mental Stress, Payment, and Dissemination

Reports also provide insight into the impacts of sextortion on the victims, which can range from discussion of whether their imagery was disseminated to mental/emotional impacts on the victim to discussion of giving into the perpetrator's demands for money. Although we only see discussion of impacts on the victim in a small set of reports (8.9% of cases), those reports can illuminate the different ways that sextortion might impact victims. Figure 3b outlines how often these different victim impacts appear in reports.

Fig 3b | **Victim impacts**

For the 8.9% of reports where any victim impacts were reported



Categories are exclusive: any instances where a report described a combination of multiple harm types is solely represented in the "multiple impacts" category.

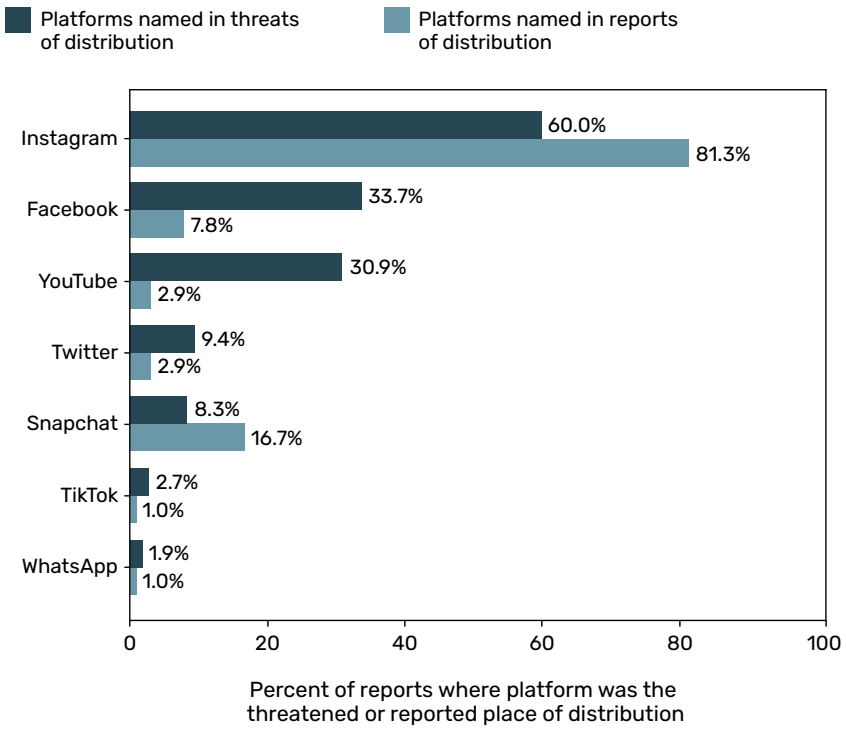
A significant category of sextortion impact is the actual sharing of the victim's imagery with friends, family, and the public. While this is a commonly mentioned outcome when victims talk about the impacts of sextortion, it remains in the minority and shouldn't be viewed as inevitable.[22] Perpetrators often claim that disseminated content will become public and potentially viral. However, in many reports where a victim states that their imagery was disseminated, the content was shared

22 The small percentage of victims stating that their imagery was shared is a lower bound: there may be more victims who did not yet know their imagery was shared or did not choose to disclose.

© 2024 Thorn

Fig 3c | **Threatened dissemination platform**

Platforms discussed as place where imagery would be disseminated



Shows platforms mentioned 30 or more times for distribution. Of 1837 reports with one or more platforms of threatened distribution, 102 confirmed distribution.

privately and with a limited number of people. This can be highlighted by looking at reports where particular platforms were coded as being the place where the victim's imagery would be shared: while it is common for Instagram, Facebook, and YouTube to be platforms where perpetrators threaten to post content (as illustrated in Figure 3c), only Instagram (and to a lesser extent, Snapchat) were commonly discussed when a victim also identified that their imagery had been shared ("Victim says images were disseminated" in Figure 3b).

81%
of reports stating that images had been shared listed Instagram as a location of that dissemination.

When mental and emotional impacts on victims do get reported, we split such content into two categories: discussions of suicidal ideation and/or self harm (mentioned in 17.5% of reports with victim impact) and a more general category of "other victim concerns" and mental stresses.[23] In both cases, only a small subset of cases include mention of a victim's situation in report texts — it can be assumed that many victims do not explicitly provide such information about the mental or emotional impact of this experience in the report.

Other sources of information, such as news articles regarding sextortion victims, can shed light on how these pressure tactics and the overall threat of image sharing can culminate in severe consequences such as suicide. In one such case, a news article notes[24] that perpetrators told the child that he "... would be labeled a pedophile. His parents wouldn't love him. He wouldn't be able to get into college or get a job. They would hurt or kill his parents." In a separate case, an article[25] quotes a parent as saying, "The information we collected shows that the pressure [the victim] was under was

27%
of victims who mentioned paying their perpetrator discussed ongoing demands.

23 Reports given the "mental stress" label included a range of experiences such as observations the victim was scared or worried to more severe discussions of panic attacks or concerns from victims about their safety if their parents were to find out. Reports with the tag "suicide/self-harm" were a subset of this group.
24 Warsmith, Stephanie. "'You might as well end it now': Terrorized by sextortion plot, a 17-year-old takes his life" USA Today, May 9, 2023. https://www.usatoday.com/in-depth/news/nation/2023/05/09/parents-spread-sextortion-warning-after-sons-death/70197048007/
25 "Starkville father goes on Fox News to warn parents about sextortion" WTVA, February 20, 2023. https://www.wtva.com/news/starkville-father-goes-on-fox-news-to-warn-parents-about-sextortion/article_cc645a-da-b14e-11ed-9aae-8f6fc94bb084.html.

© 2024 Thorn

unbearable to the point that during this exchange" the victim "... finally tells [the perpetrator], 'hey, I'm going to commit suicide, I'm going to kill myself,' and they respond with, 'go ahead, because you're already dead.'"

More than one in three (38%) reports with impact information mentioned making payments. However, these payments often did not deter continued harassment; 27% of victims who mentioned paying their perpetrator discussed ongoing demands experienced after a first payment (shown as the "paid + follow-up demands" slice in Figure 3b). This is a commonly noted trend, and some sources suggest it may be even more common than that; the report from Canadian Centre for Child Protection (C3P)[26] studying public discussions from sextortion victims on Reddit found that 93% of posts discussing payment included further demands for money after the initial payment.

Overall, we found that the median payment by a victim (gauged by the largest amount stated by the victim) was $100, and the median ask (gauged by the largest demand to a given victim) was $390. This highlights that perpetrators started with high demands but often accepted whatever amounts they could acquire from victims.

26 C3P (2022). An Analysis of Financial Sextortion Victim Posts Published on r/Sextortion. https://content.c3p.ca/pdfs/C3P_AnalysisOfFinanSextortionPostsReddit_en.pdf

© 2024 Thorn

The Role of Platforms in Sextortion

Sextortion does not happen in a vacuum; how children interact with platforms and specific design features can facilitate these sextortion events. We can see this by examining which platforms are used to initially contact children, and which were used as a secondary location that perpetrators would move the conversation to.

It is important to acknowledge this report focused on sextortion events reported by ESPs and the public to NCMEC between August 2020 and August 2023. Since this time, multiple platforms, government agencies, and NGOs have launched programs and new product features to combat the risk of sextortion confronting young people.[27] Additional research is needed to understand how these changes are impacting the likelihood of young people encountering sextortion and the outcomes for those who do experience such an event.

Established Information to Know: We add metrics to the most common use of platforms – children meeting a perpetrator on Instagram and moving to a platform such as Snapchat.

Key New Findings: We highlight the role of apps designed for random interaction with strangers – Omegle and Wizz – as places to initially meet children, and the use of other secure messaging apps like Google Chat and/or Hangouts and Telegram as secondary locations.

To avoid bias introduced by platform reporting behaviors (which platforms report and how do they report) we analyzed how platforms are used in sextortion events via explicit platform mentions in report text (such as in chat text or descriptions of the event provided by victims).[28] Any platform mentions were coded to capture how the platform was being discussed. Coding included labels such as whether perpetrators threatened to share imagery on that platform (discussed in Figure 3b), whether they explicitly discussed that the initial contact happened on that platform (e.g. "they added me on ____"), when there was explicit mention of moving to a secondary platform ("and the conversation moved to ____"). Contact Thorn for an appendix with labels and definitions.

Platforms Mentioned as Places Where Perpetrators Contacted Children

Of the 3,276 reports that discussed one of the platform uses discussed in this report, 576 were coded with an "initial contact" label; a few core platforms dominate the studied reports as places used for that initial point of contact, shown in Figure 4a.

10%
of reports with information about the platform of first meeting mentioned Omegle or Wizz as an initial contact point.

Instagram, Snapchat, and Facebook were the most common platforms mentioned — and they were also frequently cited as initial contact points. For many of these

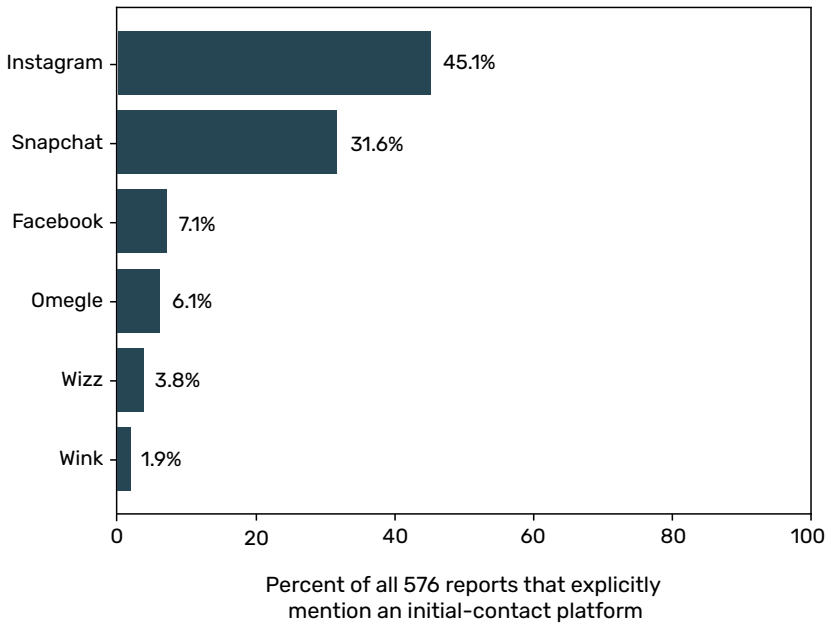
27 Examples of such changes (but not exhaustive) include restrictions on messaging with youth accounts (https://about.fb.com/news/2024/01/introducing-stricter-message-settings-for-teens-on-instagram-and-facebook/), enhanced tools flagging risky interactions (https://support.discord.com/hc/en-us/articles/18210977897239-Discord-Safety-Alerts#:~:text=Safety%20alerts%20on%20senders,-1.&text=If%20detected%2C%20Discord%20will%20notify,safety%20tips%20to%20safeguard%20themselves), cross-industry information sharing (https://www.technologycoalition.org/newsroom/announcing-lantern), and multi-sector awareness campaigns (https://values.snap.com/news/k2p-launch).

28 As described in the limitations section, while we tracked discussion of many apps and platforms, some methods of communication that are often discussed more generically (such as discussing texting without named apps, phone calls or email) — were kept out of the scope of the report due to the difficulty in discerning which tool was used. However, that should not mean that such tools may not be used for sextortion.

© 2024 Thorn

Fig 4a | **Platforms used for initial contact**

Platforms mentioned ten or more times as an initial meeting platform



A report was counted in multiple categories if multiple platforms were discussed.

mentions, a report simply stated that the child was "first contacted" on the platform or that they met on the platform. However, chat logs sometimes provide the initial conversation, which can often involve statements from perpetrators such as, "You don't know me, but your profile was recommended to me by [platform]"; such discovery systems may help perpetrators justify random connections to children. Two smaller platforms for randomly meeting strangers, Omegle[29] and Wizz, were also both explicitly mentioned as the "initial contact" point more than ten times, highlighting the role that such systems might have in enabling perpetrators to get new connections to children (while Omegle shut down in the second half of 2023, many similar competitors exist). Wizz, in turn, has been highlighted[30] by NCMEC and C3P as a common platform mentioned by the public regarding sextortion, and a recent Network Contagion Research Institute (NCRI) report[31] has found instructional videos for perpetrating sextortion over Wizz.

Platforms Mentioned as Secondary Contact

Thorn surveys of children have found that 65% of children had experienced someone attempting to get them to "move from a public chat into a private conversation on a different platform,"[32] and this is a common event in sextortion situations, with perpetrators moving children to platforms that are less likely to detect the event and/or where the child may be more likely to share content.

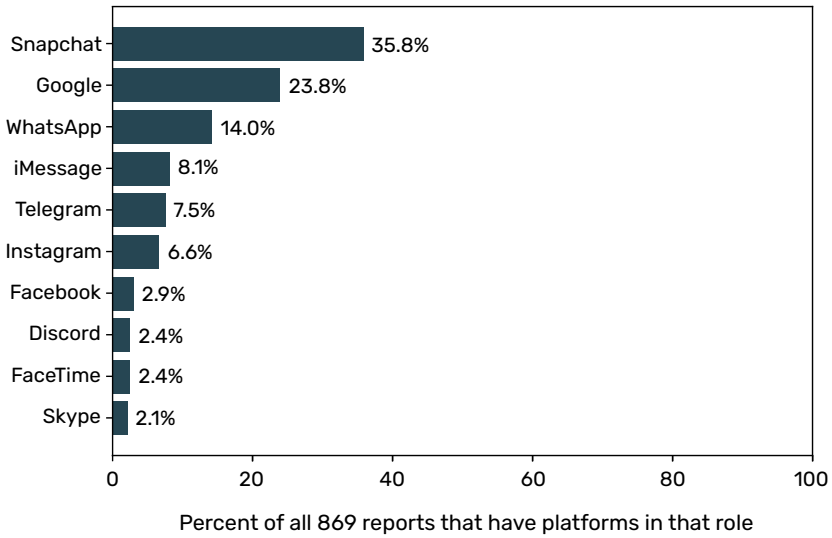
When we look at platforms that were coded as being used as a "secondary location," where a report identified that the victim was moved from one platform to another (869 reports mention such a secondary location), we found that the most common platform to which these interactions are moved is Snapchat, as shown in Figure 4b. Perhaps due to features such as disappearing Direct Messages (DMs), Snapchat is a common platform for sending self-generated CSAM (SG-CSAM or nude selfies); recent surveys of youth found that 39% of 13- to 17-year-olds who had shared their own nudes did so via DM "in apps where content disappears, like Snapchat."[33]

29 Most of the reports analyzed in this study predate the shutdown of Omegle in November 2023.
30 https://www.nbcnews.com/tech/social-media/friend-finding-app-offered-safe-space-teens-sextortion-soon-followed-rcna91172
31 Raffile et al. (2024) A Digital Pandemic: Uncovering the Role of 'Yahoo Boys' in the Surge of Social Media-Enabled Financial Sextortion Targeting Minors. https://networkcontagion.us/wp-content/uploads/Yahoo-Boys_1.2.24.pdf
32 Thorn (2022). Online Grooming: Examining Risky Encounters Amid Everyday Digital Socialization. https://info.thorn.org/hubfs/Research/2022_Online_Grooming_Report.pdf
33 Thorn (2023). LGBTQ+ Youth Perspectives: How LGBTQ+ Youth are Navigating Exploration and Risks of Sexual Exploitation Online. https://info.thorn.org/hubfs/Research/Thorn_LGBTQ+YouthPerspectives_June2023_FNL.pdf

© 2024 Thorn

Fig 4b | **Platforms used as secondary destinations**

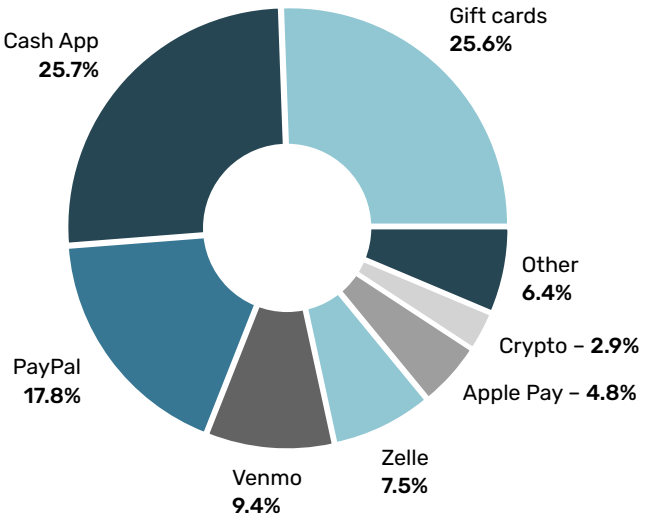
Platform mentioned as destination where conversation is moved to



Platforms included if mentioned 15 or more times in this role. A report was counted in multiple categories if multiple platforms were discussed. Google encompasses google messaging services but not Youtube.

Fig 4c | **Platforms used for payment**

Payment platforms mentioned in reports



A report was counted in multiple categories if multiple payment methods were discussed.

However, Snapchat is not the only platform used: Google messaging products (primarily Google Chat) are also commonly used, followed by WhatsApp, iMessage, Telegram and Instagram[34]. Discussions of such secondary apps generally do not provide explicit reasoning for why such a switch is done or why a specific platform is selected, but the requests are often connected to offers to exchange nudes — there are many statements along the lines of "download [platform] so that we can trade nudes / chat naked." While many of these tools share the ability to have end-to-end encrypted text messaging (and thus may be less likely to detect sextortion[35]), another notable feature of platforms such as Google messaging apps and Telegram is that they can be operated using a desktop version. This capability may be preferred by perpetrators attempting to prove their persona by "spoofing" webcams (making a saved video appear as if it is a live video chat).

Payment Platforms and Systems

Chats frequently involve mention of payment methods or platforms. We measured not only named payment platforms but also a few more general payment approaches, such as using gift cards or cryptocurrencies. The

34 Note that since the majority of sextortion reports with clear chat logs submitted to NCMEC were submitted by Instagram (discussed further in the report), it is possible that there are specific biases due to being the place of discussion: for example, two people chatting on Instagram are unlikely to suggest moving the conversation to Instagram.

35 We should acknowledge that such platform shifts, if they are successful in avoiding detection or reporting by a platform, would not show up in our platform data. They may, therefore, indicate particular blind spots in our understanding of the sextortion landscape.

© 2024 Thorn

two most common methods of payment discussed were Cash App and gift cards (which could be referred to as general gift cards or could involve specific online systems such as Steam or iTunes). These were followed by other easy-use payment apps such as PayPal and Venmo. Figure 4c shows the relative amount of mentions of each platform method: one can see that the dominant methods are gift cards and Cash App.

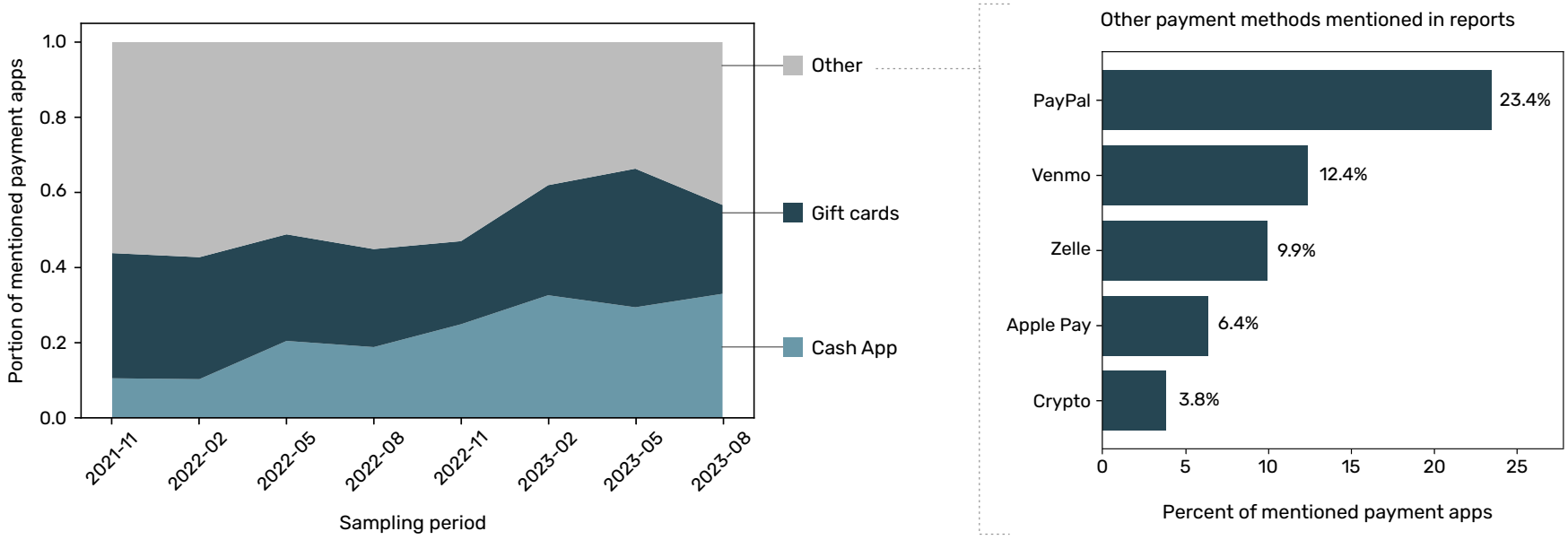
The dominant methods of payment are **gift cards and Cash App.**

These findings broadly agree with payment platform trends noted in the C3P study of disclosures on a sextortion subreddit, with some notable differences. That study — which unlike this work, included adults — found largely similar rates overall, but slightly less use of gift cards and Cash App, and higher rates of use for wire transfer/remittance systems (e.g., Western Union) and cryptocurrency. While the data from the two sources cannot be cleanly compared, it would be unsurprising to find higher rates of simple and easy-to-access payment systems in the NCMEC reports given the focus on child victims. This also suggests that while minors are likely not being targeted for sexual purposes in financial sextortion cases, tactics are being deployed specific to minor targets.

The dominance of gift cards and Cash App has slightly increased over time. If we map the use of gift cards, Cash App, and other payment platforms over time (Figure 4d), one can see the slow increase of Cash App and gift cards relative to all the other payment apps.

Fig 4d | **Payment methods over time**

Trends in largest two payment methods over time



A report was counted in multiple categories if multiple payment methods were discussed.

Platforms included if mentioned 10 or more times in this role.

© 2024 Thorn

Perpetrator Differences by Country

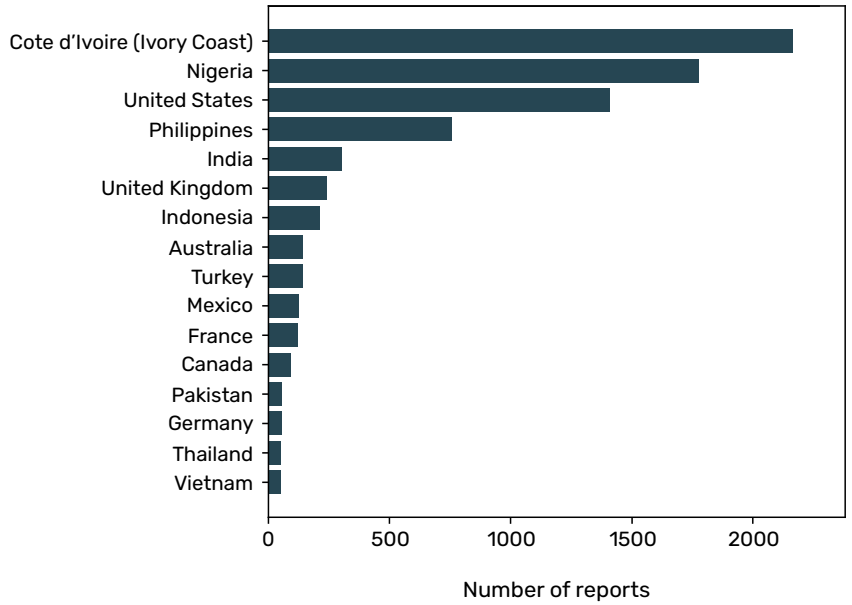
Financial sextortion appears to often be an organized endeavor, where many cases use nearly identical language (as if using the same scripts), have been reported to use the same profile pictures, and are largely resolving to a few international locations (based on information provided in ESP reports to NCMEC). The two countries linked to sextortion most often in this report are Nigeria and Cote d'Ivoire (and to a lesser extent, the United States).

Established Information to Know: A large percent of these events are from individuals based in Nigeria and Cote d'Ivoire.

Key New Findings: We highlight that perpetrators across different countries seem to differ across tactics used to influence children, which messaging apps they move the conversation to, and the victims themselves.

Fig 5a | **Reports linked to a country**

For countries linked to twenty or more reports



Country-Level Trends

Many financial sextortion cases are emanating from what seems to be organized criminal groups in two countries, Nigeria and Cote d'Ivoire,[36] and reports submitted to NCMEC are often linkable to those countries. These two locations have been postulated as the locations of organized sextortion groups.[37] Figure 5a shows the overall breakdown of all countries with more than 40 reports; for the largest sources of sextortion, such as Nigeria and Cote d'Ivoire, we have enough data to also dig into differences in how the tactics and platforms used (as discussed in prior sections) vary from country to country. Note that we discuss a report being linked to a particular country when the report lists that country in the NCMEC "international country" field. While this field often refers to the location of the perpetrator, it is not guaranteed to do so; if the only location

47%
of all reports linked to a country were linked to either Nigeria or Cote d'Ivoire.

36 https://www.justice.gov/usao-wdwa/pr/fbi-and-partners-issue-national-public-safety-alert-financial-sextortion-schemes
37 The recent NCRI sextortion report covers sextortion manuals and scripts within Nigerian perpetrators in particular – Raffile et al. (2024) A Digital Pandemic: Uncovering the Role of 'Yahoo Boys' in the Surge of Social Media-Enabled Financial Sextortion Targeting Minors. https://networkcontagion.us/wp-content/uploads/Yahoo-Boys_1.2.24.pdf

© 2024 Thorn

known is the country associated with the victim, that child's location would be provided in that field instead.

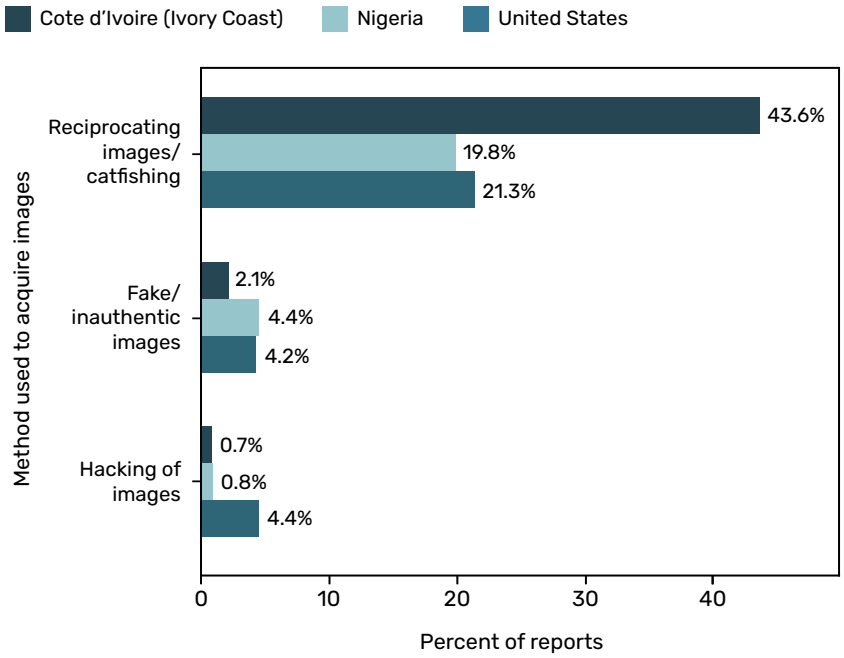
Differences in Methods by Country

We see differences in both how offenders attempt to get blackmail material of children, as well as the pressure tactics that they use once they have that material. Figure 5b outlines the labels with the most variation: from the data, it appears that accounts in the US are more likely to make use of methods like hacking to get data, whereas data from the Ivory Coast is more likely to have explicit discussion of exchanging imagery, although since a large amount of all of these cases likely involve catfishing and exchanging images, those cases may simply be more explicit about it.

We also found some trends in different tactics used to pressure children (the tactics discussed in the earlier section on victim impact). Reports linked to Cote d'Ivoire seemed more likely to pressure children using threats to get them sent to jail (e.g., scaring them with the illegality of their own nudes) and more likely to insist on constant contact (such as not disconnecting from a chat). Other tactics were more comparable — reports

Fig 5b | **Acquisition tactics by country**

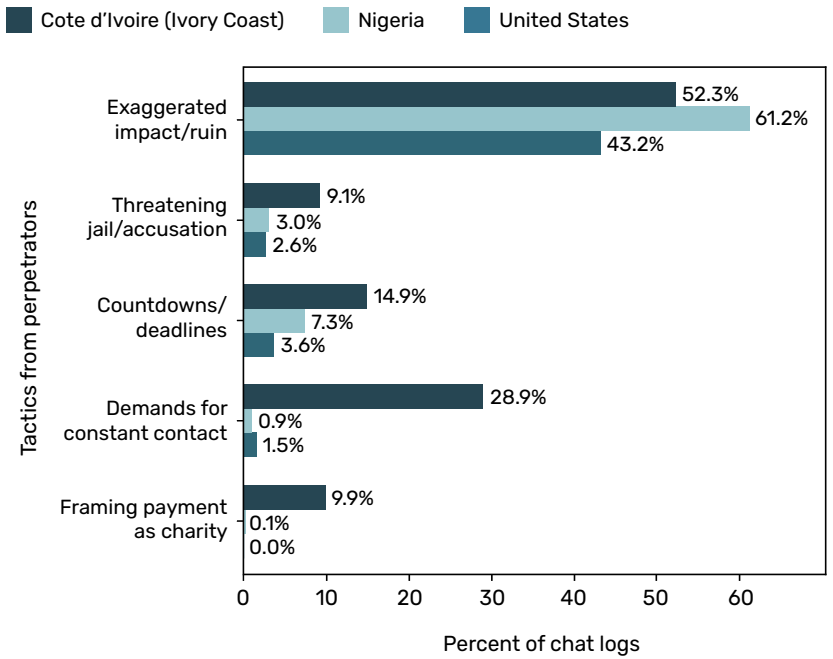
For the most common countries, as a percentage of reports to each



A report was counted in multiple categories if multiple tactics were used.

Fig 5c | **Threat tactics by country**

For the most common countries, as a percentage of reports to each



A report was counted in multiple categories if multiple tactics were used.

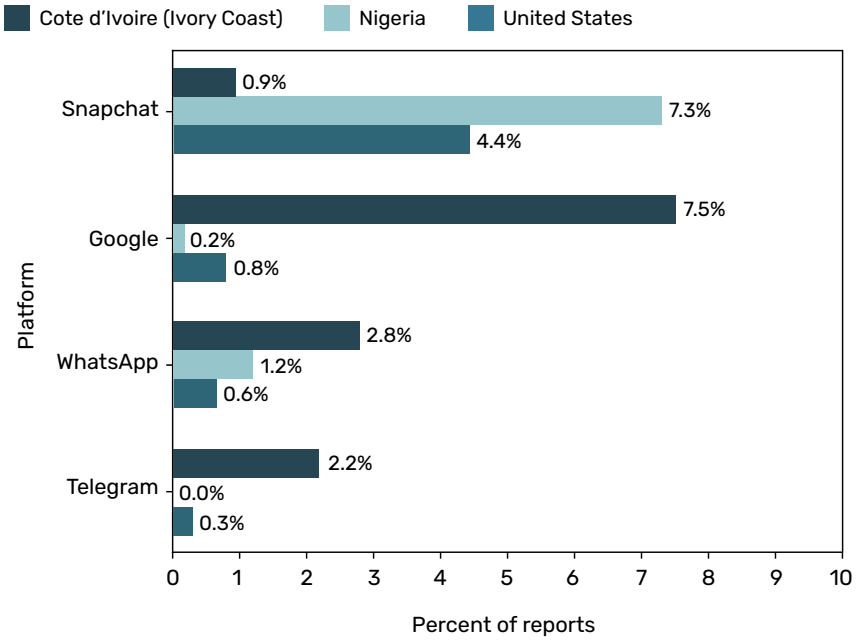
© 2024 Thorn

from both countries made use of countdowns/deadlines and extensive use of more general exaggerations of impact (such as threats to make a child's imagery go viral).

We also see differences in the platforms themselves being used, in particular, for the messaging apps where children are moved to after initial contact. Sextortion from Nigeria relied almost entirely (in the time period studied in this report) upon Snapchat, whereas data from Cote d'Ivoire also uses other messaging apps such as Google Chat/Google Hangouts, WhatsApp, and Telegram.

Fig 5d | **Secondary platform use by country**

For the most common countries, as a percentage of reports to each



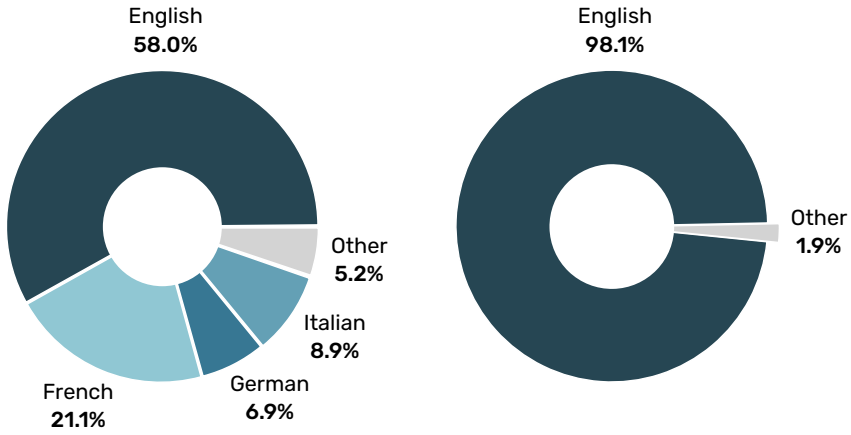
A report was counted in multiple categories if multiple platforms were discussed. Google encompasses messaging services but not Youtube.

Differences in Victim Languages Spoken

It is not surprising that there is more sextortion in French from perpetrators in the Cote d'Ivoire (which has French as an official language) and more use of English in Nigeria for the same reason. However, the distribution of languages used in chats shows the Cote d'Ivoire sextortion data is very multilingual, having cases in English, Italian, German, and Spanish, and even a long tail of other languages such as Russian and Polish. Chats in all such languages can often involve nearly identical formulaic language to what shows up in French sextortion cases, including some overly literal translations from French (such as threats with pourrir la vie, to ruin someone's life, but literally "to rot their life"). This may suggest the use of automatic translation tools to target a wider range of victims. We can see these differences in Figure 5e, which highlights the differences in language used.

Fig 5e | **Language by country**

Languages used in chat logs linked to Nigeria and Cote d'Ivoire



Language of chats linked to Cote d'Ivoire (Ivory Coast)

Language of chats linked to Nigeria

© 2024 Thorn

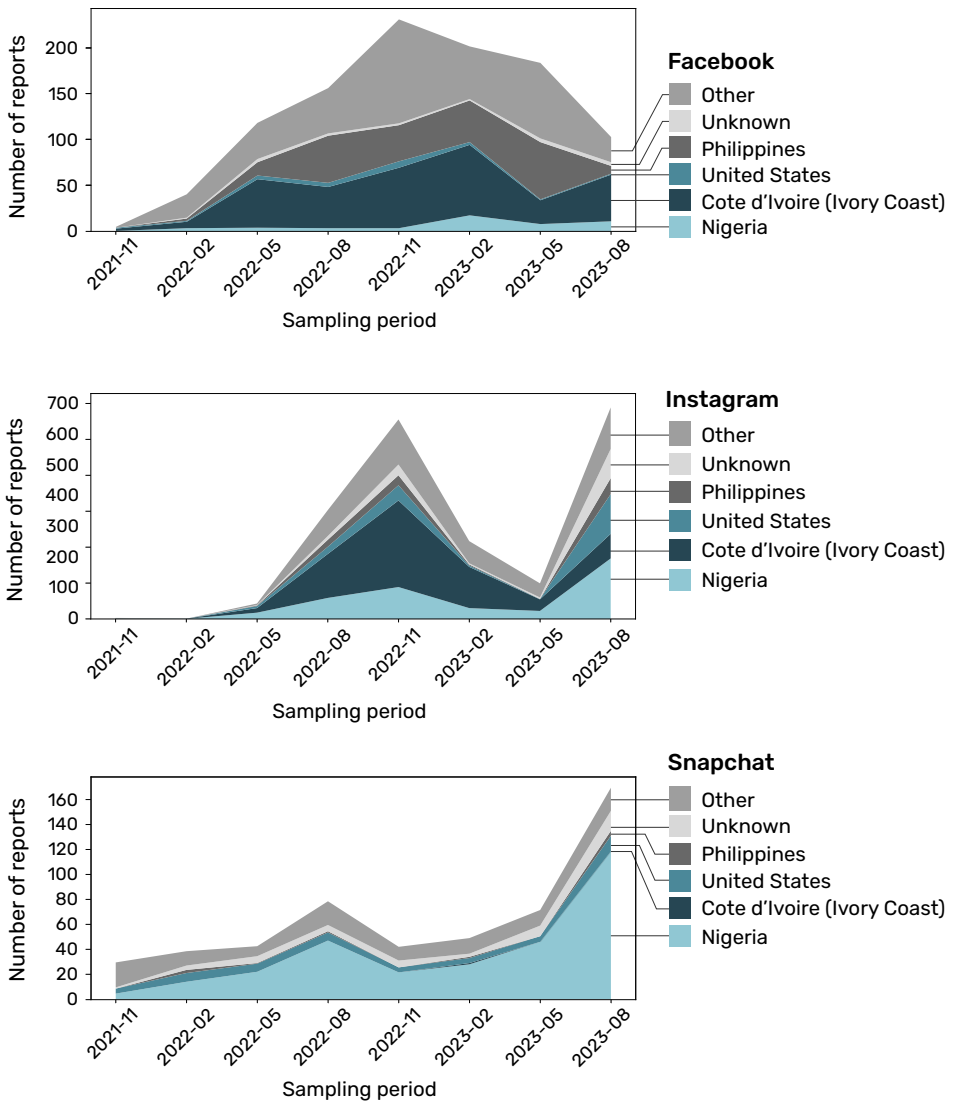
For most reports, we can speculate on victim location based on the language spoken in chats but cannot derive confident trends. Perpetrator groups based in Cote d'Ivoire do target many victims speaking languages that are common in Europe, such as French, Italian, and German, but we cannot be certain about that link; e.g., for French, many such victims may be in countries such as Canada.

Trends Over Time by Perpetrator and Country Differences

When we look at how reports submitted by platforms (ESPs) link to countries, a trend emerges in which sextortion reports submitted by Instagram tend to be linked to Cote d'Ivoire, and sextortion reports submitted by Snapchat are often linked to Nigeria (and almost never linked to Cote d'Ivoire). This reflects the trends seen above in the explicit mentions of platforms in text, which also found that reports connected to Nigeria are more likely to attempt to move victims to Snapchat. However, one can see in Figure 5f that this trend shifts in the last periods of analysis, with many reports linked to Nigeria not only in Snapchat but also Instagram. Although very preliminary checks against the latest November 2023 data do suggest that this shift may be temporary,[38] it is not clear why such shifts occurred, nor whether they are due to encouraging developments (e.g., increasing success investigating these crimes in a particular area) or other developments such as shifts in moderation systems.

Fig 5f | **Country differences by reporting platform**

Countries linked to reports submitted by top platforms



38 We do not have manual annotations of November 2023 data, and therefore can only measure reports already coded by platforms. Of reports escalated by Instagram that explicitly state "the offending account is sextorting the apparent minor", 69% were linked to Cote D'Ivoire, and only 4% linked to Nigeria.

© 2024 Thorn

Platform Reporting Landscape

Reporting activity across ESPs relating to sextortion is widely varied. ESP reports of sextortion are currently the overwhelmingly leading signal into NCMEC that a child is being sextorted, making up 85% of the total reports of sextortion during the sampled timeframe.[39] Three platforms stand out as the driving force behind these numbers: Facebook, Instagram, and Snapchat.

Importantly, the number of reports made to NCMEC should not be taken to directly equal the number of sextortion instances occurring on those platforms.[40] In fact, in some cases, increased reporting is also reflective of increased commitment to detecting and reporting such instances. In addition, it's important to acknowledge that not all sextortion cases are reported — either because they are not detected and reported by the ESP or a victim is not prepared to disclose.

Increases are not always bad
since they can be a result of better detection and reporting.

To build a more complete understanding of how sextortion is appearing on individual platforms, this study explores both the patterns in reports made by individual ESPs (such as overall volume and time between when an event occurs and when the report is made) and patterns in reports made by the public, thereby broadening our understanding of where sextortion may be occurring beyond merely if ESPs are reporting.

Established Information to Know: Reports from Instagram constitute a huge percentage of all ESP reports coming in where sextortion is reported, followed by Facebook and Snapchat.

Key New Findings: We highlight large peaks in the data from Instagram with a drop in the first half of 2023 (although we highlight that such changes may be internal rather than reflecting big drops in sextortion), and we highlight that there are platforms which might be expected to report more than they do.

Most Reports Are Submitted by a Few Companies

There are a number of notable changes in the trends over time for reporting platforms, shown in Figure 6a (these trends are purely about the reporting platform itself, and thus are not necessarily proportional to where the sextortion took place). The first is a sharp increase of cases submitted by ESPs starting in the middle of 2022; that increase could reflect the actual increase in sextortion at that time but might also reflect work that NCMEC did in raising alarms about financial sextortion to the platforms in June of 2022.

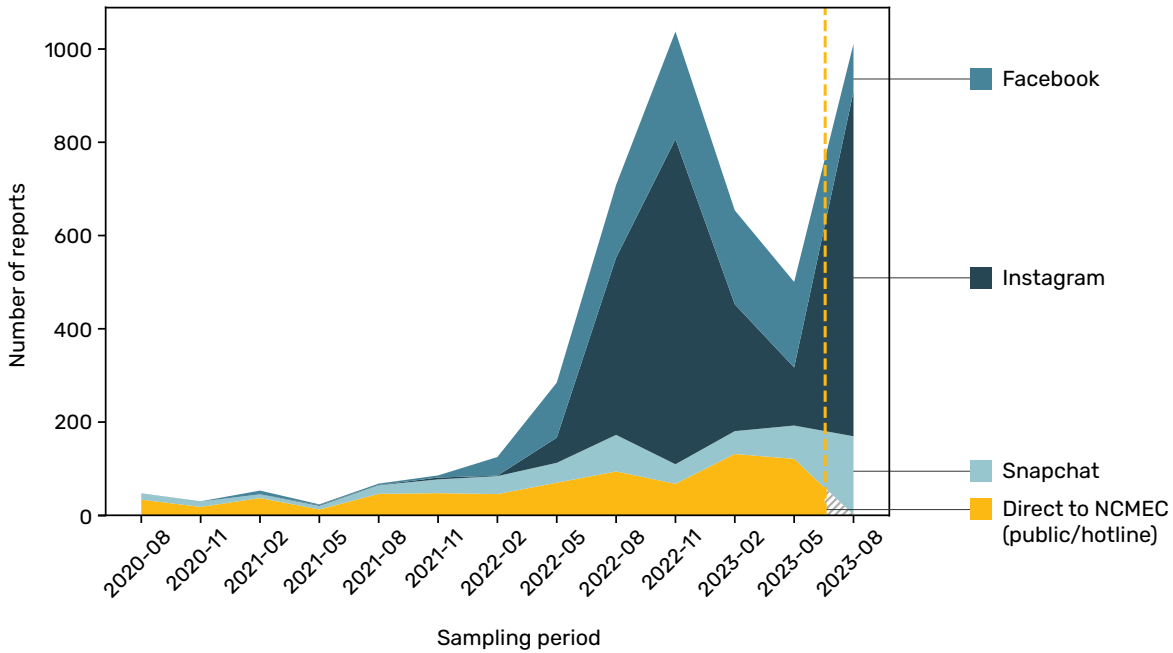
A second notable trend is the variability in the number of reports from platforms. We see two high periods from Instagram having around 700 instances per week but much lower rates in early 2023, as well as an increase in reports from Snapchat in the last period. As Instagram reports

39 Since ESP reports make up 99% of all reports to NCMEC when looking at all report types (the majority being reports of CSAM), this 85% measure is actually reflective of a relatively high number of public reports for sextortion, underscoring the critical importance of public reporting.
40 There are cases where Meta's reports are submitted by a single platform, but may cover conduct which may have occurred on multiple platforms (e.g., a Facebook report may include conduct which may have been committed on WhatsApp). In these instances, the data in this study only reflects the specific platform that made the report.

© 2024 Thorn

Fig 6a | **ESP reporting trends**

Number of reports per week submitted by an ESP or to NCMEC public form. Dashed line indicates last point of data collection of public data.



Platforms shown if they submitted 5 or more sextortion reports per week.

constitute the majority of all reported sextortion, those changes in Instagram reports can dwarf all other trends in order to define the overall trends in sextortion. A third trend, the drop-off in public reports in August of 2023, is simply missing data and can be ignored: those reports had not passed through the NCMEC analysis pipeline by the time of analysis.

57%
of all ESP reports in our data are through Instagram.

Reporting Delays and Report Informativeness

Looking at the amount of time between when these sextortion events happened and when they were reported to NCMEC by an ESP can give insight into these overall trends, pointing to differences between the overall rates of sextortion instances and the time it takes for platforms to report these cases to NCMEC. Of note, several things may influence the time between event and report to NCMEC by an ESP, not all of which are in control of the ESP submitting the report. Among them: existence and efficacy of proactive detection practices, changes in offender tactics, dependence on user reporting, and efficiency of content moderation pipelines.

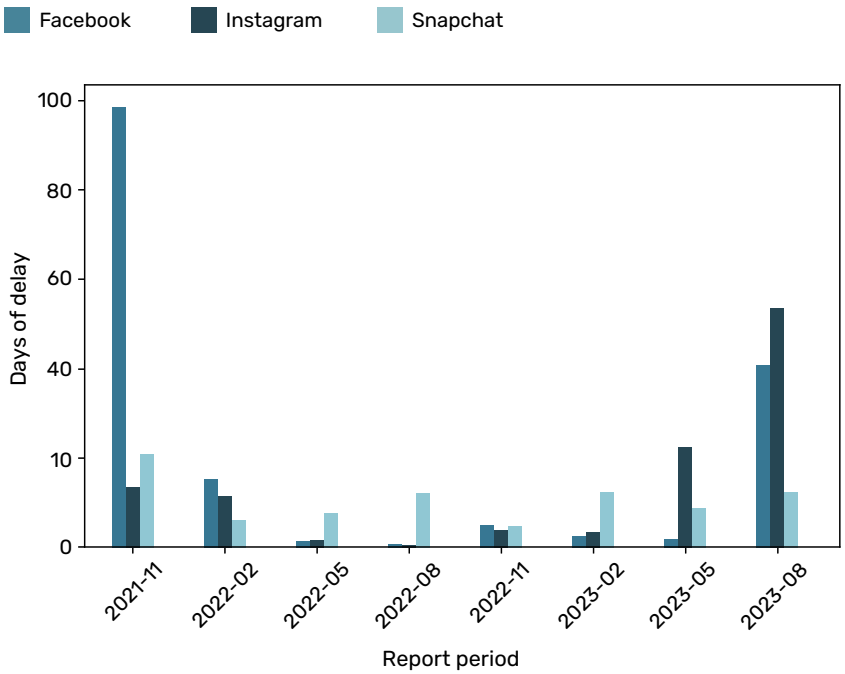
Figure 6b presents the median delay between sextortion events and their reports for the main three sextortion-reporting platforms over the last two years, showing that both Facebook and Instagram reports shifted from a pattern of rapidly submitting reports within days of an incident to more recently having a median delay of over a month. In the second half of 2022 (where Facebook and Instagram show the greatest amounts

© 2024 Thorn

of report activity), the time between event and report is the shortest (typically a matter of days). In comparison, in the sample studied in August 2023, the time between event and reports is considerably longer (over a month in some cases), suggesting this cohort of reports may be delayed reporting corresponding to the overall dip in report volume observed in the May 2023 timeframe.[41] Put another way, the data suggests the dip in reports by Instagram and Facebook in May 2023 is less a reflection of a drop in sextortion activity on these platforms and more likely a reflection of delays in the content moderation pipeline submitting these reports to NCMEC. Similarly, the data suggests the spike in report activity in late 2022 is not a result of a reporting backlog, as the time between event and report is relatively short, but may point to improved detection and/or increased sextortion activity.

Fig 6b | **Reporting delay**

Median number of days between incident and submission of report to NCMEC



Although increasing time lags in reporting are concerning, it is important to recognize the inherent difficulties in detecting and responding to sextortion cases. We do not know if these delays are primarily due to changes in content moderation, advances in detection technology, or shifts in user or perpetrator behavior. Furthermore, we should acknowledge the positive impact of periods where platforms were swiftly responding to sextortion incidents, and focus on how platforms can be encouraged to maintain such responsive reporting to NCMEC.

7 days
was the average period between incident and an ESP report to NCMEC in the last two years of data.

Beyond the speed with which a report is submitted, the level of information included in a report can determine its impact. More information — and child victim information, in particular — can be vital in deploying local emergency response services to safeguard that child in danger. When an ESP observes signals suggesting immediate risk to a minor, they may utilize an "ESP escalation" field in which a report can be flagged to NCMEC for more urgent study, with a summary characterizing the event such as "This account is sextorting an apparent minor." Apart from that escalation field, some sextortion reports include chat excerpts or additional context, which shed light on the platforms and tactics used or the situation's urgency. While 63% of ESP sextortion reports in this study had such an "ESP escalation" field (often identifying the case as a sextortion incident), that number varied across platforms, from 73% of sextortion reports being escalated in Instagram data but others providing little or no escalations.

41 Tentative data collected after the main analysis found that the time lag in November 2023 was much lower than this peak in August 2023, supporting that this may be a temporary issue.

© 2024 Thorn

Similarly, detailed report information, such as chat logs, is valuable for understanding and prioritizing these cases. While the data historically includes many reports from Snapchat without chat log data, we observe recent improvements to their reports, which now include small chat excerpts.

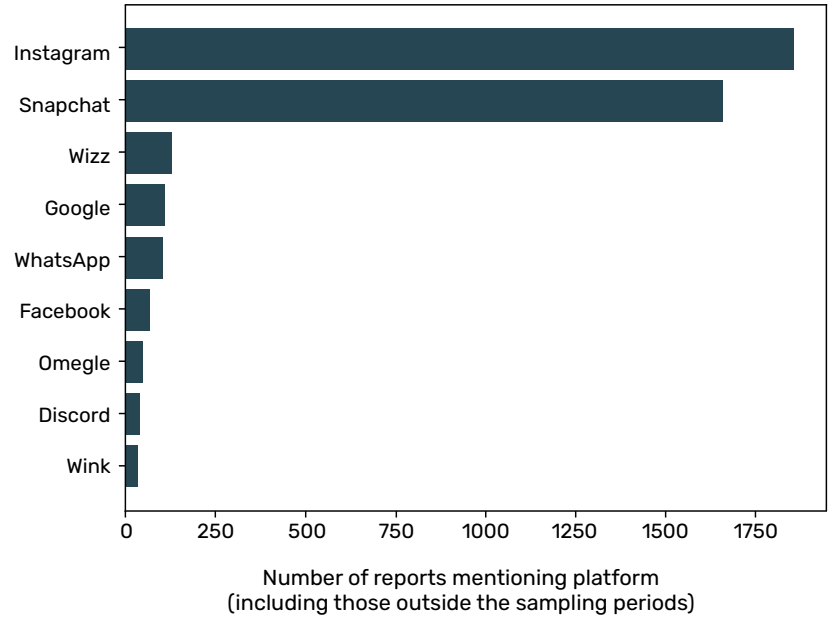
There are Many Points Where We Would Expect to See More Reports Than We Do

However, one should not give scrutiny only to the platforms that report the most, but also to platforms that report less than expected. Examining the number of times specific platforms are mentioned in public reports as compared to volume of reports made to NCMEC by the individual ESPs offers one tentative way to estimate how many reports one might expect. Figure 6c shows a distribution over how often platforms are mentioned in sextortion cases submitted to NCMEC by the public (via public form or hotline) over the last three years, as logged by NCMEC analysts[42] — it shows that Snapchat is mentioned almost as often as Instagram, and that there are a range of platforms that are mentioned more than 30 times in any sextortion reports submitted through NCMEC public form or hotline.

In this data, we see gaps between how many reports of sextortion an ESP submits to NCMEC, compared to how often that ESP is mentioned in public reports. For example, while Snapchat was mentioned nearly as often as Instagram and far more than Facebook in public reports, report volume directly from the platform is almost half that of Facebook and a quarter as many as Instagram, although that is dramatically improved in the latest (August 2023) data sample with upticks in Snapchat reporting. Similarly, Discord was mentioned roughly as often as Omegle and less than half as often as WhatsApp or Wizz, but Discord submitted far more reports of sextortion to NCMEC in the period

Fig 6c | **Platform mention in public reports**

Platforms mentioned 30 or more times in sextortion cases



Mentions of Google encompassed messaging services but not Youtube.

sampled.[43] Table 2 below provides those numbers, ordered by the ratio of the number of reports submitted each week per public mention — a rough way of approximating whether they are submitting as many sextortion reports as one might expect.

Importantly, lower rates than anticipated by comparing public and ESP reports could have several reasons. Some platforms may be better at proactive detection, so that they can find cases even if the victim did not report. Some parts of the sextortion experience may be more likely

42 NCMEC analysts coded all public reports that they analyzed (not just those within the time samples we focused on) in terms of the platforms used, and so we use that data here in order to have a larger sample size for this discussion.

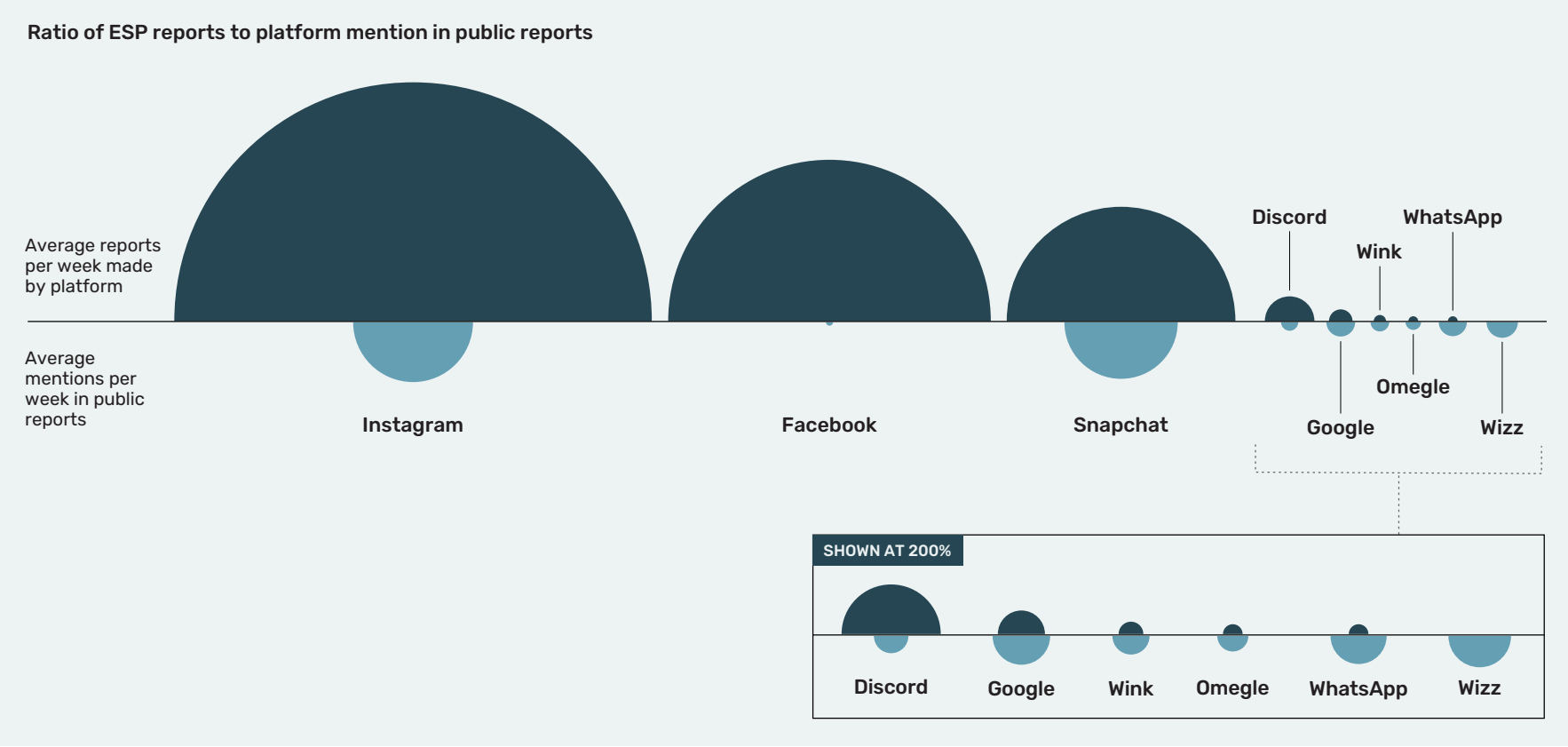
43 We should note that some companies are not based in the US, such as the France-based Wizz app, and thus may report to other agencies.

© 2024 Thorn

Table 2 | **Comparison of reporting rates to mentions**

Platform mentions in all public sextortion reports, compared to ESP report

	Instagram	Facebook	Snapchat	Discord	Google	Wink	Omegle	WhatsApp	Wizz
Average reports per week made by platform	283.94	129.62	65.12	3.06	0.69	0.19	0.12	0.12	0
Average mentions per week in public reports	17.88	0.63	15.99	0.37	1.03	0.43	0.3	0.98	1.22
Ratio of ESP reports to platform mention in public reports	15.9:1	204.3:1	4.1:1	8.4:1	1:1.5	1:2.4	1:2.3	1:7.7	0:1



No reports were submitted by Wizz. 'Google' here encompasses Google messaging products but does not count mentions of Youtube. Meta at times submits a single CyberTipline report regarding an event involving multiple Meta services (for example a report made by Facebook may include sextortion occurring on WhatsApp) - see footnote 40 for additional details).

© 2024 Thorn

to be reported by the victim (e.g., children may be more likely to report to platforms where imagery was posted or threatened to be posted). Most importantly, however, some platforms may not have reporting flows that allow victims to easily make these reports and convey that sextortion is occurring, or may fail to pass along the data clearly to NCMEC. It is important for platforms to examine aspects of reporting and moderation processes to make sure that victims of sextortion are heard and reports are escalated appropriately to get victims services in these high-risk cases of exploitation.

We should note that financial payment platforms may also report to NCMEC. These platforms play a valuable role in combating financial sextortion, especially through enhanced signal sharing between these payment companies and other platforms involved in sextortion. We observe sextortion reports from PayPal Inc., which includes both Venmo and PayPal. However, many other financial service companies are not registered to report to NCMEC or do not make substantive reports. For example, our analysis does not have any reports from Block Inc., which includes Cash App, the most commonly mentioned payment platform in sextortion reports we examined. While some cases of sextortion over a payment platform may not be easily detectable, the value of PayPal reporting sextortion instances underlines that there are actions that can be taken to keep children using financial platforms safe, whether through reporting to the CyberTipline or through other ways to respond in a targeted way to sextortion.

In addition, public reports will not fully encompass the scale or platform landscape of sextortion cases, as we know many victims of sextortion do not report their experiences. Only 43% of 13- to 17-year-old children who had experienced blackmail or threats reported it to a platform in a recent survey[44]; an older survey specific to sextortion found that 21% of victims of sextortion[45] reported their experience to a website/app and 16% to law enforcement. This means that proactive detection of sextortion may be necessary, both because many victims will not report it and in order to potentially prevent sextortion from occurring.

44 Thorn (2023). LGBTQ+ Youth Perspectives: How LGBTQ+ Youth are Navigating Exploration and Risks of Sexual Exploitation Online. https://info.thorn.org/hubfs/Research/Thorn_LGBTQ+YouthPerspectives_June2023_FNL.pdf

45 Wolak, Janis and David Finkelhor (2016) "Sextortion: Findings from a Survey of 1,631 Victims." https://www.thorn.org/wp-content/uploads/2016/08/Sextortion_Report.pdf

© 2024 Thorn

Conclusions

Increases in multi-sector and cross-industry collaborations are leading to new programming and product features to combat the risk of financial sextortion confronting minors. However, significant gaps remain as we work to defend against these dangers. We hope that the findings throughout this report inform better awareness of the issues and better prevention efforts addressing sextortion. Here are some specific ways that our findings might connect to further action:

> **Though common, financial sextortion does not exclusively target children who have shared an intimate image, and use of generative AI technologies may lead to an increase in these cases.**

This study shows the majority of financial sextortion is enabled by catfishing of teenage boys. However, it also highlights several techniques being leveraged to threaten victims without them needing to share an image directly. Reports of account hacking and use of generative AI technologies to create (or threaten to create) photorealistic explicit imagery were also identified.

While discussing the risks that come with a decision to share intimate images remains an element of many safeguarding conversations, perpetrator tactics no longer exclusively rely on a victim sharing imagery. Acknowledging these additional tactics is critical to effectively combating financial sextortion and preventing the weaponization of intimate imagery to silence and isolate victims.

> **Financial sextortion relies heavily on inflaming a victim's fears around the impact of having their nudes exposed, such as that they would go viral or send the child to jail.**

This study showed those perpetrators targeting victims for financial sextortion regularly employ language that focuses on extreme outcomes to the victim. In addition, they attempt to keep the victim silent and isolated by using countdowns and rapid, repeated, messaging, reducing the likelihood for the victim to get help.

The potential of viral spread of images, prosecution, and other life-altering outcomes are being weaponized in these cases. Groups working on prevention efforts should consider such fears during prevention campaigns and care must be taken to avoid messaging that overly relies on negative outcomes and the permanence of these outcomes, instead ensuring messaging includes opportunities to act to reduce negative outcomes and reduces the shame of victimization. In addition, the velocity of these cases demands proactive safeguarding on these topics so young people are empowered with the knowledge that deliberate attempts may be made to isolate them.

© 2024 Thorn

Perpetrators of financial sextortion often leverage accounts that pose as children or hacked child accounts.

While progress has been made by some platforms in limiting how adults can contact children on their social media platforms, a significant portion of financial sextortion threats are originating from accounts that appear to be other children, thereby evading basic platform safeguards that prevent accounts that identify as adult to interact with minors. It is essential for platforms to develop and implement safeguards for these contexts, where systems designed for detecting and limiting contact from suspicious adults might not be effective.

Financial sextortion is a global phenomenon.

We have observed that perpetrator groups from some countries are targeting victims other than English-speaking children. Detection tools, moderation endeavors, and prevention messaging should avoid focusing too narrowly on English when addressing sextortion issues, keeping the global nature of this issue in mind.

Currently available data limits our ability to explore the efficacy and extent of proactive detection practices vs. user reporting direct to platforms.

The reports made to the CyberTipline by ESPs can originate from both user reports to the ESP and proactive detection practices. Unfortunately, we are unable to differentiate which reports come from which source in this study, thereby limiting our ability to fully explore opportunities for increased impact.

This distinction becomes particularly important as shifts towards end-to-end encryption may put a halt to some forms of proactive detection, and many platforms have highlighted the reliance on user reporting to ensure these environments remain safe. It is important to have transparency regarding whether planned changes will have large impacts on sextortion prevention and reporting.

Reports from the public concerning financial sextortion suggest a far wider list of impacted platforms than are actively reporting to the CyberTipline.

Platforms play a vital role in combating online child sexual exploitation. However, worryingly, this report highlights that there are many platforms where sextortion is occurring but for which few reports are submitted.

© 2024 Thorn

Platforms should ensure they have clear policies prohibiting the use of their service for the purposes of sextortion and have scalable content moderation tools and workflows that can enforce such policies. In addition, platforms should explore ways to optimize reports to the CyberTipline for maximum impact on child safety. Finally, platforms should work closely with experts in the child safety space and other members of industry to stay current on the evolving tactics being leveraged in these cases. Intelligence sharing across the ecosystem can accelerate and improve detection of such abuses, creating safer online places for young people.

Financial sextortion continues to be a major issue, and it is important to have resources available that can address financial sextortion and help children. Some important resources are provided below:

Resources for those experiencing sextortion or worried about their imagery:

- https://report.cybertip.org/ (or contactgethelp@ncmec.org or call 1-800–THE–LOST)
- Internationally: https://www.inhope.org/EN#hotlineReferral
- https://www.stopsextortion.com/
- https://nofiltr.org/resource/what-is-sextortion/

Resources for those worried about their imagery being shared:

- https://takeitdown.ncmec.org (for minors)
- https://StopNCII.org (also for adults)
- https://www.ncmec.org/gethelpnow/isyourexplicitcontentoutthere

Additional resources and information on sextortion:

- https://www.ncmec.org/theissues/sextortion
- https://www.thorn.org/research/grooming-and-sextortion
- https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-scams-and-crimes/sextortion
- NCMEC Connect Sextortion module https://connect.missingkids.org

© 2024 Thorn

THORN

thorn.org | info@thorn.org

NATIONAL CENTER FOR MISSING & EXPLOITED CHILDREN®

ncmec.org

© 2024 Thorn