Australian Government
Australian Institute of Criminology

Trends & issues in crime and criminal justice

No. 669

Abstract | Online CSAM offending is a challenge for law enforcement, policymakers and child welfare organisations alike. The use of online warning messages to prevent or deter an individual when they actively search for CSAM is gaining traction as a response to some types of CSAM offending. Yet, to date, the technical question of how warning messages can be implemented, and who can implement them, has been largely unexplored. To address this, we use a case study to analyse the actions individuals and organisations within the technology, government, non-government and private sectors could take to implement warning messages. We find that, from a technical perspective, there is considerable opportunity to implement warning messages, although further research into efficacy and cost is needed.

How to implement online warnings to prevent the use of child sexual abuse material

Charlotte Hunn, Paul Watters, Jeremy Prichard, Richard Wortley, Joel Scanlan, Caroline Spiranovic and Tony Krone

Introduction

The rapid expansion of the internet and related technologies has seen the availability of child sexual abuse material (CSAM) grow exponentially (Holt et al. 2020; WePROTECT Global Alliance 2021; Westlake 2020). The amount of CSAM online is described as 'overwhelming' (Europol 2020: 36). Legal definitions of CSAM vary globally but typically include images, videos and texts depicting or describing infants, children and young people in sexual poses through to depictions of extreme sexual assaults and rape (Westlake 2020). The increasing availability of CSAM online has seen an upwards trend in the frequency of online CSAM offending (Europol 2020)—a trend which intensified during the COVID-19 pandemic (Interpol 2020).

The availability of CSAM online means that many internet users come into contact with CSAM (Broadhurst 2019; Westlake 2020). Contact can occur via searches of indexed content on the surface web (Westlake 2020), including searches of ostensibly

Child Sexual Abuse Material Reduction Research Program

legal adult pornography sites (Morgan & Lambie 2019), and non-indexed but mainstream peer-to-peer (P2P) networks (Wolak, Liberatore & Levine 2014). Some internet users who encounter CSAM may report it, or they may simply ignore it (Internet Watch Foundation (IWF) 2013). However, a proportion of internet users who are exposed to CSAM will go on to deliberately view it, even if the initial exposure was accidental (Protect Children 2021). Exposure represents an important factor in CSAM offending onset (Wortley & Smallbone 2012). Various pathways to exposure exist, but three broad types of CSAM offenders have been identified. These are individuals whose offending is:

- consistent with a diagnosis of paedophilia;
- part of a hypersexual disorder, with CSAM consumption part of a broader range of behaviours; and/or
- the result of impulsive risk-taking behaviour (Seto & Ahmed 2014).

Accessibility of CSAM online

Search engines are pointed to as the 'most common method' of finding CSAM online (Steel 2015: 151). Recently, Google has undertaken significant work to ensure that CSAM is not indexed through their platform (Google 2020). Yet other search engines have less developed (or undeveloped) CSAM detection programs, and researchers conclude that 'if an individual can access and use a search engine with a modicum of skill, they can assuredly find [CSAM]' (Westlake, Bouchard & Girodat 2017: 291). Publicly accessible CSAM websites, or websites that contain hyperlinks to CSAM, are 'overt' and 'do little to hide their intended purpose' (Westlake, Bouchard & Girodat 2017: 289). Research (Morgan & Lambie 2019) also suggests that CSAM is accessible via searches of popular legal adult pornography websites. Further, studies examining keyword searches of P2P networks show that these networks are frequently used to search for and share CSAM (Bissias et al. 2016; Interpol 2020). And, together with P2P networks, the darknet (eg Tor) is a popular means of accessing and sharing CSAM (Europol 2022).

Challenges in tackling CSAM online

In response to the growing problem, the capacity of law enforcement agencies (LEAs) to detect, investigate and prosecute CSAM offenders has become increasingly sophisticated, with attention rightly focusing on the 'most immediate and serious threats' (Holt et al. 2020: 6). Human and resourcing constraints limit the capacity of LEAs to respond to CSAM offending (Holt et al. 2020; see also Broadhurst 2019; Carr 2017; Westlake, Bouchard & Girodat 2017). These constraints have only tightened since the COVID-19 pandemic (Europol 2020; Interpol 2020). Recognition of this, together with an acknowledgement that CSAM offending is not a problem that LEAs can 'arrest their way out of' (Quayle & Koukopoulos 2019: 348), has been a catalyst for researchers, the technology industry and LEAs to explore other responses to tackle CSAM offending. An area attracting particular interest is the potential for early interventions to prevent an escalation from viewing to more serious forms of CSAM offending (WePROTECT Global Alliance 2019).

In addition, industry plays a part in the detection and reporting of CSAM, as electronic communication service providers in the United States are required to report instances of detected CSAM offending to the CyberTipline of the National Center for Missing and Exploited Children under

18 US Code § 2258A. In 2021, 29.1 million reports (of a total 29.3 million) were made under this Code (National Center for Missing and Exploited Children 2022). While 230 service providers submit reports to the CyberTipline, the vast majority of reports were made by Facebook (22,118,952 reports) followed by Instagram, WhatsApp and Google, with Apple making only 160 reports. In August 2021 Apple announced that it would scan photos in its iCloud Photos for CSAM (Schneider 2021a). While providers already scan photos shared on platforms, the Apple proposal would allow the scanning of photos on phones, raising concerns about privacy (Green & Stamos 2021). Following public debate, the proposal was apparently dropped (Schneider 2021b).

Value of automated responses to CSAM online

Against this background, researchers have stressed the value of automated methods of tackling CSAM online, informed by theoretical models of crime prevention (Prichard et al. 2019; Quayle & Koukopoulos 2019; Smallbone & Wortley 2017). Automated methods of identifying CSAM and responding to offending have been developed and implemented by the technology industry in recent years, often in collaboration with LEAs and/or non-government organisations (NGOs). For example, the use of hashing technology has reduced the need for visual inspection to identify previously identified CSAM images by automating identification of duplicate copies of a CSAM image based on a library or database of hashes (eg Microsoft's PhotoDNA technology; Microsoft 2022). Other examples of automated responses include URL blocking (Carr 2017) and filtering software (Quayle 2013).

Warning messages as a prevention strategy

Automated online warning messages are a form of secondary, offender-focused prevention, as they target an at-risk group to 'prevent an offence before it occurs' (Wortley & Smallbone 2017). Currently, this is attempted through deterrence content, which alerts users to the illegality of CSAM, and referral messages, which refer users to therapeutic services that can address their attraction to CSAM (Prichard et al. 2022a). A warning message is displayed to an individual within the online environment in response to a user's conduct at the very time the user is contemplating engaging in illegality.

Warning messages have typically been used to respond to two types of user conduct:

- when an internet user enters a CSAM keyword as a search query into a search engine (eg Google 2020); and
- when an internet user attempts to access a URL which has been removed due to CSAM content (Bailey, Squire & Thornhill 2018).

The first type of warning message typically appears at the top of the list of search results (see Quayle & Koukopoulos 2019), while the second type of warning message appears as an HTML 'stop page' (Wortley & Smallbone 2012) or 'splash page' (see Bailey, Squire & Thonhill 2018). Both LEAs (eg the Norwegian police) and search engine operators (eg Google) have used messages that warn internet users of the potential criminality of their behaviour (Smallbone & Wortley 2017; Wortley & Smallbone 2012). Search engine operators along with NGOs (eg the Lucy Faithfull Foundation) have also used messages to encourage users to report CSAM (Google 2020; Steel 2015) and/or seek help (Bailey, Squire & Thornhill 2018).

Additionally, some pornography companies have recognised the role they can play in dissuading users from attempting to access illegal content. Notably, MindGeek, which operates PornHub, uses warning messages and a chatbot to reduce the likelihood that users access the results of search terms that might lead to CSAM (see Prichard et al. 2022a).

The display of an automated warning message is the result of a manipulation of the content sent by the web server in response to a particular user action. Webpages displayed in a web browser require a programming language, such as JavaScript, to create a dynamic webpage interface. By manipulating the programming language behind the webpage, a warning can be automatically displayed when the user performs a particular action. The warning message can appear as either an outbound or inbound display. An outbound display occurs when a warning is presented to the user when they send a request for content over the internet (ie when they click 'search')—before the request leaves the user's device. An inbound display occurs when a warning is presented on the request's return journey to the user but before (or instead of) the search results being given to the user.

The action that triggers a warning message can take a number of forms, as programming languages provide numerous mechanisms—for example, by detecting when a button is clicked, when text is entered into a field, or when a browser window is closed. There are also a number of ways to present a warning message to the user, including as a pop-up, an alert, or as an HTML page (eg a stop or splash page). The warning message could be small and designed to automatically recede (like reminders to update software) or it could fill the entire screen and require a user action to make it disappear. In addition, the content of a warning message could include text, static images, moving images (eg GIFs), audio or video.

Keyword searches are commonly the trigger action for displaying a warning message to a user. Several methods for identifying and monitoring keyword searches for CSAM exist (Belbeze et al. 2009). Research shows that when keyword search terms are regularly monitored and updated they are a reliable indicator of CSAM content (Steel 2015; Westlake, Bouchard & Frank 2017). A number of organisations maintain keyword lists, including the IWF, which updates its *Keyword List* monthly (IWF 2020). In addition, lists of URLs identified as containing CSAM are maintained by a number of organisations including LEAs (eg Interpol) and NGOs such as the IWF (IWF 2020). URLs and filenames may also contain metadata indicative of the content, which may be indexed on large-scale file hashing databases (effectively a library of known CSAM images) maintained by law enforcement and other agencies (Sanchez et al. 2019).

A number of studies suggest that warning messages can prevent undesirable behaviour offline (Hammond 2011) and online (Maimon et al. 2014). Although research into the efficacy of warning messages to deter online CSAM offending is scarce (Prichard et al. 2019; Wortley & Smallbone 2012), recent empirical studies have found that warning messages dissuaded internet users from viewing 'barely legal' pornography online (Prichard et al. 2021) and sharing potentially illegal sexual images (Prichard et al. 2022b). Relevant to the focus of this paper, the practicalities of implementing online CSAM warnings has received little attention in the academic literature.

Implementing online warning messages

In this section we analyse the types of actions a wide range of organisations across the technology, government and private sectors could take to implement warning messages to prevent CSAM use. By way of identifying the range of actors involved, consider the following scenario describing an individual searching for CSAM using keyword searches before their behaviour progresses to the use of more complex technology including virtual private networks (VPNs) and the darknet (eg Tor):

> An individual (the user) activates their device—for instance, a computer or a smartphone. The user opens a web browser through their operating system. Their device connects to the internet via a broadband modem using an internet service provider (ISP). At this point, the user navigates to a search engine website, P2P website or other website with a search option (eg a pornography website). Next, the user enters a search term associated with CSAM. If the user enters the term into a search engine, the search engine will process the query and provide a set of resulting links to webpages or uniform resource locators (URLs). The URLs will point to the particular internet web server that hosts the content most closely matching the query. The user can then click on a link which will take them to the particular website (URL) where they can view CSAM. Similarly, if the user accessed a P2P website or other searchable website, the query will be processed and a set of results will be displayed to the user. Again, they can click on the material they want to view. For the purpose of this case study, we assume that, at a certain point, the user obscures their identity by downloading and using a VPN and/or Tor software.

In this scenario, the user may hold the ISP account, or they may be using a shared network with the account held by another private individual (eg the user is using a shared network as part of a family or share-house arrangement) or a public/private institution. In what follows, we identify the actors involved in the above case study and describe the actions they may be able to take to implement warning messages.

Account holders

An account holder is the individual or entity that pays the ISP for access to the internet. This can include a private individual or an institutional, corporate or other business account holder that allows others to use that account. An account holder may require identity verification before their network is accessible to others. For example, a corporation may require users to register to access the internet, or an account holder may require registration but not identity verification, as in the case of a food outlet that provides free public wi-fi to customers. Alternatively, access may be provided with neither registration nor identity verification. An account holder could take a number of actions to implement automated warning messages, including installing security software and employing a link-checker or proxy server.

- **Security software**—End-point security has the potential to be programmed to display a warning message in response to a user typing a keyword into a web browser or attempting to access a banned URL. Most common forms of security software work by monitoring all data on a user's computer. There are a variety of types of security software. For example, 'Net Nanny' type software monitors what users type into web browsers and blocks inappropriate results. This type of security software could be extended to not only block access but also display a warning message in response to the user action described in the above scenario.

- **Link-checker service**—A link-checker service, which provides redirection, is a cloud-based service that checks links that a user may try to access—for example, a link contained in an email. By employing a link-checker service, an account holder uses a third party to check all links that appear in its users' emails and, if the link is defined as malicious, generates a warning. This type of service could be extended to check the links generated when a user conducts a CSAM keyword search or attempts to access a banned URL.
- **Proxy server**—A proxy server stands between the user's computer and the rest of the internet. A proxy server can monitor outbound traffic from a user's computer on a network and, if a user searches for a keyword or attempts to access a banned URL, it could be programmed to deliver a warning message to the user and/or the account holder.

Operating system developers and vendors

Operating system (OS) developers and vendors provide the basic platform on which applications run, including web browsers. The three most common operating system are Microsoft Windows, macOS and Linux. The operating system manages and coordinates the computer's hardware (ie the computer's data storage) and the software (ie applications including web browsers such as Google Chrome, Safari and Firefox). An OS has the capacity to take various actions. For example, because an OS has access to all memory that is used by the browser, an OS could be programmed to detect when a user types a CSAM search term into a browser and generate a warning, displayed in either a separate window or as a pop-up window within the browser itself.

Browser developers and vendors

Browser developers and vendors create the software, or interface, between a user's computer and the internet. For an existing computer program, a browser developer could develop a 'plug-in'—a customised software component—to monitor the terms that a user types into a search engine and generate a warning message in response to the above scenario. For example, a range of plug-ins are currently available to protect users, including AdBlock, which blocks all internet ads. A vendor could also make such a plug-in a native feature of their browsers, thereby making it available to anyone who uses that browser.

Search engine operators

Internet search engines—including Google, Bing and Baidu, to name but a few—are software systems through which a user can systematically search the internet using a text-based query (ie a keyword). To display 'relevant' advertising, many search engines maintain a full set of data about a user's search history (Price 2021). By matching search terms with a list of suspicious or known CSAM-related terms, a search engine can either display a warning message to a user when they search for a particular keyword or provide a redirection to a third-party link-checking service, as described above, for the links returned as part of the search results. This latter option may be particularly relevant for search engine operators that do not have the resources of the larger companies like Google. The feasibility of search engine operators displaying warnings has been demonstrated by Google, which presents deterrence ads when an individual searches for CSAM-related terms in a number of countries (Google 2020). Another option is that, if a URL page in the search results is known to contain CSAM, a warning message could be displayed if the user clicks on the link, in addition to the URL being blocked.

Internet service providers

Commercial internet service providers, such as Telstra, Vodafone, Optus and iiNet (among others), sell internet connections and services to private individuals and organisations, including institutions and corporations. The customers of ISPs are the account holders. ISPs play the crucial role of connecting a user, either as the account holder or via an account holder's account, to the internet. ISPs have the capacity to monitor all inbound and outbound internet traffic between the user and internet servers using the internet protocol (IP). ISPs also have the capacity to block URLs, as further discussed below, and theoretically to monitor search terms being sent by users to search engines.

Virtual private network vendors

VPN services enable a user to send and receive data within an encrypted private network using a shared or public internet network. In effect, a VPN creates a secure tunnel between a user and the local network they want to access (eg a workplace network). Vendors of VPNs fall into two types: those providing secure VPN services based in Australia and those operating beyond Australia's regulatory environment. While options for implementing warning messages are more limited, a warning message could, in theory, be displayed to a user if they tried to initiate a connection to a VPN. However, only ISPs in some jurisdictions maintain a list of IP address ranges belonging to VPNs, which may make identification difficult.

Domain name service providers

Domain name service (DNS) providers translate a readable domain address into an IP address—that is, a numerical identifier. When a URL is typed or web page bookmark is clicked, the computer sends a DNS query to look up the IP address (ie the numerical identifier of the server it is attempting to contact). This look-up matches the human readable address, such as www.google.com, to an IP address like 142.250.70.142. The computer uses the latter to communicate with that server. DNS providers are in a position to block access to CSAM material by not responding to DNS queries to websites that have been previously identified as containing CSAM. DNS providers are also able to redirect the traffic to a different location, and this provides an opportunity for warning messages to be displayed or for users to be referred to support services. As such, DNS providers are well placed to remove access to websites that appear on a deny list, or fail a check with a link-checking service.

Tor software

Tor software, named after the original software project, The Onion Router, enables anonymous communication through layers of encryption (like the layers of an onion), hiding the source and destination addresses from observers such as ISPs or government entities. The Tor network requires the use of a modified internet browser (of which there are several) which supports its encryption protocols to enable access to what is termed the darknet. These websites are not otherwise accessible. There are essentially two parts to Tor: the web browser and the darknet itself. It would be challenging to insert messages directly through the Tor network, as the content is encrypted multiple times. However, as the Tor client runs on an operating system, it may be feasible to deliver a warning message to a user via the operating system notification window—that is, to display the warning message at the point where traffic enters and exits the Tor network.

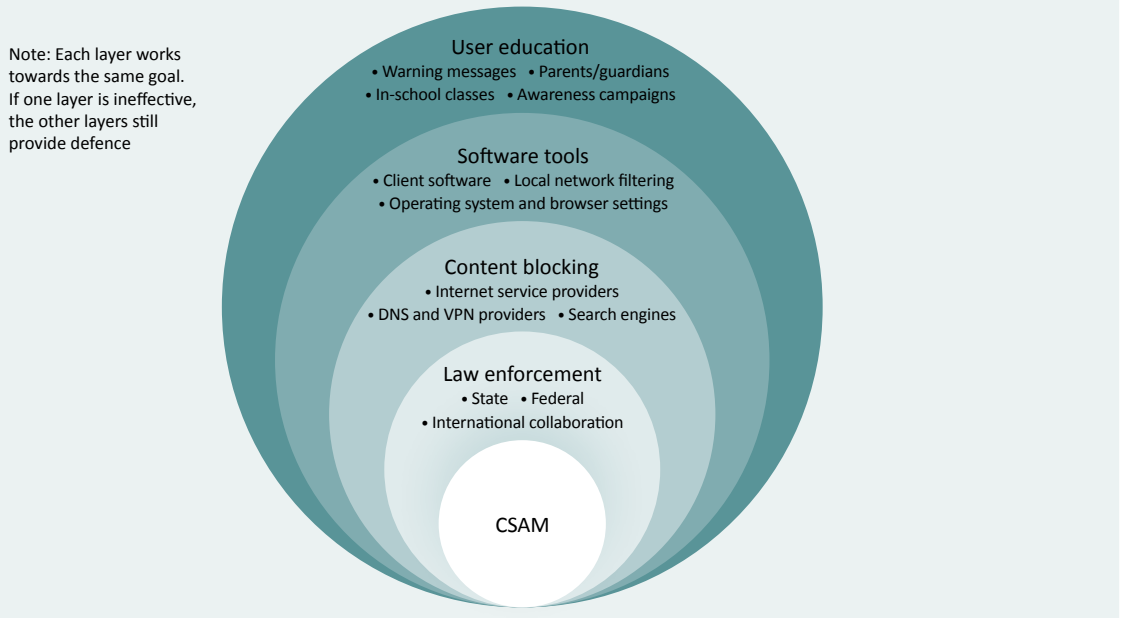
Third parties

Third parties include government departments, statutory bodies (eg Office of the eSafety Commissioner), agencies (eg LEAs), NGOs, and for-profit companies who operate in the child protection sector. The action of third parties with regard to warnings has largely been limited to collaborative action—for example, between LEAs and ISPs (Williams 2005) or between ISPs and NGOs (eg Google and INHOPE). However, while the capacity of third parties to implement warnings unilaterally may be limited, some third parties may be able to develop or adapt consumer software packages. One example would be the adaption of filtering software to include warnings in response to specific user actions (eg keyword searches), while another is the development of a downloadable plug-in that could operate, effectively, as an in-browser version of filtering software. Admittedly, both these options are premised on voluntary adoption by the relevant user, whether a private individual or an institutional, corporate or other business account holder. Moreover, indirectly, governments can enact laws requiring key actors in the technology industry to implement warnings. Indeed, as discussed below, a key example from the Australian context is the legislative requirement that ISPs block Australian internet users' access to URLs that contain some types of CSAM (Conroy 2012).

Discussion

From a technical perspective, the best approach to implement an automated warning message system is likely to be one based on the cooperation of multiple actors using a combination of actions—that is, an approach to cybersecurity based on the theory of defence in depth (Coole, Corkill & Woodward 2012), as illustrated in Figure 1. This requires the cooperation of multiple actors within the technology industry, and the implementation of measures by account holders. The premise of the approach is that, if a single layer of defence fails, the other layers of defence can still provide protection. New methods of circumventing protective measures will always be created, so having multiple measures in place increases the effectiveness and resiliency of the system.

Figure 1: An illustration of the concept of defence in depth



Yet even unilateral actions by individual actors may offer a partial solution. As described above, an account holder could take a number of actions to effectively implement warning messages. At present, the key limiting factor is the lack of relevant consumer products available for private individual account holders. Private individuals tend not to have the level of expertise needed to customise software. As such, actions by account holders are likely to be restricted mostly to the larger institutional, corporate or other business account holders who have the resources to purchase such products (eg NetClean ProActive: https://www.netclean.com/proactive/) or who can acquire the expertise necessary to customise or develop new products. For example, with regard to the use of proxy servers, an institution or corporation could require all network access within the organisation to be undertaken through a proxy server and, as described above, the proxy server could monitor all outbound traffic from computers on the network, displaying a warning message if a user searched for a CSAM keyword or attempted to access a banned URL. Such actions may, however, require an impetus for action, such as a requirement for warnings to be explicitly included within an organisation's cybersecurity policy.

The actors who can perhaps take the most obvious actions to implement warning messages are in the technology industry. Resource and expertise constraints may limit the capacity of some actors within this industry to take action (Thorn 2020). For others, such constraints are less likely to be an issue, in particular for those who develop and/or run the infrastructure where CSAM may be present (Holt et al. 2020). Indeed, in March 2020, several of the largest actors in the technology industry (including Microsoft, Yahoo! and Vodafone) committed to pursue the prevention of CSAM offending through voluntary principles developed by the WePROTECT Global Alliance (2020).

More generally, however, the technology industry has been criticised for being less than enthusiastic about 'proactively policing' CSAM online (Holt 2018). For example, while the vital role that ISPs can play in this context is well recognised (Holt et al. 2020), with few exceptions, ISPs have been condemned for having 'sat mostly at the sidelines' (Westlake 2020: 1236; World Health Organization 2020). One exception is where there is a legislative requirement to take action. Since 2010, Australia's largest ISPs have been required under the Commonwealth *Telecommunications Act 1997* s 313(1), to block URLs on Interpol's 'worst of' list—that is, URLs containing the most severe forms of CSAM (Conroy 2012). If an Australian internet user attempts to access a blocked URL, a stop page is displayed which provides reasons for the block and contact details for follow-up (AFP and telecommunications targeting online crime, *About the House*, February 2015: 11. http://classic.austlii.edu.au/au/journals/AboutHouseMag/2015/8.html). We note that the Commonwealth *Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018* amended the *Telecommunications Act 1997* to authorise communication providers to give voluntary technical assistance when requested by law enforcement and that this could extend to the adoption of warning messages.

As noted above, it would also be technically possible for warning messages to be implemented in response to the use of VPNs and Tor software. A range of strategies have been proposed on the Tor network to counter CSAM, with the potential to uncover at least partial identity information (Abbott et al. 2007; Loesing, Murdoch & Dingledine 2010). Even so, tackling nefarious activities on the Tor network presents an ongoing challenge for law enforcement and government.

Third parties already play a vital role in maintaining the data necessary for a warning to be delivered (eg keyword and URL lists). Moreover, as mentioned above, a number of third parties including NGOs and LEAs have collaborated with the technology industry to develop and implement warning messages. There are also examples of key actors within the technology industry providing financial support to third parties in this area. For example, Google offers paid advertising credits to NGOs and charities who run reporting hotlines for CSAM (Google 2020). Going further, some third parties may be in a position to commission the development of a range of services for individual and other users, including security software, link-checker services and proxy servers to implement warning messages. A potentially simpler option would be for a third party to commission the development of a plug-in for web browsers, which could be programmed to generate a warning message when any user searches for a CSAM-related term or attempts to access a banned URL using that browser.

Conclusion

In this paper, we have provided a technical overview of the actions that could be taken by various actors to display a warning message to a user when they search for CSAM online. In doing so, we have shown that, from a technical perspective at least, there is considerable opportunity for a range of actors across the technology, government and private sectors to implement warning messages. The context for this overview is that the availability and accessibility of CSAM online means that the average internet user may come into contact with CSAM, and even inadvertent exposure may lead to further offending. Moreover, the capacity of LEAs to respond to every instance of CSAM offending—particularly viewing or accessing behaviours—is necessarily limited. In providing this technical overview, we acknowledge that we have not presented a complete picture of the issues associated with implementing warning messages—in particular, questions about the efficacy, scalability and reach of warning messages, and the cost of implementing them, fall outside the scope of this paper (but see Prichard et al. 2021). Further research examining these factors is needed.

References

URLs correct as at November 2022

Abbott T, Lai K, Lieberman M & Price E 2007. Browser-based attacks on Tor. In N Borisov & P Golle (eds), *Privacy enhancing technologies.* Lecture Notes in Computer Science vol 4776. Springer: 184–199. https://doi.org/10.1007/978-3-540-75551-7_12

Bailey A, Squire T & Thornhill L 2018. The Lucy Faithfull Foundation: Twenty-five years of child protection and preventing child sexual abuse. In R Lievesley, K Hocken, H Elliott, B Winder, N Blagden & P Banyard (eds), Sexual crime and prevention. Cham: Palgrave Macmillan: 57–82. https://doi.org/10.1007/978-3-319-98243-4_3

Belbeze C et al. 2009. Automatic identification of paedophile keywords. *Measurements and Analysis of P2P Activity Against Paedophile Content Project*

Bissias G et al. 2016. Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks. *Child Abuse & Neglect* 52: 185–199. https://doi.org/10.1016/j.chiabu.2015.10.022

Broadhurst R 2019. Child sex abuse images and exploitation materials. In R Leukfeldt & T Holt (eds), *Cybercrime: The human factor.* Routledge: 310–336. https://doi.org/10.4324/9780429460593-14

Carr J 2017. A brief history of child safety online: Child abuse images on the internet. In J Brown (ed), *Online risk to children: Impact, protection and prevention*. Wiley: 5–21. https://doi.org/10.1002/9781118977545.ch1

Conroy S 2012. *Child abuse material blocked online, removing need for legislation*. Media release, 9 November. https://parlinfo.aph.gov.au/parlInfo/search/summary/summary.w3p;adv=yes;orderBy=customrank;page=0;query=%E2%80%9CChild%20abuse%20material%20blocked%20online,%20removing%20need%20for%20legislation%E2%80%9D

Coole M, Corkill J & Woodward A 2012. Defence in depth, protection in depth and security in depth: A comparative analysis towards a common usage language. *Proceedings of the 5th Australian Security and Intelligence Conference*. https://ro.ecu.edu.au/asi/

Europol 2022. Child sexual exploitation. https://www.europol.europa.eu/crime-areas-and-trends/crime-areas/child-sexual-exploitation

Europol 2020. *Internet organised crime threat assessment 2020*. https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2020

Google 2020. Fighting child sexual abuse online. https://protectingchildren.google/intl/en/

Green MD & Stamos A 2021. Apple wants to protect children. But it's creating serious privacy risks. *New York Times*, 11 August. https://www.nytimes.com/2021/08/11/opinion/apple-iphones-privacy.html

Hammond D 2011. Health warning messages on tobacco products: A review. *Tobacco Control* 20(5): 327–337. https://doi.org/10.1136/tc.2010.037630

Holt TJ 2018. Regulating cybercrime through law enforcement and industry mechanisms. *Annals of the American Academy of Political and Social Science* 679(1): 140–157. https://doi.org/10.1177/0002716218783679

Holt TJ, Cale J, Leclerc B & Drew J 2020. Assessing the challenges affecting the investigative methods to combat online child exploitation material offenses. *Aggression and Violent Behavior* 55: 101464. https://doi.org/10.1016/j.avb.2020.101464

Internet Watch Foundation (IWF) 2020. Keywords list. https://annualreport2020.iwf.org.uk/tech/keyservices/keywords

Internet Watch Foundation (IWF) 2013. New study reveals child sexual abuse content as top online concern and potentially 1.5m adults have stumbled upon it. https://www.iwf.org.uk/news/new-study-reveals-child-sexual-abuse-content-as-top-online-concern-and-potentially-1-5m-adults

Interpol 2020. *Threats and trends child sexual exploitation and abuse: COVID-19 impact*. https://www.interpol.int/en/News-and-Events/News/2020/Interpol-report-highlights-impact-of-COVID-19-on-child-sexual-abuse

Loesing K, Murdoch SJ & Dingledine R 2010. A case study on measuring statistical data in the Tor anonymity network. In *Financial cryptography and data security*. Lecture Notes in Computer Science vol 6054. Springer: 203–215. https://doi.org/10.1007/978-3-642-14992-4_19

Maimon D, Alper M, Sobesto B & Cukier M 2014. Restrictive deterrent effects of a warning banner in an attacked computer system. *Criminology* 52(1): 33–59. https://doi.org/10.1111/1745-9125.12028

Microsoft 2022. *PhotoDNA*. https://www.microsoft.com/en-us/photodna

Morgan S & Lambie I 2019. Understanding men who access sexualised images of children: Exploratory interviews with offenders. *Journal of Sexual Aggression* 25(1): 60–73. https://doi.org/10.1080/13552600.2018.1551502

National Center for Missing and Exploited Children 2022. *2021 CyberTipline reports by electronic service providers (ESP)*. https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata

Price C 2021. 20 great search engines you can use instead of Google. https://www.searchenginejournal.com/alternative-search-engines/271409

Prichard J et al. 2022a. *Online messages to reduce users' engagement with child sexual abuse material: A review of relevant literature for the reThink chatbot*. Report for the Lucy Faithfull Foundation. Hobart: University of Tasmania. https://eprints.utas.edu.au/46223/

Prichard J et al. 2022b. Warning messages to prevent illegal sharing of sexual images: Results of a randomised controlled experiment. *Trends & issues in crime and criminal justice* no. 647. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78559

Prichard J et al. T 2021. Effects of automated messages on internet users attempting to access "barely legal" pornography. *Sexual Abuse* 34(1): 106–124. https://doi.org/10.1177/10790632211013809

Prichard J, Krone T, Spiranovic C & Watters P 2019. Transdisciplinary research in virtual space: Can online warning messages reduce engagement with child exploitation material? In R Wortley, A Sidebottom, N Tilley & G Laycock (eds), *Routledge handbook of crime science.* London: Routledge: 309–319. https://doi.org/10.4324/9780203431405

Protect Children 2021. ReDirection launch: 'Protecting children through prevention'. https://suojellaanlapsia.fi/2021/09/27/redirection-launch-protecting-children-through-prevention/

Quayle E 2013. Organizational issues and new technologies. In M Erooga (ed), *Creating safer organizations: Practical steps to prevent the abuse of children by those working with them*. John Wiley & Sons: 99–121. https://doi.org/10.1002/9781119943228.ch6

Quayle E & Koukopoulos N 2019. Deterrence of online child sexual abuse and exploitation. *Policing: A Journal of Policy and Practice* 13(3): 345–362. https://doi.org/10.1093/police/pay028

Sanchez L, Grajeda C, Baggili I & Hall C 2019. A practitioner survey exploring the value of forensic tools, AI, filtering, & safer presentation for investigating child sexual abuse material (CSAM). *Digital Investigation* 29: S124–S142. https://doi.org/10.1016/j.diin.2019.04.005

Schneider J 2021a. Apple confirms it will scan iPhone photo libraries to protect children. PetaPixel, 5 August. https://petapixel.com/2021/08/05/apple-confirms-it-will-scan-iphone-photo-libraries-to-protect-children/

Schneider J 2021b. Apple wipes controversial child abuse photo scanning from its site. PetaPixel, 16 December. https://petapixel.com/2021/12/16/apple-removes-all-references-to-csam-photo-scanning-from-its-site/

Seto MC & Ahmed AG 2014. Treatment and management of child pornography use. *Psychiatric Clinics of North America* 37(2): 207–214. https://doi.org/10.1016/j.psc.2014.03.004

Smallbone S & Wortley R 2017. Preventing child sexual abuse online. In J Brown (ed), *Online risk to children: Impact, protection and prevention*. Wiley: 143. https://doi.org/10.1002/9781118977545.ch8

Steel CM 2015. Web-based child pornography: The global impact of deterrence efforts and its consumption on mobile platforms. *Child Abuse & Neglect* 44: 150–158. https://doi.org/10.1016/j.chiabu.2014.12.009

Thorn 2020. The road to Safer: Equipping industry to end CSAM. https://www.thorn.org/blog/announcing-safer-built-by-thorn-eliminate-csam/

WePROTECT Global Alliance 2021. *Global threat assessment 2021: Working together to end the sexual exploitation of children online.* https://www.weprotect.org/global-threat-assessment-21/#report

WePROTECT Global Alliance 2020. *Voluntary principles to counter online child sexual exploitation and abuse*. https://www.weprotect.org/library/voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse/

WePROTECT Global Alliance 2019. *Global Threat Assessment 2019: Working together to end the sexual exploitation of children online.* https://www.weprotect.org/issue/global-threat-assessment/

Westlake B 2020. The past, present, and future of online child sexual exploitation: Summarizing the evolution of production, distribution, and detection. In T Holt & A Bossler (eds), *The Palgrave handbook of international cybercrime and cyberdeviance.* Cham: Palgrave Macmillan: 1225–1253. https://doi.org/10.1007/978-3-319-78440-3_52

Westlake B, Bouchard M & Frank R 2017. Assessing the validity of automated webcrawlers as data collection tools to investigate online child sexual exploitation. *Sexual Abuse* 29(7): 685–708. https://doi.org/10.1177/1079063215616818

Westlake B, Bouchard M & Girodat A 2017. How obvious is it? The content of child sexual exploitation websites. *Deviant Behavior* 38(3): 282–293. https://doi.org/10.1080/01639625.2016.1197001

Williams KS 2005. Facilitating safer choices: Use of warnings to dissuade viewing of pornography on the internet. *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect* 14(6): 415–429. https://doi.org/10.1002/car.920

Wolak J, Liberatore M & Levine BN 2014. Measuring a year of child pornography trafficking by US computers on a peer-to-peer network. *Child Abuse & Neglect* 38(2): 347–356. https://doi.org/10.1016/j.chiabu.2013.10.018

World Health Organization 2020. *Joint leaders' statement: Violence against children: A hidden crisis of the COVID-19 pandemic.* https://www.who.int/news/item/08-04-2020-joint-leader-s-statement---violence-against-children-a-hidden-crisis-of-the-covid-19-pandemic

Wortley R & Smallbone S 2012. *Internet child pornography: Causes, investigation, and prevention*. Santa Barbara, CA: Praeger

Dr Charlotte Hunn is a Research Fellow in the Faculty of Law, University of Tasmania.

Dr Paul Watters is Adjunct Professor of Cybersecurity at La Trobe University.

Dr Jeremy Prichard is Associate Professor in the Faculty of Law, University of Tasmania.

Dr Richard Wortley is Professor at University College London and the University of Waikato.

Dr Joel Scanlan is Senior Lecturer at the University of Tasmania and Associate Professor of Maritime Cyber Security at Western Norway University of Applied Sciences.

Dr Caroline Spiranovic is Senior Lecturer in the Faculty of Law, University of Tasmania.

Dr Tony Krone is Associate Professor at the University of Canberra.

This project is part of the AIC's Child Sexual Abuse Material Reduction Research Program, funded under section 298 of the Commonwealth *Proceeds of Crime Act 2002*.

General editor, *Trends & issues in crime and criminal justice* series: Dr Rick Brown, Deputy Director, Australian Institute of Criminology. Note: *Trends & issues in crime and criminal justice* papers are peer reviewed. For a complete list and the full text of the papers in the *Trends & issues in crime and criminal justice* series, visit the AIC website: www.aic.gov.au

ISSN 1836-2206 ISBN 978 1 922478 89 4 (Online)
https://doi.org/10.52922/ti78894

GPO Box 1936
Canberra ACT 2601, Australia
Tel: 02 6268 7166

www.aic.gov.au