Investigating Malware Epidemiology and Child Exploitation Using Algorithmic Ethnography

Paul A. Watters
La Trobe University
P.Watters@latrobe.edu.au

Abstract

A number of recent studies have investigated the role that advertising plays in funding and sustaining piracy sites. In this study, the composition of advertising on piracy websites hosting Child Exploitation Material (CEM) in Brazil was analysed, measuring both mainstream and "high-risk" ads[1] using samples generated from the most-complained about sites, as well as the most popular sites, using algorithmic ethnography. The results indicated that high-risk ads pose significant problems for Brazilian users, especially children, who may be harmed by exposure to ads promoting the sex industry, gambling, scams and malware. Links between Brazil-focused piracy websites and CEM are documented. Policy options for dealing with this issue are discussed. CEM poses a threat not only to the children depicted in it but also to children who might be exposed to it. Indeed, any viewer who is exposed to CEM may experience adverse psychological and behavioral consequences.

1. Introduction

Child exploitation is a serious and growing problem online. Whilst paedophiles are always present in the general population at large, the growth of the internet, and the easy availability of Child Exploitation Material (CEM), has increased supply, and has ultimately led to a corresponding increase in demand, especially from people who would not fit the traditional psychological profile of a paedophile. The behavioural mechanism that results in an internal escalation of sexual arousal may be related to incidental or non-intentional exposure to CEM. For example, a user of mainstream pornography sites may incidentally view images from the "teen" category, and

---

[1] Defined as categories which include malware, downloading sites, gambling, scams and the sex industry.

– in a state of heightened sexual arousal – develop arousal to these images through classical conditioning [24]. The novelty of the stimulus is especially important during this process, and most people would not have seen CEM previously. An intriguing question is whether other forms of incidental exposure – such as seeing CEM advertising – might also trigger such a response, especially since the viewing is not intentional. In this paper, we investigate the links between advertising and CEM, but focusing on the extent to which CEM might be supported (unwittingly) by the advertising industry, using a new methodology known as algorithmic ethnography [2]. This method tries to understand how we can infer the parameterisations of algorithms by making naturalistic observations of how they behave "in the wild", while seeking to reduce extraneous variation.

The data used in this study are drawn from piracy websites, which attract a very significant amount of advertising revenue in rapidly developing economies, such as those of the BRIC countries (Brazil, Russia, India and China). A "piracy website" can be defined as one which is primarily (or exclusively) devoted to the distribution of copyright infringing materials, whether at little or no cost [25]. While some of these sites charge a subscription fee, many others provide access at no cost: thus, their commercial model, like many other websites, relies on advertising to generate revenue.

To what extent might piracy websites be contributing to the contemporary challenges concerning the prevention of child exploitation in BRIC countries, and cybercrime proliferation more generally (including malware distribution)? Many websites, including piracy websites, use advertising as a revenue source to fund their operations [1]. They make available "space" on their pages (similar to

HICSS

traditional highway banners or billboards) for advertisers to promote their products and services. Historically, advertisers paid websites directly to "lease" this space, and they would pay revenue to the website based on the number of pageviews, or the number of users who clicked on the ad. Now, the buying and selling of website advertising space is very complex, as a number of intermediaries are involved – advertising networks take ad placement requests from advertisers, and display them across a range of websites, based on matching the interests of the advertiser and the user. Every user's interests can be tracked by the use of cookies, cross-referenced with other sites that they visit, or search terms that they enter. This enables ad networks that also operate search engines to implement "behavioural advertising". A further complication is that advertisers can now engage an ad exchange, which seeks to place advertising across multiple ad networks. Each of these developments has placed further distance between the advertiser and the website which is selected to host their advertising. This had led to greater marketplace efficiencies, but also to a loss of control by advertisers.

A number of recent studies have looked at the links between online advertising of various kinds, and how profits from displaying these ads fund and sustain piracy [2]. These studies have either focused on mainstream advertisers [3], or "high risk advertising", where the ads have the potential to harm the user, including malware [4]. The term "high risk advertising" was first introduced by [4].

High risk ad categories include malware, downloading sites, gambling, scams and the sex industry. Studies into high risk ads have been undertaken within many countries in the Asia Pacific, including Australia and New Zealand [5]. The results of these studies have been used by governments, NGOs and the private sector to guide appropriate policy responses, including regulation (whether by government or by the advertising industry itself), as well as building awareness around the risk of visiting these sites, especially by children.

In trying to understand the scale and scope of high risk advertising in high growth economies, a number of approaches have been developed. These aim to model how a user might encounter different types of

advertising or be exposed to CEM or links to CEM while searching for pirated material.

In executing the research, representative snapshots were constructed - typically comprising several thousand samples of the kinds of ads most likely to be served to users in a particular country. Two approaches for analysis were developed. The first focuses on the *most-complained about sites*. These are sites indexed by Google that rights holders have identified as being largely concerned with piracy, and where Google has investigated and found that they are in breach of the Digital Millennium Copyright Act (hereinafter DMCA). This provides independent verification[2] of the rogue nature of such sites, but it does not always follow that the most-complained about sites are always the most popular, particularly in local language sites. This is the approach utilised by Watters et al [5] for a New Zealand case study. The second approach focuses on the *most popular sites*. By using an information-seeking behaviour model, researchers can make use of the Google search engine to identify those sites which are the most relevant given a search term, such as "download free movies". Furthermore, a snowball sampling technique can be used to construct a set of terms for further searches, to identify more sites. Such data can also be supplemented through the use of site lists provided by experts, e.g. lists of sites which have been the subject of civil action or criminal prosecution.

By combining measures of popularity and alleged infringement, it is possible to arrive at an informed snapshot of high risk ad prevalence in a particular market or economy. In some countries, researchers have looked only at the most-complained about sites (especially in relation to Hollywood movies and TV), but this study *also* investigated the most popular sites when using the local language for searching.

In this paper, both types of data collection and analysis were utilised, to understand the advertising

---

[2] A number of studies, eg [12], have shown that Google sometimes processes DMCA mistakenly. According to its own documentation, Google removes 97% of the sites which are reported to it. Google also notifies the site owners (where it is able to) regarding the takedown, and provides an appeal process where false positives have been flagged for removal. An error rate of 3% is thought to be reasonable.

landscape, the mainstream and high risk aspects of that environment, and the links that exist between piracy sites and CEM.

1.1. Background

As one of the BRIC countries, Brazil is the eighth largest economy in the world as measured by GDP.[3] Total ad spend in Brazil is now US$20.64b,[4] with 14.7% of that being digital spend, representing an increase of 15% year-on-year. Mobile advertising is also growing rapidly, at 120% p.a., and now represents 18.1% of digital ad spending. Even a small fraction of this advertising market could be used to fund significant piracy operations, and make significant profits for organised crime [11].

Being a large and vibrant economy, Brazil has faced a number of challenges in policing intellectual property. A National Council against Piracy and Intellectual Property Crimes (CNCP) has the overall remit to promote intellectual property rights through education, enforcement and policy initiatives.[5] Additionally, concerns regarding online child exploitation and trafficking have frequently been expressed by Brazilian law enforcement, Government and NGOs [12]. In a recent operation (Operation Darknet), 55 people were arrested in Brazil for participating in a significant CEM "ring".[6] Six children were rescued from their abusers during the operation. This incident was not an isolated one, and points to a significant and emerging problem for Brazil, as well as many other countries globally. Whether the problem has existed for a long time, and is only now receiving investigative attention, or whether the internet is fuelling a rise in technology-enhance crime, systematic research is required to understand the scale and severity of the problem.

---

[3] http://tinyurl.com/osxljfm
[4] http://www.emarketer.com/Article/Total-Media-Ad-Spending-Brazil-Pass-20-Billion-This-Year/1012339
[5] http://www.wipo.int/wipo_magazine/en/2006/05/article_0003.html
[6] http://www.dailymail.co.uk/news/article-2795299/brazilian-police-crack-open-haul-child-pornography-dark-internet-rescue-six-children-abuse.html

In the following sections, separate studies targeting the most-complained about sites, the most popular sites, and links to child exploitation sites are explored from within Brazil. A disturbing snapshot emerges. Policy options which may help to ameliorate (or minimise) the harms due to advertising, in particular, are outlined. Firstly, however, the literature investigating the links between CEM and advertising is reviewed.

1.2. Advertising and Child Exploitation Material

As early as 1997, Burke [19] identified the economic motives involved in the commercial production of CEM, when discussing whether such advertising would enjoy First Amendment rights. Legal judgements in the US have not supported the view that CEM is a form of artistic expression that may be protected, and therefore legitimately be able to advertised. Even 20 years ago, it was clear that governments would need to consider regulatory measures to prevent the success of the CEM business model. In a criminological context, Wortley and Smallbone [20] first drew a link between the promotion of CEM and advertising, by identifying SPAM messages sent to potential customers through email. An alternative view was presented by Wolak et al [21], where advertising was not just seen a mechanism to elicit interest in CEM, but could potentially be utilized to deliver messaging around the risks and harms of CEM. This would be consistent with the "alerting the conscience" strategy predicted by situational crime prevention theory [15]. Advertising does not need to be limited to commercial, for-profit sites: much CEM is traded and "advertised" on P2P search sites, where it can be downloaded for free, facilitated by the appropriate choice of keywords [22]. Indeed, as shown by [8], searches for CEM on one of the most popular P2P search engines were more popular than Harry Potter movies during the time of their initial release.

The commercial model of P2P sites relies on advertising income, so understanding the links between CEM and advertising may provide further insight into how to disrupt the illicit trading of this material. A broader point has been made about the relationship between advertising and the sexualization of girls [23]: commercial entities use highly sexualized of pre-teen girls to promote non-CEM products and services routinely. Surprisingly, there has been no research systematically analyzing the direct linkages between CEM and advertising; it is likely that such advertising will likely appear in the "dark side" of the internet, rather than in plain sight.

2. Most Complained About Sites

The approach used in this study is based on algorithmic ethnography: broadly speaking, how to understand online culture by studying the outputs of algorithms, where we seek to understand how they operate in real-world contexts. In this setting, we seek to observe how online advertising algorithms operate to place CEM or malware ads on certain websites, and certain audiences, by testing a range of input parameters. These could include geographic location, search terms, browser types etc While ethnographic approaches are often qualitative, we have chosen to use a quantitative approach, to simply answer the question: if you were a naïve user, what kinds of materials would you be exposed to when using these sites, and with what frequency? While frequency is a simplistic measure, it does provide an indication of the categories of experience that a user may be exposed to.

To build a database of the sites most-complained about by rightsholders, the Google Transparency Report[7] was used. This report, among others, lists all of the URLs removed from the Google Index, whenever a complaint from a rightsholder is upheld, i.e., when Google determines that the URL does contain or links to allegedly pirated material.[8] In the United States, this process is governed by the Digital Millennium Copyright Act (DMCA). It is interesting to note that the number of complaints upheld by Google almost doubles every year. Note that even when Google chooses to remove a URL from its index, that site is still accessible through the Internet, and could also potentially be found by using other search engines.

Using this data, it is possible to identify the top domains against which complaints have been upheld. These sites range from torrent sites (which host links to pirated content), through to streaming sites, and file locker sites, where URLs for pirated content are made accessible through search engines, but physically downloaded from the locker site. In some cases, the same logical site is hosted on a number of different physical / top-level domains, in order to circumvent legal avenues for takedowns and DMCA complaints.

The algorithmic ethnography methodology used in this study was first outlined by [4]. The basic process is that a URL from each top infringing domain is downloaded a number of times (typically 10 page impressions), and the advertising networks and advertisers are identified. Furthermore, each advertisement is assigned a category (mainstream or high risk), and the high risk ads are further divided into a number of sub-categories: malware, sex industry, gambling, download site or scams. A semi-automated system has been developed to assist in processing the large volumes of data involved in this kind of analysis.

In this study, the Top 100 most-complained about websites (for the month of September 2015) identified by Google were analysed. A simulated user was created using a Virtual Private Network (VPN) connected to Brazil, such that each requested website (and its advertising networks) would "see" a Brazilian user, and content would be customized appropriately.

From the 1,000 pages sampled, a total of 709 advertising items were identified. Table 1 shows the number of advertising items. Note that the automated process for identifying advertising items relies on a database of known multinational advertising networks, and may not be reliable for all locations. This is why every page downloaded is manually analysed for visible ads.

A typical page on one of these sites might include a number of banner ads to the top, left and right of the page, with the main text instructing a user how to download a pirated movie occupying the mid-section. For example, on one search for the site extratorrent.work, with the search term "Disney", a link to a torrent containing a number of Disney children's movies is returned, and a preview of one movie is embedded in the page. Overlaid on this movie preview is an ad for the "Reality Kings" pornography website, with a penis depicted entering a woman's mouth. The

---

[7] http://www.google.com/transparencyreport/

[8] Note that Google only undertakes basis informational rather than legal checks (http://www.google.com/transparencyreport/removals/copyright/faq/). Google does not adjudicate for fair use, and webmasters may not be notified their pages are removed from search, and counter-notification processes may be beyond the means of many ordinary web users.

same search on thepiratebay.gd also provides a list of Disney torrents, but with further explicit images of nude and semi-nude women promoting pornography websites. Disney was selected as a search term because it is the brand that most parents would identify with as encapsulating childhood. Pirates also know this, and make available pirated versions of many of Disney's movies. There is no suggestion that Disney is in any way negligent or responsible for the behaviour of pirates or ad networks. Using Disney as a search term provides insights into ad networks that display pornography in response to searches for children's movies.

The point about "visible ads" may require some explanation. A great many ad networks are active on websites collecting data about browsing habits which they pass on or use to improve their own algorithms, without this being apparent to the browser. Herps et al [6] found that Google, for example, had trackers on 86% of the top sites accessed by Australians, and this pattern would be similar globally.

In this study, the actual number of visible ads was 1,402. Of these, some 1,315 were categorized as high risk (93.79%), while 87 were identified as mainstream (6.21%). These results are comparable to other countries where the most-complained about sites have been investigated. A further breakdown of high risk ads into different categories is shown in Table 2.

It can be noted that the most frequent high risk ads were found to belong to the sex industry category (45.15%), followed by download services (25.55%) and scams (13.69%).

Table 3 provides a list of the most prevalent advertised brands identified. These are provided to indicate the range of vertical industries that typically advertise on these sites.

Table 1 – Top Advertising Networks (Most Complained About)

Advertising Network	Items
ads.exoclick.com	103
main.exoclick.com	83
syndication.exoclick.com	83
ads.torrentco.com	80

waptrick.me	80
ad.zumads.com	30
ad.propellerads.com	30
track.adnetwork.vn	20
delivery.adnetwork.vn	20
www.adcash.com	20
unblockw.com	20
ads.clicksor.com	20
www.liveadoptimizer.com	20
ads.livedirsets.org	10
srv.juiceadv.com	10
ww.googletagservices.com	10
adserver.adreactor.com	10
cdn.popcash.net	10
stafaband.info	10
games.waptrick.me	10
iweb2.mangapicgallery.com	10
go.arbopl.bbelements.com	10
www.mangago.me	10

Table 2 – High Risk Ads Categories (Most Complained About)

Category	%
Sex	45.15
Malware	8.82
Download	25.55
Gambling	3.80
Scams	13.69

Table 3 – Most Advertised Mainstream Brands (Most Complained About)

Advertiser	N
Radiorage	12
Sling	10
Allin1Convert	10
Alibaba	10
Google Apps	8
imesh	3
ERV Travel	3
Amazon	1
CfS Medical	1
Weatherblink	1
Godaddy	1
Cadillac	1
Lifelock	1
Identity Guard	1
Paypal	1
Crowne Plaza	1
Marriott	1

VectorVest	1

.

3. Most Popular Sites

In this study, a user model was developed, to replicate how people search for pirated content by means of search engines within Brazil, *only* in the local language. The approach was based on the snowball sampling technique, where a Portuguese seed term was used to find piracy sites on Google Brazil. When a piracy site is identified, any new Portuguese terms associated with piracy were then added to the term list, until no further sites were found (up to a limit of 50). Ten impressions were made of each page.

A total of 2,088 advertising items were identified in this study. The advertising networks and other tools responsible are shown in Table 4. Once more major brands are listed. Yet in almost the reverse situation of the most-complained about sites, there were relatively few visible ads – 320 in total.

Also, the proportion of mainstream to high risk was reversed: a majority of ads in the most popular sites were mainstream (65.63%) versus 34.38% for high risk. Perhaps the most astonishing result was that there was no sex advertising at all on the local language sites, even though there was significant local language sex advertising on the most complained about sites. Table 5 provides a list of the mainstream advertisers identified in this study, and Table 6 provides a breakdown of the high risk ad categories.

Table 4 – Top Advertising Networks (Most Popular)

Advertising Networks	N
www.baixarcdstorrent.com.br	1540
i.imgur.com	50
adf.ly	48
goo.gl	40
www.baixaki.com.br	40
4.bp.blogspot.com	30
ads.egrana.com.br	25
cdn.adf.ly	20
humornerd.com	20
ad2.adsafiliados.com.br	20
zoomads.org	19
www.adcash.com	18
cdn.popcash.net	10

downloadfilmesgratis.org	10
www.cpmaffiliation.com	10
baixarfilmestorrents.org	10
www.comando-filmes.com	10
www.gospelparabaixar.com	10
baixarfilmesviatorrents.com	10
www.themediafire.com	10
scarytorrent.com.br	10
www.baixarcdstops.com	10
adprovider.adlure.net	10
www.downloadsfull.net	10
bdv.bidvertiser.com	10
1.bp.blogspot.com	10
www.doutormp3.com	10
www.bidvertiser.com	10
www.grandetorrent.com	10
www.animaniacos.org	10
ww.facebook.com	10
torrentgames.biz	10
www.vytorcds.com	10
anoxxx.com	8

Table 5 – Most Advertised Mainstream Brands (Most Popular)

Advertiser	N
Ford	60
Buscape	50
SEB COC	10
Brastemp	10
Xero	10
Netshoes	10
Nike	10
GVT	10
Dell	10
Jeep	10
Testlife	10
Zoom	10

Table 6 – High Risk Ads Categories (Most Popular)

Category	%
Sex	0.00
Malware	27.27
Download	45.45
Gambling	0.00
Scams	27.27

4. Links to Child Exploitation Material

Piracy websites are known to provide links to CEM [8]. In a longitudinal study, Prichard et al [9] found that the term "pthc" (pre-teen hardcore) was consistently more frequently searched for than Harry Potter movies on what was the then #2 most popular torrent site. Other studies have consistently linked piracy sites and CEM [9].

In the current two studies reported in the previous sections, some quite startling differences were observed: when the most complained-about sites were reviewed, the sex industry was the top high risk advertiser (and high risk advertising was the dominant category), yet for the most popular local language sites, there was no sex industry advertising. It is not clear why there should be such a difference: perhaps rights holders do not fully use the available intelligence to generate a very clear view of the sites which are actually making the most revenue and/or pose the greatest risk [26].

One particular category of sex industry website was quite prominent from this sample: manga, or Japanese comics. This category or genre of manga is known to depict simulated sexual acts between adults and children, between children or between children and animals. In many countries, such as the United States and the United Kingdom, obscenity laws cover the depiction of child sexual abuse in cartoon format; in a highly-publicised case, Christopher Handley, an American citizen, pleaded guilty to possessing manga books depicting child sex abuse, even though he possessed no other types of CEM[9]. He received a sentence of 6 months. Yet in Japan, while the possession of CEM has recently been criminalized, it only concerns materials where "real children are being depicted", therefore this does not cover CEM in manga.[10] The UN Special Rapporteur on the sale of children, child prostitution and child pornography recently urged the Japanese authorities to outlaw manga containing extreme child pornographic content.[11]

In the study of the most complained about sites, a number of manga piracy sites appear in the Top 100 (including mangapark.me, ranked #23). While many of the titles appear with non-CEM themes, mangapark.me hosts 487 "adult" titles, in which a number deal directly with CEM.

The prevalence of mainstream versus high risk advertising on the mangapark.me site was examined for the pages included in the sample, and 100% of the ads displayed were mainstream. Table 7 below shows the list of brands advertised.

Table 7 – Most Advertised Mainstream Brands on a CEM title on Mangapark.me

Advertiser	N
Teach.org	5
Ford	4
Legalzoom	3
Godaddy	2
Vectorvest	2
Microsoft	2
Dodge	2
Verizon	2
Fuelperks	2
Adobe	1
Cathay Pacific	1
Chevrolet	1
Wyndham Grand Resorts	1
Clear Care	1
Gerber Life	1
Olive Garden	1
UPS	1
Nissan	1
SmokeyBear.com	1
Zatarains	1
CBS All Access	1
Unopar	1
Westwing Home and Living	1
Salvation Army	1
Western Pest Services	1
Discovertheforest.org	1

---

[9] http://www.wired.com/2009/05/manga-porn/
[10]
http://www.aljazeera.com/indepth/opinion/2014/08/manga-anime-japan-still-treatin-201484145420634173.html

[11]
http://www.theguardian.com/world/2015/oct/27/japan-urged-to-ban-manga-child-abuse-images

The results indicate tentative links between piracy sites, CEM and advertising; these links are not exclusive, since advertising of some kind appears on all piracy sites. Consider the case of a CEM user who may use a search engine to identify torrents that contain CEM, and then visit the piracy site. On the piracy site, the user is able to download a torrent, or in some cases, view the material by clicking on a series of links to reveal a preview gallery. At each step, the user is exposed to a range of advertising which the advertising networks deem to be relevant to the user. In some cases, information from the search engine may be combined with referrer pages and other observable data linked through cookies to refine the relevance further. Also, once the user is on the piracy site, they can click on the username of someone sharing CEM material, and see other files that this person has uploaded. As files become more popular with users in terms of downloads, they then tend to rank higher in search results. Further research is required to whether these results are robust across different countries, languages, cultures etc

The unexpected link to CEM comes in the form of image previews, which users also tend to upload in order for other users to verify that they will be downloading CEM (since there are a large number of fake files on piracy sites. This means that mainstream ad networks placing ads on image hosting sites used for CEM previews may place ads alongside CEM images, especially when there is a "match" found between the interests of the user, and the advertiser.

To illustrate, one of the terms used by Fournier et al [9] in their analysis of the CEM peer-to-peer ecosystem, was used in this study to seed an image search on a popular search engine. This returned many results which contained links to piracy sites. The first of these was a link to torrenthound.com. On the results page, only sex industry advertising was present. Taking the name of the image sets, a set of results was also returned from thepiratebay.la. The advertising displayed here was also high risk – sex industry and malware downloads. Clicking through further to the detail page for the first torrent, again sex industry advertising. Note that in most cases, the advertising is localized for Brazilian users, in Portuguese.

Unexpectedly, while users could just click the magnet link or download the torrent to obtain the material, a link to an image hosting service is also supplied (imgbox.com). This allows the user to verify that they will be downloading CEM. Yet it is at this stage that mainstream advertising once again (unwittingly) becomes linked to CEM: mainstream ads are placed on the pages where image previews are made available.

Thus, mainstream advertisers are being placed alongside CEM. Two particular examples merit a mention: ads for Nissan motor vehicles and Jockey underwear both were found on this preview page for CEM material. Mainstream advertisers would clearly not want their brands to be damaged in this way, yet the advertising networks and image hosting services have become unsuspecting players in the CEM ecosystem. The images displayed appear to have been produced by a US corporation which lists its WHOIS registration data as being in Russia. The lack of verification or validation of WHOIS registration data is a serious and ongoing impediment to cybercrime investigations, as much of it is fake [10].

5. Discussion

This paper analysed piracy sites that are available to Brazilians, by simulating users searching for infringing content on either the most complained about sites (globally), or by searching using the country's official language on a local search engine, using algorithmic ethnography. While the most complained about sites had primarily high risk advertising, of which a majority was promoting the sex industry, the local language sites had mainly mainstream ads, with nothing displayed from the sex industry. A key question which needs to be answered by further research is whether Brazilians tend to use global piracy sites, or prefer to search for pirated content in Portuguese on local sites. Given the significant amount of sex industry advertising on piracy sites, it is likely that this may become a pathway for young people to become exposed to CEM [15].

The manga sites in the most complained about category only had mainstream advertisers, including a range of technology companies, car manufacturers and not-for-profit organisations. Hopefully, the realisation that their advertising spending is facilitating access to

CEM will ensure that such organisations pay closer attention to their marketing policies. This may be easier said than done: historically, advertisers simply leased ad space on a set of known websites, and when this approach didn't scale with the rise of internet advertising, advertising networks took over the role of bringing together advertisers and ad space. Yet, further complications have arisen over time: new entities, such as ad exchanges play a further intermediary role, allowing advertisers to purchase ad space targeted at quite specific demographics identified by behavioural profiles constructed by these exchanges. Ad networks, agencies and exchanges may be able to differentiate themselves in the marketplace by offering guarantees that major brands will not appear alongside pornography, gambling, or indeed, be used to support the distribution of CEM.

In terms of policy responses, there are a number of options. From a harm minimisation perspective, given that the most complained about sites contain mostly high risk ads, approaches such as regulatory site blocking are likely to have an impact. A potential risk from this strategy is that users are pushed further towards the fringes of the internet, leaving them more exposed to advertising which may be even more harmful.

Another approach could be to have some sort of reporting mechanism through which the advertisers could be made aware that their ads are being placed on CEM sites, so that they could take direct action. As a result, their marketing teams would have to give a public explanation as well as identify the advertising networks who actually supply the ads. If the advertising networks withdrew from placing ads on these sites, then the sites would have no revenue to fund their illicit operations, and might be forced to close. This approach has been pioneered in the UK, where the City of London Police maintain a list of known illegal sites, known as the Infringing Website List (IWL), that advertising firms can use to ensure that advertising is not sent to those sites[12]. It has resulted in a reduction of mainstream advertising on piracy sites when viewed by British users.

6. Conclusion

This paper points to ways in which piracy websites are being nurtured by revenues from both mainstream and high risk advertisements and how these same sites – which are known to be popular with children and young people – are thereby exposing the latter to materials which are likely to be extremely damaging to them as well as to adults.

The Brazilian advertising industry has a major responsibility to act against local advertising networks which are placing ads on known piracy websites. If self-regulation is failing, then the Federal Government should step in. Further research will be needed to see whether similar patterns are observed in other rapidly developing economies, and/or to see how local results may compare globally. Furthermore, only two possible categories of sites have been analysed – the most complained about, and the most popular – yet there are other possible categories, such as quality and recency – which may be as salient. One can imagine – particularly on the dark web – that subscribers to niche interest sites may be low frequency, but high in monetary value, and high in potential harm to victims.

The top ad networks identified by undertaking this analysis have high potential economic impact: exoclick, for example, is the fourth largest advertising network in the world, with more than 5 million banners being displayed daily[13].

In addition the brands themselves need to take more responsibility for ensuring that their advertising spend does not unintentionally support sites which pose these type of threats to children and young people. How ironic that advertisers like Teach.org – setup to promote the teaching of young people – should be associated with a CEM title on a piracy site.

The difference between the types of advertising appearing on the most complained about sites and the sites which are most popular with Brazilians seems to suggest a disparity in the level of scrutiny and attention which ought to be addressed.

---

[13]

https://w3techs.com/technologies/overview/advertising/all

---

[12] http://www.bbc.com/news/technology-26788800

Bearing this in mind, could a system be established which regularly monitors the advertisements appearing on piracy websites in Brazil and provides timely advice both to the brands and to the advertising networks?. This would be consistent with situational crime prevention strategies [15], such as increasing the effort, and reducing the rewards. Given the amount of detailed information available about the owners and operators of CEM sites [16], further co-ordination between law enforcement, lawmakers, NGOs and the community will be essential to set standards for combating CEM online [17, 18] – and particularly the links to advertising – in the future.

10. References

[1] Garz, M., Rott, A., and Wass von Czege, M. (2015), "The Online Market for Illegal Copies of Magazines: A German Case Study", *Journal of Broadcasting & Electronic Media*, *59*(1), 169-183.

[2] Watters, P. (2015), "Censorship is ~~futile~~ possible but difficult: A study in algorithmic ethnography", *First Monday*, *20*(1).

[3] Taplin, J. (2013), "USC Annenberg Lab Ad Transparency Report", downloaded from http://www.annenberglab.com/sites/default/files/uploads/USCAnnenbergLab_AdReport_Jan2013.pdf

[4] Watters, P.A. (2014), "A Systematic Approach to Measuring Advertising Transparency Online: An Australian Case Study", in *Proceedings of the Second Australasian Web Conference: Second Australasian Web Conference,* Vol. 155, eds. Cranefield S., Trotman, A., and Yang, J., 59-67.

[5] Watters, P. A., Watters, M., & Ziegler, J. (2015). Malicious Advertising and Music Piracy: A New Zealand Case Study. In *Proceedings of the Fifth Cybercrime and Trustworthy Computing Conference (CTC)*, pp. 22-29

[6] Herps, A., Watters, P., and Pineda-Villavicencio, G. (2013), "Measuring Surveillance in Online Advertising: A Big Data Approach", *Cybercrime and Trustworthy Computing Workshop (CTC), 2013 Fourth*, IEEE, 30-35.

[7] Fournier, R., Cholez, T., Latapy, M., Chrisment, I., Magnien, C., Festor, O., and Daniloff, I. (2014), "Comparing pedophile activity in different P2P systems", *Social Sciences*, *3*(3), 314-325

[8] Prichard, J., Spiranovic, C., Watters, P., and Lueg, C. (2013), "Young people, child pornography, and subcultural norms on the Internet", *Journal of the American Society for Information Science and Technology*, *64*(5), 992-1000

[9] Rutgaizer, M., Shavitt, Y., Vertman, O., and Zilberman, N. (2012), "Detecting pedophile activity in bittorrent networks", *Passive and Active Measurement*, Berlin Heidelberg: Springer, 106-115

[10] Watters, P., Herps, A., Layton, R., and McCombie, S. (2013b), "ICANN or ICANT: Is WHOIS an enabler of cybercrime?", *Cybercrime and Trustworthy Computing Workshop (CTC), 2013 Fourth*, IEEE, 44-49.

[11] Treverton, G. F. (2009), "Film piracy, organized crime, and terrorism", Rand Corporation, vol. 742.

[12] Urban, J. M., & Quilter, L. (2006). Efficient Process or'Chilling Effects'? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act. *Santa Clara Computer and High Technology Law Journal*, *22*, 621.

[13] Hazeu, M. and van Kranen, F. (2014), "Sexual Exploitation of Children in Brazil: Putting a Spot on the Problem", Terre des Hommes Netherlands, ECPAT Netherlands, Plan Netherlands, Free a Girl, downloaded from https://www.terredeshommes.nl/sites/tdh/files/uploads/140509_sexual_exploitation_of_children_in_brazil_putting_a_spot_on_the_problem.pdf

[14] Ferraro, M.M., and Casey, E. (2004), "Investigating child exploitation and pornography: The internet, law and forensic science", Academic Press

[15] Clarke, R.V.G. (ed.) (1997), "Situational crime prevention", Monsey, New York: Criminal Justice Press, 53-70

[16] Watters, P. A., Lueg, C., Spiranovic, C., and Prichard, J. (2013a), "Patterns of ownership of child model sites: Profiling the profiteers and consumers of child exploitation material", *First Monday*, *18*(2).

[17] Quayle, E. (2017). Over the Internet, Under the Radar: Online Child Sexual Abuse and Exploitation–a brief literature review. http://www.cycj.org.uk/resource/over-the-internet-under-the-radar-online-child-sexual-abuse-and-exploitation-a-brief-literature-review/

[18] Taylor, M. & Quayle, E. (2008). Criminogenic qualities of the internet in the collection and distribution of abuse images of children. *The Irish Journal of Psychology, 29(1–2),* 119–130.

[19] Adler, A. (2001). The perverse law of child pornography. *Columbia Law Review*, 209-273.

[20] Wortley, R. K., & Smallbone, S. (2006). *Child pornography on the internet*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

[21] Wolak, J., Finkelhor, D., & Mitchell, K. J. (2005). Child-Pornography Possessors Arrested in Internet-Related Crimes: Findings From the National Juvenile Online Victimization Study.

[22] Steel, C. M. (2009). Child pornography in peer-to-peer networks. *Child Abuse & Neglect*, *33*(8), 560-568.

[23] Merskin, D. (2004). Reviving Lolita? A media literacy examination of sexual portrayals of girls in fashion advertising. *American Behavioral Scientist*, *48*(1), 119-129.

[24] Laws, D. R., & Marshall, W. L. (1990). A conditioning theory of the etiology and maintenance of deviant sexual preference and behavior. In *Handbook of sexual assault* (pp. 209-229). Springer US.

[25] Treverton, G. F. (2009). *Film piracy, organized crime, and terrorism* (Vol. 742). Rand Corporation.

[26] Lee, S.J. & Watters, P.A. (2017). Cyber budget optimization through security event clustering. *Proceedings of the 7th IEEE International CYBER Conference,* Hawaii, HI.