How Risky Are Piracy Websites For Indonesians? The Role of Search

Paul A. Watters

Massey University

Executive Summary

Recent research indicates that users of rogue websites face significant risks, since much of the advertising has the potential to cause harm to users. This type of advertising is called “high risk”, and includes ads from the sex industry, gambling, malware and scams. Some studies have focused on naming and shaming mainstream advertisers on rogue sites (Taplin, 2013), and there are significant numbers of these advertisers in Asia, placing ads through ad networks that focus on local language and local content. A number of studies have recently investigated the role played by mainstream internet advertising in supporting the revenue of rogue websites. Such advertising by household names – including multinational corporations, governments and charities - generates enormous profit margins for operators of these websites.

However, the advertisements which have the potential to cause the greatest harm are also those which are forbidden under Islamic jurisprudence. While vigilance against these harms is often promoted within the physical environment in Indonesia (eg, Jakarta Post, 2014), policing the online environment to protect citizens from *haram* activities (Assyaukanie , 2009) may be more challenging.

A first step in understanding the risk to users is to measure the prevalence of such material. Recent studies have measured the scale of the problem in the Asia-Pacific region, including Singapore, Malaysia, Hong Kong, Taiwan, Canada, Australia and New Zealand. Generally speaking, the pattern seems to be that high risk advertisers tend to target those sites which are promoting Hollywood content in English, whereas mainstream advertisers tend to promote their goods and services on local language sites, with local content.

The goal of this study was to measure the problem in Indonesia. A sample of 1,000 webpages from rogue sites was downloaded and analysed, to reveal the prevalence of both mainstream and high-risk advertising. The sample was derived from expert opinions as well as the use of a snowball technique to generate a list of sites using vocabulary terms linked to illegal downloading. Snowball sampling replicates the path taken by users as they search for infringing content by using search engines. It also acts as a proxy measure for popularity, since the most “relevant” search terms are always returned by ranking.

Analysis of the data reveals that 16.05% of the ads were mainstream, and 83.95% were high-risk, of which 75% were gambling ads. The level of gambling advertising was unprecedented and not observed in any other country. The results are discussed in the context of policing the online environment with regard to Indonesian social policy.

Keywords

Infringing content, internet advertising, Digital Millenium Copyright Act (DMCA), internet safety.

Definitions

Internet Advertising. Ads are typically placed as "banners" on a website, which direct a user to another site when clicked. The contents of the ad are similar to a highway billboard, except that they can incorporate interactive elements such as animation. Ads on the same page are often rotated through a predetermined or random sequence, depending on the advertising plan that an advertiser has subscribed to. While some sites host and manage their own banners, most often, these are managed by a third-party advertising network. These ad networks act as an intermediary between an advertiser and many hundreds, thousands or millions of sites, allowing an advertiser to increase their reach to potential consumers while only dealing with a single agency. Advertisers typically operate either a "pay per impression" or "pay per click" model, billing an advertiser every time a user views or clicks on a banner ad respectively.

Mainstream Advertising. Mainstream ads are those placed by legitimate businesses that operate within the formal economy. Such businesses operate through a corporate structure and offer goods or services which fall outside the black market, grey market or underground economy.

High-Risk Advertising. High-Risk ads are those promoting goods or services which fall outside the legitimate economy or white market, may be illegal or restricted within certain jurisdictions but not others, or may be fake or counterfeit. Examples include the sex industry, gambling and suspicious software/malware, such as anti-virus software which actually installs a Trojan Horse on a user's system. Many of the ads are likely to fall into scam categories described by Stabek et al (2009).

Advertising Network. Ad networks facilitate the placement of an advertiser's ads on numerous websites according to a specific revenue model. Ad networks specialise in anticipating consumer's needs and wants by building up profiles of users who click most frequently on certain ad categories on certain page themes, which can lead to more targeted, personalised, and relevant advertising. For the purposes of this paper, sites that host advertising on behalf of external / third-party advertisers are also grouped under this category, even if they only provide banners on sites within their own domain. For example, isohunt.com provides their own ad network exclusively for their own site, and not to other sites; they also host banners from other ad networks.

Internet Advertiser. A business, government, association or individual that desires to sell goods or services, or provide information to, a target group of consumers. Internet advertising competes with traditional advertising for marketing budgets. Indonesia's online user base was 71.19m in 2013[1], with US$25 spent on each user[2], making the approximate advertising market value of US$ 1.779b. The market is growing at a rate of 75% pa[3], suggesting current revenue of around US$3.11b[4] at the end of 2014.

[1] http://www.thejakartapost.com/news/2014/01/15/number-ri-internet-users-increases-7119-million-2013-apjii.html
[2] http://www.statista.com/statistics/242843/digital-ad-spending-per-internet-user-in-indonesia/
[3] http://www.iab.net/media/file/GlobalSummit2013-Presentations/Insights_on_Digital_Growth_and_Opportunities_from_the_Global_IAB_Network-Asia_Pacific.pdf
[4] http://read.pwc.com/i/179445/12

Rogue Site. A website which provides an index and search capability for torrents of infringing content, a "file locker" site which provides hosting for such material, or a "link site" which provides direct links to content on third party sites. The primary motivation for users visiting these websites is to access infringing content. These sites can all use advertising as either primary or secondary sources of income.

Digital Millenium Copyright Act (DMCA). The DMCA provides US ISPs with indemnity against liability for copyright infringement, provided that they agree to co-operate in "takedowns" of material which is alleged to be infringing, typically after being notified by a rightsholders. Google provides a report of requests that they have received and actioned on behalf of rightsholders in order to provide transparency to their users.

Introduction

Online advertising has a 20 year long history (Medoff, 2000), progressing from simple ad banners displayed on a fixed rotation schedule, through to personalised, behavioural advertising networks, which use profiles of individual users to present the most "relevant" advertisements (McStay, 2011). Such technologies make extensive use of "tracking cookies" (Watters, 2012) and the linkages between advertising networks and cookies have recently been monitored and explored for the most popular websites in Australia (Herps et al, 2014). The most interesting result from this study was that the number of cookies stored on a user's computer from any of the Top 50 most-visited sites for Australians ranged between 0 and 86. The sophistication and the extent to which user behaviour is tracked and experiences customised is only going to increase over time, as is the overall volume of advertising. Indeed, in 2012, online advertising spending in the US reached US$39.6b, exceeding the amount spent on traditional print advertising for the first time (eMarketer, 2012), and predicted to reach $44.74b for 2013 and $51.01b in 2014.

Furthermore, some companies are in a unique position to know "everything" about their customers. Google, for example, has the capacity to monitor almost all of the world's information, including personal emails, YouTube movies, Android phones, news services, images, shopping, blogs and so on (Cleland, 2013). Through its acquisition of Doubleclick, Google controlled an estimated 69% of the online advertising market (Browser Media, 2008), however, the rise of social media advertising (especially through Facebook) has seen this reduce to 56% (Womack, 2013). Clearly, there is a potential confluence of capability and opportunity to maximise the number of "eyeballs" exposed to online ads.

What are the implications of this massive rise in advertising expenditure, which coincides with an increased ability for online advertising networks to be able to best "place" ads to suit specific customers? One particular type of website – those associated with file sharing of infringing content – appears to have wholeheartedly embraced advertising. Indeed, advertising revenues provide the commercial motivation for criminal syndicates to operate such 'rogue' web sites. While the connection between film and television piracy and organised crime has been explored elsewhere, in terms of direct revenues (Treverton et al, 2009), there has been far less publicity about the advertising revenues generated from sites that appear to offer infringing content for free, or at least, offer torrents that enable users to download such material. Certainly, the links between the

underground economy and the internet have been criticised for facilitating sexual exploitation and human trafficking through organised crime – in the classic paper in this field, Hughes (2000) highlighted how global advertising and marketing of prostitution have led to increases in volume globally. Furthermore, Hughes identified that a *lack of regulation of internet advertising* was the key policy failure in preventing harm to women and children.

The Pirate Bay is one of the most popular sites for providing torrents to infringing content, and has been the subject of criminal proceedings against its operators in Sweden. In the 2009 trial of its operators, their expenses were estimated to be US$110,000 p.a (Olsson, 2006; Kuprianko, 2009), with advertising revenues in the order of US$1.4m p.a (Sundberg, 2009) – in other words, an extremely profitable business with gross margins of 1272%! A recent study (Detica, 2012) indicated that there are six different business models operating within the pirate site marketplace, ranging from advertisement and donation funding, through to subscriptions and freemium sites, where subscribers can gain faster access to illicit content by paying a subscription fee. 83% of the sites in that study operated using a central website. Selling advertising on file locker and torrent search sites is the major source of revenue for such sites. The Pirate Bay, for example, regularly features in the Top 50 sites (as computed by alexa.com) , and so it is a potentially attractive space for advertisers and ad networks, since the number of potential "eyeballs" is very high. Other rogue sites with high Alexa rankings include Kickass torrents (rank 103) and Torrentz (rank 153)[5].

Maximising "eyeballs" leads to clicking, which drives revenue for the ad networks (if they operate a Pay Per Click revenue model), and sales for the advertisers. A key question for advertisers and ad networks is the extent to which they wish to be associated with this type of activity; indeed, due to the complex algorithms which decide which ads to display to which users, advertisers may not be aware of every site that their ads are being displayed on.

Being able to quantify the scale of advertising on these sites is important, since informing and making advertisers aware of the integrity of the sites on which their ads are being displayed can then be undertaken. Advertisers will thus be able to make more informed choices about their use of online advertising networks (the companies who provide aggregation of space on web sites) who are supporting piracy by selling ad space on torrent and file locker sites. A recent set of best practice guidelines for ad networks to address piracy and counterfeiting have recently been released[6], and early indications are that most of the world's major web companies will participate[7].

There have been few systematic studies investigating the relationship between piracy and advertising, and most have been concerned with the impact of interventions to reduce piracy. For example, Sheehan et al (submitted) identified that increasing the perception of legal risk for college students was most likely to influence downloading behaviour, while Gopal et al (2009) weighed up the ethical predispositions of downloaders and their beliefs in justice and law to the money potentially saved by downloading infringing content. Indeed, it is this appeal to justice as the

[5] http://au.ibtimes.com/articles/533033/20140106/pirate-bay-popular-torrent-site-top-10.htm#.Uvk8g_mSw3I
[6] http://2013ippractices.com/bestpracticesguidelinesforadnetworkstoaddresspiracyandcounterfeiting.html
[7] http://torrentfreak.com/tech-giants-sign-deal-to-ban-advertising-on-pirate-websites-130715/

primary virtue of social behaviour (Rawls, 1999) that may concern ethical advertisers if their advertising expenditure was being used to fund illicit activities.

Recent studies have indicated users were much more likely to be exposed to "high risk" advertising on such sites, relative to mainstream ads. Australia (Watters, 2014a), Singapore (Watters, 2013a), Canada (Watters, 2013b), and New Zealand (Watters, 2014c) all had mainstream prevalence rates of 1-10%, while high risk ads had prevalence rates of 90-99%. High risk ads are those which have the potential to cause harm to users, and include pornography, gambling, malware and scams.

These studies all investigated web pages that were sampled from Google's ad transparency report for movies and TV shows or music downloads, having been verified as being in breach of the Digital Millennium Copyright Act (DMCA). However, the Google report is heavily biased towards Hollywood TV/movies and music in English, so another study (Watters, 2014d) investigated Hollywood content in Taiwan, where the sites were presented in Chinese. It was found that mainstream advertising was much more prevalent for locally-developed sites with Hollywood content, compared to viewing Hollywood content sites in the other countries examined. 61% of ads were Mainstream, while 39% were High Risk for local content. A follow-up study in Hong Kong (Watters, 2014e) investigated whether Mainstream advertising would also be more prevalent not just for sites written in the local language, but also promoting local content (eg, a Chinese language website providing links to Chinese language titles. For local content, 61.36% of movie and TV ads were Mainstream, while 38.64% were High Risk. In contrast, 3.84% of Hollywood ads were Mainstream, while 96.16% were High Risk.

A subsequent study directly measured prevalence rates for High Risk versus Mainstream ads for local versus Hollywood content for movies/TV in Malaysia (Watters, 2014e). A sample of expert-identified rogue sites hosting local content was identified, and all ad banners comprising part of the sample were downloaded and identified, along with the ad network serving each banner. A comparable sample was taken from Google's ad transparency report. It was found that for local Malay content, 72.1% of movie and TV ads were Mainstream, while 27.9% were High Risk. In contrast, 8.24% of Hollywood ads were Mainstream, while 91.76% were High Risk. Like Hong Kong and Taiwan, this suggests that Mainstream advertisers in Asia are being drawn to local language sites, whereas Mainstream advertising rates for Hollywood titles are similar to other countries.

The first peer reviewed paper in this field was published by Watters (2014a). That study outlined a fully replicable algorithm for sampling rogue sites to provide a much clearer view of advertising network behaviour in different countries, jurisdictions, languages etc. The major difference between this study and the USC study was that it examined all advertisements, not just the Mainstream ones. In doing so, Watters was able to establish the relative proportion and prevalence of Mainstream ads versus High Risk ads.

In summary, local content sites were many times more likely to be displaying mainstream ads when compared to Hollywood content sites. The levels of mainstream advertising were almost identical to Taiwan and Hong Kong for local content, and were similar to Canada, Singapore, Australia and New Zealand for Hollywood content. In this study, local rogue websites in Indonesia are the focus, and it is predicted that the prevalence of mainstream and high risk ads for local content will be consistent with the results from Malaysia and Hong Kong.

Methods

A sampling procedure can be used to identify a representative subset of URLs, and the advertisements on each page can be downloaded along with their metadata. In the case of simple banner ads, it is then relatively easy to identify the advertisers concerned; in the case of each distinct advertisement, a rule can be generated using SQL or similar to identify all advertisements with the same metadata. However, some advertising networks use JavaScript obfuscation and a series of redirects to obscure the ultimate destination for the advertising banner; in this case, manual inspection must be performed, in the absence of a general purpose image/logo recognition system. The overall prevalence of a particular advertiser on each network can be then be computed and ordered by frequency.

Furthermore, it may be of interest to separate out "Mainstream" advertisements as opposed to "High-Risk" advertising, since the Annenberg reports indicate a flight by Mainstream advertising this year from sites that host infringing content. Advertisers who may otherwise be unable to place their ads on a Mainstream site can then take advantage of increasing "eyeballs" by occupying display space. Results are thus reported for the High-Risk and Mainstream categories, with the former including categories such as:

- Sex Industry, which includes adverts for:
    - Penis length extension medication
    - Fake personal/dating sites
    - Pornography of various kinds
    - Dating and "foreign bride" sites
- Online Gambling
- Malware, including
    - Fake software incorporating Trojan horse malware (numerous alerts were raised by anti-virus software during the data collection process due to "drive by downloads" of malware)
    - Fake anti-virus or anti-scamware
    - Suspicious software such as fake video codecs or video players that replicate existing functions within Microsoft Windows. The purpose of such downloads is unclear, although it is possible that they could host Trojans or provide backdoor access to systems.
- Scams, as defined by Stabek et al (2010), such as:
    - Premium rate SMS scams
    - Fake competitions where no prizes are offered
    - Investment scams
    - Employment scams

Two sources of data were used to source rogue sites for this study:

- An expert-derived list of known rogue sites which had been monitored for infringing content and/or had been issued takedown notices; and

- A list of sites generated from a snowball sample, used to identify the *most popular* rogue sites matching an initial "seed" term. The goal of snowball sampling is to build a sample that is often hidden or hard to measure directly. In the case of infringing content, there is no central database of sites that is readily available, so instead, we seek to replicate user behaviour in searching for such sites. The role of search engines here is critical – they provide a set of search results that act as a proxy for popularity based on the relevance of those results. Thus, by entering terms into the engine, and building a term list (i.e., those terms relevant to illegal downloading) and a site database (of infringing sites used within Indonesia), we can replicate the path that users take in discovering infringing content.
- To generate the snowball sample, Google Translate was used to identify "mendownload film" as the best seed (meaning "download film"). This term was entered into the term database, and the top ten sites from Google Search were then identified and added to the site database. Bahasa terms relating to illegal downloading were then identified from the text of the pages in the site database. This resulted in the terms terbaru (latest), gratis (free), baru (new), situs (site), DENGAN TERJEMAHAN (with translation), cara (method), bioskop (cinema) being added to the site database. Further Google searches were performed with these terms, and any rogue sites identified were added to the site database, until 100 sites were registered. A set of screenshots depicting the sequence of activities is shown in Appendix C.

The data was collected as follows:

1. A data collection system is installed physically or logically to attract advertising for a specific geographical/country segment. For this study, Indonesia was selected.
2. The sample was generated according to the algorithm described above.
3. For each site in the sample, 10 page impressions were taken, as banner ads often change between refreshes. This gives a total of 1,000 web pages to be downloaded (the sample) for local content.
4. Each of the 1,000 web pages in the sample is downloaded, and a screenshot is taken, showing the ads being served. Note that for technical reasons, pop-up ads are not captured.
5. For each web page in the sample, the code blocks that contain advertising are parsed and extracted. This can be achieved by matching against the Easy List[8] (used by Adblock Plus for filtering), for known URL patterns and hostnames of advertisers. Some pages in the sample will have no ads, while others will have multiple ads.
6. For each advertising code block, the domain of the advertising network being used is identified, by stripping extraneous code and links from the code block, and counting the frequency of appearance of each ad network domain. If an ad network has fewer than 5 occurrences, the items are discarded. The rationale for exclusion is that errors in coding, extraneous links etc can result in false positives being included in the list.
7. For each identified advertisement, an attempt is made to identify the actual advertiser, by analysing metadata, following the link and extracting the domain of the actual

[8] http://easylist.adblockplus.org/en/

advertiser, or through visual inspection. A list of all identified advertisers is then generated.

8. The category of the advertisement is then determined by an expert as either being high risk or mainstream.

Results

Appendix A contains a list of the sites identified in Step 2. From the 1,000 pages analysed in Step 3, a total of 2,585 advertising items and 2,175 visible ads were identified in Step 5[9]. Postprocessing of the identified domains were performed to ensure that all ad blocks were correctly identified, for example, by removing port numbers that were included as part of a URL by using a regex filter. 46 unique domains for advertising networks were identified, indicating an average 47.28 visible ads per network in the sample. Appendix B contains the complete list of advertising networks detected. Note that no merging of distinct services was performed, eg, the several domains of The Pirate Bay were not aggregated, to preserve the literal domains as observed. Also, where a domain appears within an ad block, this is a technical definition as per the methodology in Steps 6 and 7, ie, if the site or known ad URL appears in the block, then it will be counted. This could include Facebook social plugins, for example, rather than Facebook ads.

The analysis is presented by reviewing the High Risk ads first, followed by the Mainstream ads.

Table 1. Frequency Analysis by Advertising Network – Top 10 [10]

Advertising Network	**N**	**%**
tujuh.co.id	270	12.41%
yllix.com	264	12.14%
innity.net	260	11.95%
bidvertiser.com	230	10.57%
adplus.co.id	207	9.52%
ads.ec	148	6.80%
popcash.net	120	5.52%
adsensecamp.com	120	5.52%

[9] Advertising items include any scripts, images, spacers etc being referenced from an Adblock domain, in addition to visible ads

[10] Note that some ad networks like isohunt.com and sumotorrent.com do not display their ads outside their own domain; they are ranked highly because of the high number of DMCA complaints against their site.

ad4game.com	90	4.14%
maingames.co.id	80	3.68%

High-Risk Advertising – Top 10

There were 1,789 advertising items in this sub-sample. Each of these advertisements was downloaded, visually inspected and categorised. Mainstream advertisements were excluded, and only unique advertising were then attempted to be downloaded. The results indicate that gambling, malware, scams (including employment, investment and SMS premium rate), and the sex industry were the most popular distinct advertising types in Indonesia for the Top 10 networks. Note that some of these ad networks may display predominantly mainstream ads, or that they may employ obfuscation to try and block identification.

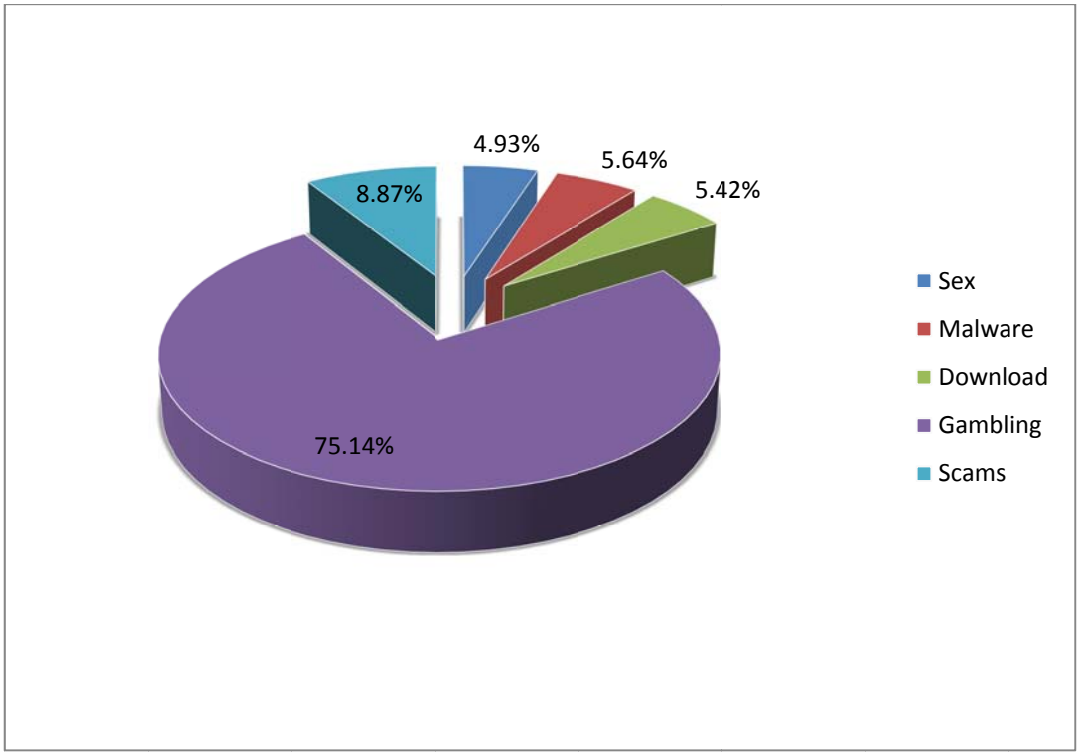
High-Risk Advertising

Table 2 shows the breakdown of the most common ad categories for High Risk ads across all networks. Each advertisement was downloaded, visually inspected and categorised. The results indicate that gambling, the sex industry, malware, downloading sites or scams (including employment, investment and SMS premium rate) were the most popular distinct advertising types. The categories are summarised in Figure 1.

Table 2. Frequency by Ad Category – High Risk Ads

	Sex	**Malware**	**Download**	**Gambling**	**Scams**
N	90	103	99	1372	162
%	4.93%	5.64%	5.42%	75.14%	8.87%

Figure 1. High-Risk advertising


4.93%
5.64%
5.42%
8.87%
75.14%
Sex
Malware
Download
Gambling
Scams

Mainstream Advertising- All Sites

Table 3 contains the results of the step 8 results obtained by visually inspecting every advertisement in the sample to identify whether it contained any Mainstream advertising. Typically, a rogue site will have 3-4 ad panels, and in many cases, the ads were tailored to the local geographic context. In some cases, advertisements were blocked with an image stating the site was "blocked" for offshore users, indicating further evidence of geographic customisation for the advertising content. In some cases, domains associated with file sharing were "parked" and advertising displayed, even if no infringing content was actually displayed – especially where such sites had terms like "warez", "anon" and "rapidshare" in their domain name.

Only 16.05% of the ads sampled consistently showed evidence that Mainstream advertisers are targeting Indonesian users via rogue piracy sites. Some ads and/or advertisers were only detected once. In a sense, this represents a type of leakage, since the Mainstream ads were a minority of the overall ads displayed (which were overwhelmingly High-Risk). A breakdown by industry category is shown in Figure 2, and the relative composition of Mainstream to High-Risk ads is shown in Figure 3.

Table 3– Mainstream Advertisers Detected

Advertiser	**N**
iklanpop.com	40
indosniper.com	30
cloudiro.com	30
iflazz.com	30
Air Asia	20
tiket.com	18
Yahoo.com	18
ANZ Bank	18
Givology	14
fukuichi clothes	10
Okane Clothes	10
Kolaski Clothes	10
bunglononline.com	10
gelartikar.com	10
ikofx.com	10
Godaddy	10
hostgator	10
bluehost	10
namecheap	10
adopuskids	8
vip.com	5
jd.com	3
red cross	3
Smokey the Bear	3
Discover The Forest	3

mudah.my	3
family violence prevention	2
Pick Gig	1
cahayavisual.com	1

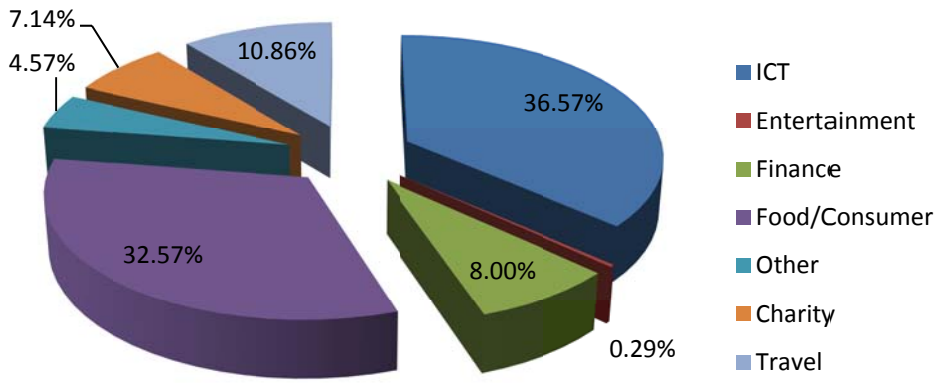


Figure 2. Mainstream advertising

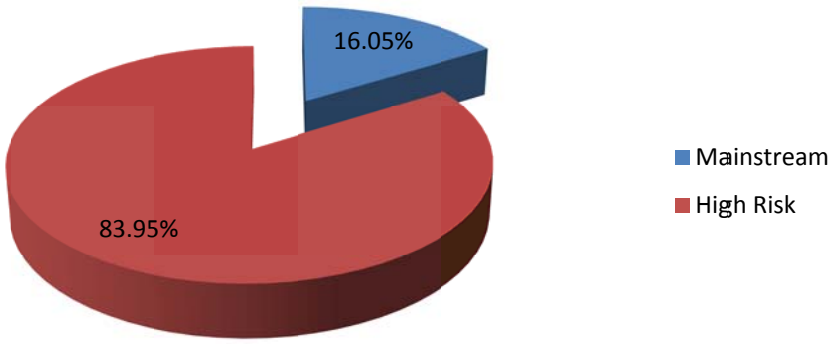


Figure 3. Mainstream versus High-Risk Advertising

Mainstream Advertising – Top 10 Ads

Table 4 shows the frequency distribution for the ten most frequently detected Mainstream ads. The key difference to note is the absence of Google ads for this sample and geographic location compared to the original Australian sample, where 87% of the Mainstream ads were served up by Google ads (Watters, 2013). This may be due to local advertising conditions, network restrictions (eg, blocking of certain ad networks) or a reduction in placement of Google ads onto rogue sites as a matter of policy. As some ad networks increase or decrease their presence on rogue sites, other ad networks often move in to fill the void, resulting in a type of displacement. Criminological theory suggests that displacement does not necessarily always result in negative outcomes. For example, if a more serious crime type is displaced by a less harmful type, then displacement can be positive (Felson & Clarke, 1998).

Table 4– Mainstream Advertisers Detected (Top 10)

Advertiser	**Site**	**Ad Network**	**%**
iklanpop.com	www.animekompi.web.id	ads.ec	11.46%
indosniper.com	www.icinema3satu.com	onclickads.com	8.60%
cloudiro.com	www.icinema3satu.com	onclickads.com	8.60%
iflazz.com	www.icinema3satu.com	onclickads.com	8.60%
Air Asia	workshop-movie.blogspot.com	idblognetwork.com	5.73%
tiket.com	forum.indowebster.com	tujuh.co.id	5.16%
Yahoo.com	forum.indowebster.com	tujuh.co.id	5.16%
ANZ Bank	forum.indowebster.com	tujuh.co.id	5.16%
Givology	dnacrewz.fantastique.tv	adnxs.com	4.01%
fukuichi clothes	www.icinema3satu.com	onclickads.com	2.87%

Conclusion

In contrast to other studies of local content on rogue sites in Asia, this study found that the majority of advertising on Indonesian sites was high risk in nature. In particular, 75% of high risk ads were for gambling. While the Taiwan and Malaysian studies found that more mainstream advertising was found for local content (Chinese or Malay), this study levels of mainstream advertising on rogue sites promoting comparable with other countries including Australia, New Zealand, Canada, and Singapore

The key findings from the analysis of this first-ever Indonesian data set are discussed below:

- For local content, 16.05% of movie and TV ads were Mainstream, while 83.95% were High Risk.

- Advertising for the gambling industry comprised 75% of the high risk ads. Sex industry ads, malware and scams were also common.
- The top ad networks serving ads to Indonesian residents for local content included tujuh.co.id, yllix.com, innity.net, bidvertiser.com, adplus.co.id, ads.ec, popcash.net, adsensecamp.com, ad4game.com and maingames.co.id.
- A number of household name brands in Indonesia are choosing to advertise on sites and their pages which are promoting the distribution of local infringing content. Further investigation is needed to uncover the mechanics of how these ads are selected to appear; are advertisers engaging directly with ad networks, or are ad networks operating at a wholesale level and distributing ads to other networks through a resale programme? Who, eventually, has control over the display of this type of advertising space?
- Some advertisers may be unwittingly placing ads which contain no obvious textual references to piracy, yet these pages do contain links to infringing content.

Drawing together these findings, some key lessons can be drawn:

- Indonesian residents have a higher chance of viewing high risk ads compared mainstream ads. Such ads are disproportionately promoting gambling, an activity which is illegal in Indonesia.
- Indonesian ads do not appear to be filtered, as sex industry ads were quite prevalent. Indonesia should investigate applying further controls that are text based as well as image based (eg, Ho & Watters, 2004).
- Advertisers need to have better mechanisms to control where their ads are eventually displayed on ad networks. Better systems for operational assurance and detection of misplaced ads need to considered, whether they operate using a whitelist or a blacklist (Ho & Watters, 2005).
- Regulatory approaches need to be considered to control the revenue flowing to rogue websites, and to minimise harm to users. A proposed code of conduct (Dredge, 2013) would be a first step to isolating rogue websites. Advertisers recently succeeded in pressuring Facebook, for example, to remove offensive ad placements by threatening to remove ads (as a group; Cellan-Jones, 2013).
- Other types of rogue content have been managed effectively by legal sanctions in the past. For example, paid search results for pharmaceuticals without prescriptions (O'Donnell, 2013) were removed by Google after they paid a very significant fine. However, Google's organic search results continue to display results from rogue drug sellers, ranging from marijuana through to MDMA and ecstasy (Watters & Phair, 2012). Searching Google for "buy ecstasy" returns numerous pages such as http://buyecstasypillsonline.wordpress.com/2013/07/27/buy-cheap-ecstasy-pills-online/ where users can order illicit drugs and have them delivered to order. Regulation of this type of advertising can be effective but more needs to be done.
- Since cyber criminals are very effective at exploiting jurisdictional differences, a global, industry wide code may have a greater impact on revenue flows for rogue websites. However, industry codes need to engage with ad networks that are placing ads for High Risk advertisers. At this stage, none of the top advertising networks supporting rogue websites

appear to be involved in the proposed codes of conduct[11]. Also, no additional burden should be placed on rightsholders to police the internet for offensive material.

- Policymakers, parents and educators need to be aware that the sex industry and online gambling sites specifically target torrent search and file locker sites for advertising their services. These material are *haram* for the majority of Indonesians, and the fact that they are so prevalent on sites promoting local content should be a great concern to everyone. There are absolutely no age warnings on these pages, and no attempt is made by these sites to verify if users are adults. Parents need to be aware that this is the type of content that piracy sites will serve up to their children, even if they are only intending to download torrent for music or less offensive content. The absence of traditional regulatory mechanisms for effectively controlling online content mean that new subcultural norms are rapidly being established online, and these can have profoundly negative consequences; for example, a progression model of rising interest in child exploitation material has been linked to the rise of the online porn culture, particularly where young users are inadvertently exposed to pornography through advertising (Prichard et al, 2013).
- Finally, this study illustrates the key role that search plays in the "illicit value chain" of (a) building up term lists to use to find infringing content, and (b) providing a list of ranked items in order of relevance to allow users to find the infringing content most relevant to their interests. While a small number of users will already know the domain name of the site that they wish to visit to download infringing content, search engines are facilitating the discovery of such sites by other users. Therefore, while domain blocking may be an effective strategy to prevent rogue sites providing infringing downloads, it can only be successful if replacement sites are unable to be "found" through search engines. As Mike Weatherly MP has recently noted in his discussion paper on search engines and piracy "removing a domain from search results will not solve piracy – although it would be a very important step in the right direction"[12]. Perhaps, as Mr Weatherly suggests, consumers expect search engines to direct them toward legitimate content.

References

Assyaukanie, L. (2009). *Islam and the Secular State in Indonesia*. Institute of Southeast Asian Studies

Browser Media (2008). DoubleClick deal means Google controls 69% of the online ad market. Downloaded from http://www.browsermedia.co.uk/2008/04/01/doubleclick-deal-means-google-controls-69-of-the-online-ad-market/

Cameron, N. (2013). Hong Kong's online advertising market valued at $17.1bn. Downloaded from http://www.cmo.com.au/article/466022/Hong Kong_online_advertising_market_valued_17_1bn/

[11] http://www.bbc.co.uk/news/technology-23325627

[12] http://www.mikeweatherleymp.com/2014/05/29/search-engines-and-piracy-a-discussion-paper-by-mike-weatherley-mp/

Cellan-Jones, R. (2013). Facebook removes ads from controversial pages to avoid boycott. Downloaded from http://www.bbc.co.uk/news/technology-23097411

Cleland, S. (2013). Why Google is Big Brother Inc. – A One-Page Graphic. Downloaded from http://www.precursorblog.com/?q=content/why-google-big-brother-inc-%E2%80%93-a-one-page-graphic-part-33-google-disrespect-privacy-series

Detica (2012). A data driven study of websites considered to be infringing copyright. Downloaded from http://www.prsformusic.com/aboutus/policyandresearch/researchandeconomics/Documents/TheSixBusinessModelsofCopyrightInfringement.pdf

Dredge, S. (2013). Google, David Lowery and the BPI talk ad-funded piracy. Downloaded from http://musically.com/2013/05/28/live-google-david-lowery-and-the-bpi-talk-ad-funded-piracy/

eMarketer (2012). US Online Advertising Spending to Surpass Print in 2012. Downloaded from http://www.emarketer.com/Article/US-Online-Advertising-Spending-Surpass-Print-2012/1008783

Felson, M and Clarke RV (1998). Opportunity Makes the Thief: practical theory for crime prevention Police Research Series Paper 98. London: Home Office

Gopal, R. D., Sanders, G. L., Bhattacharjee, S., Agrawal, M., & Wagner, S. C. (2004). A behavioral model of digital music piracy. Journal of Organizational Computing and Electronic Commerce, 14(2), 89-105.

Herps, A., Watters, P.A. & Pineda-Villavicencio, G. (2014). Measuring the prevalence of behavioral advertising using tracking cookies. *Proceedings of the 4th Cybercrime and Trustworthy Computing Workshop.*

Ho, W. H., & Watters, P. A. (2004, October). Statistical and structural approaches to filtering internet pornography. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on* (Vol. 5, pp. 4792-4798). IEEE.

Ho, W. H., & Watters, P. A. (2005, April). Identifying and blocking pornographic content. In *Data Engineering Workshops, 2005. 21st International Conference on* (pp. 1181-1181). IEEE.

Jakarta Post (2014). Students protest prostitution, gambling dens (28/01/14). http://www.thejakartapost.com/news/2014/01/28/students-protest-prostitution-gambling-dens.html

Kuprijanko, A. (2009). Försvaret: verksamheten är laglig. Sydsvenskan. Downloaded from http://archive.is/omksR.

McStay, Andrew. The mood of information: a critique of online behavioural advertising. Continuum, 2011.

Medoff, Norman J. Just a click away: Advertising on the Internet. Allyn & Bacon, Inc., 2000.

Olsson, S. (2006). Pirate Bay drar in miljonbelopp. Svenska Dagbladet. Downloaded from http://www.svd.se/nyheter/inrikes/pirate-bay-drar-in-miljonbelopp_334410.svd

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

Rawls, J. (1999). A Theory of Justice. Belknap Press.

Sheehan, B., Tsao, J., Bruno, E., Crider, D., Cutrone, J., Jones, C., & Serra, A. (Submitted). Improving the effectiveness of anti-digital music piracy advertising to college students.

Stabek, A., Brown, S., & Watters, P. A. (2009, July). The Case for a Consistent Cyberscam Classification Framework (CCCF). In Ubiquitous, Autonomic and Trusted Computing, 2009. UIC-ATC'09. Symposia and Workshops on (pp. 525-530). IEEE.

Stabek, A., Watters, P., & Layton, R. (2010, July). The seven scam types: mapping the terrain of cybercrime. In Cybercrime and Trustworthy Computing Workshop (CTC), 2010 Second (pp. 41-51). IEEE.

Sundberg, S. (2009). TPB har tjänat tio miljoner om året" (blog) (in Swedish). Svenska Dagbladet. Downloaded from http://www.webcitation.org/6D8mmNnUX

Taplin, J. (2013). USC Annenberg Lab Ad Transparency Report – January. Downloaded from http://www.annenberglab.com/sites/default/files/uploads/USCAnnenbergLab_AdReport_Jan2013.pdf

Treverton, G., Matthies, C., Cunningham, K., Goulka, J., Ridgeway, G., & Wong, A. (2009). Film Piracy, Organized Crime and Terrorism. RAND Corporation. Downloaded from http://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG742.pdf

Watters, P.A. (2014a). A Systematic Approach to Measuring Advertising Transparency Online: An Australian Case Study. In S. Cranefield, A. Trotman, & J. Yang (Eds.)Vol. 155 (pp. 59 - 67). *Proceedings of the Second Australasian Web Conference*

Watters, P.A. (2014b). Sweet as? Advertising on rogue websites in New Zealand. Available at SSRN: http://papers.ssrn.com/abstract=2466696

Watters, P.A. (2014c). Mainstream Ad Support for Online Piracy in Taiwan. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2405281

Watters, P.A. (2014d). Mainstream Advertising on Rogue Websites in Hong Kong: A Comparison of Chinese and Western Titles. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2468700

Watters, P.A. (2014e). Mainstream Advertising on Rogue Websites in Malaysia: A Comparison of Local and Foreign Content. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2469606

Watters, P.A. (2013a). The Prevalence of High-Risk and Mainstream Advertisements Targeting Canadians on Rogue Websites (December 2, 2013). Available at SSRN: http://ssrn.com/abstract=2389850

Watters, P.A. (2013b). Measuring Advertising Transparency in Singapore: An Investigation of Threats to Users. Available at SSRN: http://ssrn.com/abstract=2362626 or http://dx.doi.org/10.2139/ssrn.2362626

Watters, P.A. (2012). Taming the Cookie Monster: How Companies Track us Online. Centre for Internet Safety, University of Canberra. ISBN 978-1-922017-04-8.

Watters, P.A. & Phair, N. (2012). Detecting Illicit Drugs on Social Media Using Automated Social Media Intelligence Analysis (ASMIA). CSS 2012: 66-76.

Womack, B. (2013). Google Is Projected to Expand Lead in Online-Ad Market. Downloaded from http://www.bloomberg.com/news/2013-06-13/google-is-projected-to-expand-lead-in-online-ad-market.html

Appendix A – Rogue Sites in the Sample

www.hao123.com
www.icinema3satu.com
www.wardhanime.net
www.oploverz.net
www.bestusefuldownloads.com
www.downloadfilmbaru.com
www.jalantikus.com
www.animekompi.web.id
www.cinemaindo.com
forum.viva.co.idgfilm-barat
ganool.com
forum.indowebster.com
www.indowebster.com
icinema3satu.com
www.bersosial.com/forum/movie
icinema3satu.com
nontonmovie.com
bioskop25.blogspot.com
www.downloadfilmbaru.com
idfl.me
cepatnya.com
downloadfilmbaru.com
sidofi.com
downloadfilem.com
www.sinema-21.com
www.forumbebas.com
balqishotspot.blogspot.com
www.ryemovies.com
cupux-movie.com
movie76.com
300mbfilms.com
www.comotin.com
bioskopkita.com
www.gudangfilm.net
www.cinema81.com
www.indonesia-movie21.blogspot.com
cinemaindo.net
www.tempatdownloadfilm.com
sejutafilm.com
bioskopgratis.com

21sinema.com
www.film-baru.com
fajarnoah.blogspot.com
situsmovie.heck.in
www.m2cinema.com
rudihartoyo.blogspot.com
demuvia.com
newdownloadfilm.blogspot.com
unknown-green.blogspot.com
downloadfilmgratis-filmterbaru.blogspot.com
www.m-cinema21.com
davit-wae.blogspot.com
gallery-movie.blogspot.com
gudangfilmhoror.com
infobagus-foryou.blogspot.com
www.indonesianmoviesonline.com
joey87.heck.in
cinema-xcinema.blogspot.com
gratis-film-21.blogspot.com
www.video-movies21.com
filmseo.blogspot.com
fullmovies31.heck.in
cibitungknowing.com
donmies.blogspot.com
www.indo-movie.com
film-horor.blogspot.com
bioskopkita.net
jaiifanii.mywapblog.com
downloadfilmz.com
mas-acep.com
www.klikmovie.com
movie-downloadzz.blogspot.com
ktimen.mywapblog.com
4themovie.blogspot.com
situsdownloadfilm.com
fdmovie.com
supernova21cinema.blogspot.com
cyberload.us
alkisah-nolove.blogspot.com
indonesiamovie1.blogspot.com
indianfilmbare.blogspot.com

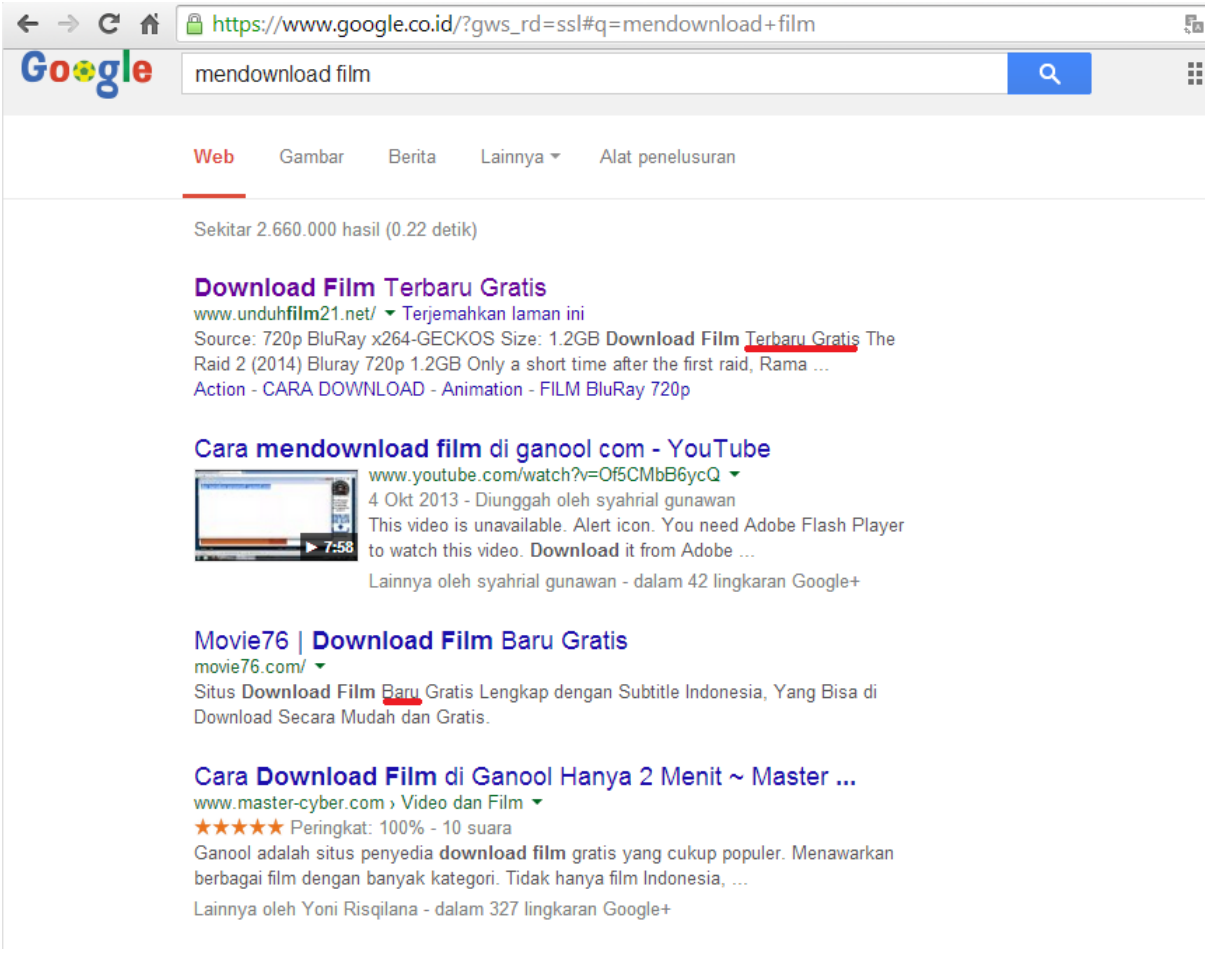
blackeditiondownload.blogspot.com
movieforlife.com
www.letshared.com
moviestudio090.blogspot.com
dnacrewz.fantastique.tv
www.icinema27.com
workshop-movie.blogspot.com
bioskop.juraganfoto.net
filmonline79.blogspot.com
bluray21movie.blogspot.com
www.ngunduhfilm.net
watchonline.downloadmoviefree.info
moviedownloadstation.blogspot.com
movieindo.com
cekdot.blogspot.com
gudangfilmeyncom.blogspot.com
www.legendafilm.com

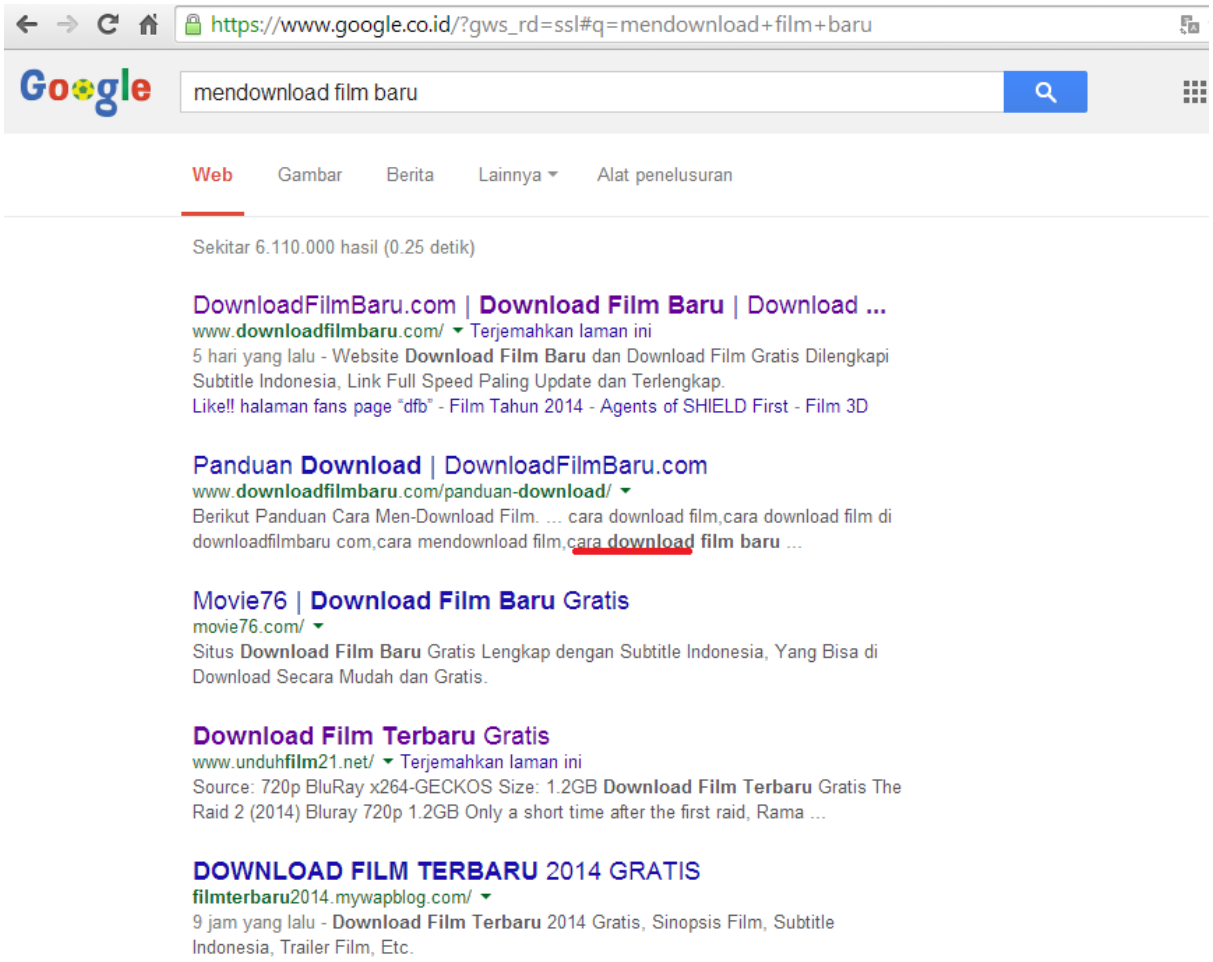
Appendix B – Full List of advertising networks detected

Ad Network	Frequency
tujuh.co.id	270
yllix.com	264
innity.net	260
bidvertiser.com	230
adplus.co.id	207
ads.ec	148
popcash.net	120
adsensecamp.com	120
ad4game.com	90
maingames.co.id	80
777seo.com	70
kliksaya.com	64
buzzcity.net	60
postimg.org	50
popads.net	40
adspserving.com	40
flashvortex.com	32
adbrite.com	30
onclickads.net	30
clicksor.com	30
adcash.com	30
ppcindo.net	30
emailduit.com	20
cinema81.com	20
21sinema.com	20
adroll.com	20
bp.blogspot.com	20
olb365.com	10
hopfeed.com;	10
adreactor.com	10
sarana99.com	10
hopfeed.com	10
agent-pialadunia.com	10
picadmedia.com	10
oploverz.net	10
ipoker8.com	10
icinema27.com	10
altervista.org	10
goo.gl	10
googleusercontent.com	10
besthitsnow.com	10

onclasrv.com	10
agenbola123.com	10
forum.idws-static.com	10
cmcpoker.com	10
juaraqq.com	10

Appendix C – The Role of Search




https://www.google.co.id/?gws_rd=ssl#q=mendownload+film+baru
Google
mendownload film baru
Web
Gambar
Berita
Lainnya
Alat penelusuran
Sekitar 6.110.000 hasil (0.25 detik)
DownloadFilmBaru.com | Download Film Baru | Download ...
www.downloadfilmbaru.com/ Terjemahkan laman ini
5 hari yang lalu - Website Download Film Baru dan Download Film Gratis Dilengkapi Subtitle Indonesia, Link Full Speed Paling Update dan Terlengkap.
Like!! halaman fans page "dfb" - Film Tahun 2014 - Agents of SHIELD First - Film 3D
Panduan Download | DownloadFilmBaru.com
www.downloadfilmbaru.com/panduan-download/
Berikut Panduan Cara Men-Download Film. ... cara download film,cara download film di downloadfilmbaru com,cara mendownload film,cara download film baru ...
Movie76 | Download Film Baru Gratis
movie76.com/
Situs Download Film Baru Gratis Lengkap dengan Subtitle Indonesia, Yang Bisa di Download Secara Mudah dan Gratis.
Download Film Terbaru Gratis
www.unduhfilm21.net/ Terjemahkan laman ini
Source: 720p BluRay x264-GECKOS Size: 1.2GB Download Film Terbaru Gratis The Raid 2 (2014) Bluray 720p 1.2GB Only a short time after the first raid, Rama ...
DOWNLOAD FILM TERBARU 2014 GRATIS
filmterbaru2014.mywapblog.com/
9 jam yang lalu - Download Film Terbaru 2014 Gratis, Sinopsis Film, Subtitle Indonesia, Trailer Film, Etc.