

Mainstream Advertising on Rogue Websites in Hong Kong: A Comparison of Chinese and Western Titles

Dr. Paul A. Watters, Massey University

JULY 2014



MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

CONTENTS

EXECUTIVE SUMMARY 	1
INTRODUCTION 	3
METHODS 	7
CONCLUSION 	17
REFERENCES 	19

EXECUTIVE SUMMARY

A number of studies have recently investigated the role played by mainstream internet advertising in supporting the revenue of rogue websites (Taplin, 2013). Such advertising by household names – including multinational corporations, governments and charities – generates enormous profit margins for operators of these websites, which present an ongoing threat to the viability of Hong Kong's creative industries.

However, a recent study by Watters (2014a) 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, 2014b) all had mainstream ad 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, 2014c) 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 key question remains 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).

In this study, we directly measured prevalence rates for High Risk versus Mainstream ads for local versus Hollywood content for movies/TV in Hong Kong. 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. 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. Like 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.

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 for local content, and were similar to Canada, Singapore, Australia and New Zealand for Hollywood content.

The policy implications of this result and future research directions, including methodology enhancements, are discussed.

Keywords

Infringing content, internet advertising, Digital Millennium 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. Hong Kong’s online advertising market was valued at HKD\$7.3b in 2013¹ and is growing at a rate of 13% pa.

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 ISPs in the USA 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.

1 <http://www.scmp.com/business/economy/article/1160657/advertising-spending-hong-kong-see-modest-gains-year>

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 Hong Kong (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). Assuming the same growth rate as Hong Kong, this will reach \$44.74b for 2013 and \$51.01b in 2014.

Selling advertising on file locker and torrent search sites is the major source of revenue for such sites.

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

Internationally, advertisers and ad networks are joining in creating “best practice” guidelines², to allow advertisers to better control their brand images and avoid association with piracy syndicates.

2 See “ICC Policy Statement; Safeguarding Against the Misplacement of Digital Advertising”, International Chamber of Commerce Document No. 240/52-707, Paris, March 6 2014

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 accessed by Hong Kong residents (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)³.

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 activities on illegal websites, including rogue piracy sites; 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. Several recent sets of best practice guidelines for ad networks to address piracy and counterfeiting have recently been released⁴, and early indications are that most of the world’s major web companies will participate⁵.

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

3 http://au.ibtimes.com/articles/533033/20140106/pirate-bay-popular-torrent-site-top-10.htm#_Uvk8g_mSw3l

4 An example can be found here: <http://www.jicwebs.org/agreed-principles/latest-news/133-jicwebs-approves-industry-principles-aimed-at-growing-safer-online-ad-placement>

5 <http://torrentfreak.com/tech-giants-sign-deal-to-ban-advertising-on-pirate-websites-130715/>

6 Note that the USC reports are not peer reviewed

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

A number of studies have recently examined the relationship between piracy sites and online advertising networks. The USC report (Taplin, 2013) provides a method for revealing the advertisers whose ads are most likely to be served up on these sites, which may be occurring without the direct knowledge of the advertiser. While the objectives of USC research are significant, the monthly rankings of the “top ten” advertising networks responsible for placing the most ads on web sites that support infringing content are surprisingly variable – Google, for example, was ranked at #2 in January 2013, but did not appear at all in the February and March 2013 lists at all. One interpretation of the result could be that the January report achieved its goal of sensitising advertising networks, and that Google subsequently withdrew from placing ads on those sites. Alternatively, the variation could be due to biases inherent in studies using an observational methodology, including selection bias, information bias and recall bias. The lack of detail in measures like the “top 500” sites prevent the study results from being directly replicated, which would be the standard required for peer review by other researchers⁶. By not providing this level of detail, the credibility of the USC report may be called into question by the very vocal critics of any research in the anti-piracy field.

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. High risk ads are those which have the potential to cause harm to users, and include pornography, gambling, malware and scams. After examining the ads being served to Australians in that study, it was found that 99% of the ads from the “top 500” sites were High-Risk, while only 1% were Mainstream. Subsequent studies measured the prevalence rates for

Hollywood movies/TV in Singapore, Canada, and New Zealand, finding that prevalence rates for mainstream varied between 1-10%, while high risk ads had prevalence rates of 90-99%.

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, so another study (Watters, 2014c) investigated "local" content in Taiwan (Chinese language), and compared the prevalence rates for high risk versus mainstream ads with Hollywood content. It was found that mainstream advertising was much more prevalent for local content sites compared to viewing Hollywood content sites in the other countries examined. 61% of ads were Mainstream, while 39% were High Risk for local content. However, no direct comparison was undertaken to see whether Taiwanese viewers were more likely to see mainstream advertising on local content sites versus Hollywood sites.

The aim of this study is to undertake such a comparison, by measuring the prevalence rates for High Risk versus Mainstream ads for local versus Hollywood content for movies/TV in Hong Kong. It is predicted that, for local content, the prevalence rates will be similar to Taiwan, but for Hollywood content, the prevalence rates will be similar to Australia, New Zealand, Canada and Singapore.



METHODS

The main goal of the methodology is to identify the advertising networks and advertisers from a sample of DMCA complaints, which have been ranked in terms of the number of complaints upheld by Google (through their Transparency Report)⁷, or from a list of expert-identified rogue sites.

The DMCA complaints typically relate to the availability of search results for a wide range of potentially infringing content; by only selecting the most complained about and subsequently upheld complaints as assessed by a third-party (Google), the results should be robust against criticisms that there is no proof that the sites in question were hosting torrents of infringing content or infringing content directly, in the case of a file locker site. The methodology operates by downloading each page from the “top 500” complaints submitted to Google within the previous month, ordered by the number of upheld complaints. Since each DMCA notice can contain many thousands of individual URLs, 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. An example obfuscated URL is shown below, from adcash.com:

```
var ct_SuLoaded = 1; var ct_SuUrl='http://'+
+ 'www.adc' + 'ash.com' + '/script/pop_
packcpm.php?k=52e9e900315f7850142.1482334&h
=e0004318879f95130e6cf975c6b761523a69e256&i
d=0&ban=850142&r=128995&ref=&data=&subid=';
var ct_nSuUrl='http://' + 'www.adc' + 'ash.
com' + '/script/pop_packcpm.php?k=52e9e9003
15f7850142.1482334&h=e0004318879f95130e6cf9
75c6b761523a69e256&id=0&ban=850142&r=128995
&ref=&data=&subid=&new=1';
```

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

⁷ Google.com.hk is ranked as the #1 site for Hong Kongers to visit; <http://www.alexa.com/siteinfo/>

The algorithm works as follows:

1. A data collection system is installed physically or logically to attract advertising for a specific geographical/country segment. For this study, Hong Kong was selected.
2. The current Google Transparency Report⁸ is downloaded, which lists all of the DMCA requests for a specific time period⁹. This list provides one means of identifying sites involved in sharing pirated material. In parallel, a list of expert-identified rogue sites targeting Hong Kong local content (Chinese language) was compiled.
3. The dataset is sorted by the number of URLs removed, retaining the “top 500” DMCA requests (the request list) by complaint category. For this study, the complaint categories were movies and TV shows; other complaint categories such as pirated software, adult material, music etc were excluded.
4. For each report in the request list first 10 URLs are extracted as a representative sample of all of the URLs contained within the report. This gives a total of 5,000 web pages to be downloaded (the sample) for Hollywood content. For local content, 38 sites were downloaded from an original list of 98¹⁰, so 380 pages were downloaded.
5. Each of the 5,380 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.
6. 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 List¹¹ (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.
7. 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.

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

Results – Hollywood Movies/TV

Appendix A contains a list of the DMCA notices identified in Step 3, including TV and movies from major Hollywood studios such as Fox, Warner Bros etc during the first two weeks of December 2013. From the 5,000 pages analysed in Step 4, a total of 7,097 advertising items and 2,213 visible ads were identified in Step 6¹². 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. 279 unique domains for advertising networks were identified, indicating an average 7.93 visible ads per network in the sample (keeping in the mind that the distribution – shown in Table 1’s Top 10 advertising networks – is highly non-uniform). 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, i.e., 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.

8 <https://www.google.com/transparencyreport/removals/copyright/data/>

9 The DMCA list for May-July 2013 was used in this analysis
10 Local content sites were excluded if they contained no advertising, duplicated an existing site or did not respond to HTTP requests on the sampling date

11 <http://easylist.adblockplus.org/en/>

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

Table 1. Frequency Analysis by Advertising Network – Top 10¹³

Advertising Network	Frequency	% of Ads
propellerads.com	1,363	19.21%
torrentus.to	1,103	15.54%
onclickads.net	329	4.64%
tumejortv.com	294	4.14%
admxr.com	261	3.68%
filestube.to	234	3.30%
adexprt.com	234	3.30%
sharelab.org	226	3.18%
adcash.com	216	3.04%
fulldls.com	178	2.51%

High-Risk Advertising – Top 10 (Hollywood Movies/TV)

Table 2 contains a summary of the results from the Top 10 ad networks. There were 2,185 advertisements in this sub-sample. Each of these advertisements was downloaded, visually inspected and categorised. The results indicate that malware, scams (including employment, investment and SMS premium rate), and the sex industry were the most popular distinct advertising types in Hong Kong for the Top 10 networks.

An example of malware downloaded is provided by the advertising link <http://isohunt.com/a/adclick.php?bannerid=493&zoneid=&source=btDetails-banner&dest=http%3A%2F%2Fip.ncdownloader.com%2Fexact%2F%3Fq%3DCannonball+Run+II.+1984>. Upon visiting this page, a download is initiated to the user’s computer containing the file Cannonball Run II. 1984.exe which is only 292K in size – much smaller than a typical video file of at least 700M. Running this file through the online scanner virscan.org – which analyses suspicious files using 36 different products – the file is verified as ADWARE/Adware. Gen (<http://v.virscan.org/ADWARE/Adware.Gen.html>) by AntiVir 8.2.10.202 and as Adware.Downware.1166 by ClamAV (<http://v.virscan.org/Adware.Downware.1166.html>). A review of the other known filenames associated with this malware indicates a typical strategy of associating a desirable filename with the malicious code, ie, using a filename that users desiring to download infringing content will click on, including **Mortal Kombat – Complete Edition Crack (2013) Download.exe** and **Transformers 3 - Dark of the Moon (2011) [1080p].exe**.

13 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.

Table 2. High-Risk ad type frequencies by network

Ad Network	Ad URLs	Distinct Ad URLs	Sex Industry	Malware	Downloading	Gaming/ Gambling	Scams	TOTAL
propellerads.com	1,363	103		44				44
torrentus.to	1,103	6		1				1
onclickads.net	329	21						0
tumejortv.com	294	74	6	102		13	36	157
admxr.com	261	186		38	1		8	47
filestube.to	234	113		90				90
adexprt.com	234	84	2	64			7	73
sharelab.org	226	43			39			39
adcash.com	216	48		7	1	2		10
fulldls.com	178	99		76		1		77
TOTAL	4,438	777	8	422	41	16	51	538

High-Risk Advertising – (Hollywood Movies/TV)

Table 3 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 the sex industry, malware, downloading sites, gambling or scams (including employment, investment and SMS premium rate) were the most popular distinct advertising types. The categories are summarised in Figure 1.

Figure 1. High-Risk advertising

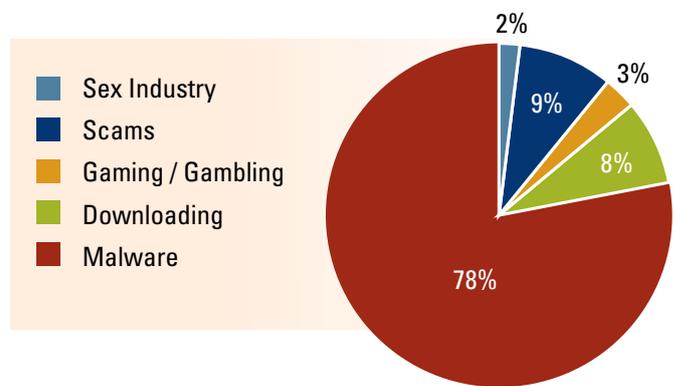


Table 3. Frequency by Ad Category – High Risk Ads

key: ■ Sex Industry ■ Scams ■ Gambling ■ Malware ■ Downloading Sites

	Sex Industry	Malware	Downloading Sites	Scams	Gambling
N	8	422	41	16	51
%	1.03%	54.31%	5.28%	2.06%	6.56%

Mainstream Advertising – All Sites (Hollywood Movies/TV)

Table 4 contains the results of the step 8 results obtained by visually inspecting every advertisement in the sample (comprising 10 pages from each of the Google Ad Transparency Top 500 complaints) 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 3.84% of the ads sampled consistently showed evidence of targeting Hong Kong users through the presentation of Mainstream advertising. 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.

Figure 2. Mainstream advertising

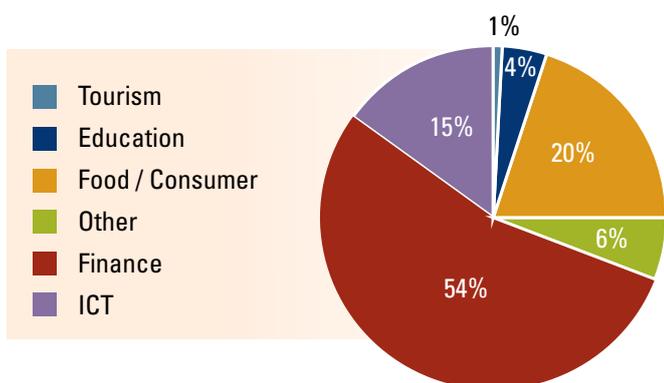
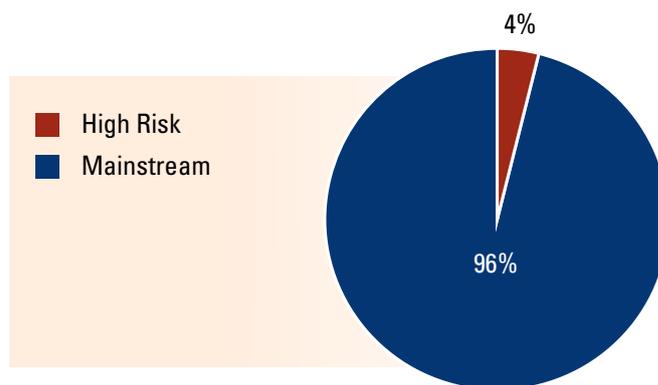


Figure 3. Mainstream versus High-Risk Advertising



Mainstream Advertising – Top 10 Ads (Hollywood Movies/TV)

Table 5 shows the frequency distribution for the ten most frequently detected Mainstream ads. The key difference to note is the relative decline in 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.

The relevance of some sites with a Polish domain name (eg, chomikuj.pl) to Hong Kong users might not be apparent. Yet, chomikuj.pl is ranked by Alexa globally at 1,538, and 37.4% of its traffic is referred by Google¹⁴. This makes it very appealing to a global audience, including users in Hong Kong. Furthermore, the operations of the site are geographically segmented: the technical contact is given as FS File Solutions Ltd in Cyprus, and the domain name is registered by Instra Corporation Pty Ltd in Australia.

14 <http://www.alexa.com/siteinfo/chomikuj.pl>

Table 5. Mainstream Advertisers Detected (Top 10)

Advertiser	Ad Network	Example Site Where Displayed	Frequency	%
Options Now	propellerads	limetorrents.com	46	16.09%
Crazy Apps Maker	adk2.com	torrentroom.com	11	9.20%
Interia	gemius.pl	chomikuj.pl	4	6.90%
Nissan	gemius.pl	chomikuj.pl	3	5.75%
Groupon	bannerplay.com	majaa.net	3	4.60%
Graha Wish	adlickmedia.com	tobrut.com	2	4.60%
Mbank	gemius.pl	chomikuj.pl	2	4.60%
Porsche	gemius.pl	chomikuj.pl	2	3.45%
Scholl	gemius.pl	chomikuj.pl	2	3.45%
Renault	gemius.pl	chomikuj.pl	2	3.45%

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).

Results – Hong Kong Movies/TV

Appendix C contains a list of the expert-identified rogue sites which make available torrents or links to infringing content for download. From the 380 pages analysed in Step 4, representing ten page impressions from each site, a total of 3,530 advertising items and 590 visible ads were identified in Step 6¹⁵. 27 unique domains for advertising networks were identified, indicating an average 21.85 visible ads per network in the sample. Appendix D contains the complete list of advertising networks detected.

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

Table 6. Frequency Analysis by Advertising Network – Top 10¹⁶

Advertising Network	Frequency	%
xemphimso.com	2	62.04%
bp.blogspot.com	500	14.16%
adnetwork.vn	190	5.38%
bidvertiser.com	160	4.53%
thegioiphim.com	150	4.25%
kenh88.com	90	2.55%
1stdrama.com	20	0.57%
kpopstyle.net	20	0.57%
tlvmedia.com	20	0.57%
v9bet.com	20	0.57%

High-Risk Advertising – Top 10 (Hong Kong Movies/TV)

Table 7 contains a summary of the results from the Top 10 ad networks. There were 3,360 advertisements in this sub-sample. Each of these advertisements was downloaded, visually inspected and categorised. 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 Hong Kong for the Top 10 networks.

¹⁵ Advertising items include any scripts, images, spacers etc being referenced from an Adblock domain, in addition to visible ads

¹⁶ 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.

Table 7. High-Risk ad type frequencies by network

Ad Network	Ad URLs	Distinct Ad URLs	Sex Industry	Malware	Downloading	Gambling	Scams	TOTAL
xemphimso.com	2,190	155				155		155
bp.blogspot.com	500	47	1					1
adnetwork.vn	190	14						0
bidvertiser.com	160	9						0
thegioiphim.com	150	9						0
kenh88.com	90	9					1	0
1stdrama.com	20	2		2				2
kpopstyle.net	20	6						0
tlvmedia.com	20	2						0
v9bet.com	20	1						0
TOTAL	3,360	254	1	2	0	155	1	158

High-Risk Advertising – All Sites (Hong Kong Movies/TV)

Table 8 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 or scams (including employment, investment and SMS premium rate) were the most popular distinct advertising types. The categories are summarised in Figure 1. High-Risk ads accounted for 38.64% of all visible ads.

Figure 4. High-Risk advertising

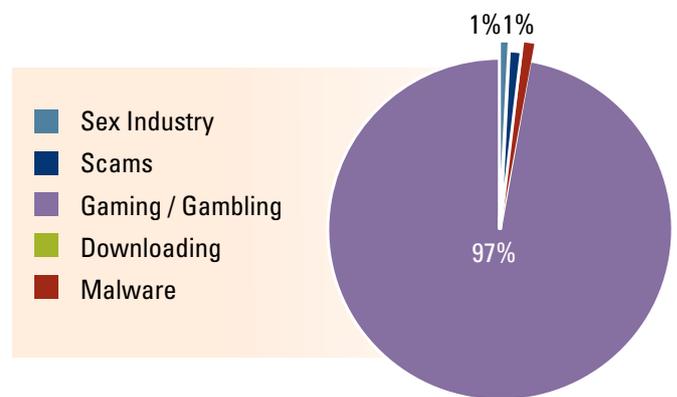


Table 8. Frequency by Ad Category – High Risk Ads

	Sex	Malware	Download	Gambling	Scams
N	1	2	0	155	1
%	0.63%	1.27%	0.00%	98.10%	0.63%

Mainstream Advertising – All Sites (Hong Kong Movies/TV)

Table 9 contains the results of the step 8 results obtained by visually inspecting every advertisement in the sample (comprising 10 page impressions from each of the rogue sites) to identify whether it contained any Mainstream advertising. Some 61.36% of the ads sampled consistently showed evidence of targeting Hong Kong users through the presentation of Mainstream advertising. A breakdown by industry category is shown in Figure 5, and the relative composition of Mainstream to High-Risk ads is shown in Figure 6.

Table 9. Mainstream Advertisers Detected

Advertiser	Frequency
Dramastyle	65
Baidu	53
Taoweibtrongoi	31
galaxy macau	27
Xdeal	20
Camel shoes	20
Ambient Digital	10
Dropbox	10
Hairgrow	10
Loverscare	10
La Miu	10
Qihoo360	10
Mua Ngay	10
Clinton Foundation	9
Redcross	9
MediaV	9
Emule	8
PS4book.com	7
Shanghai Ren-Ai Hospital	7
S	5
Adoptuskids.org	4
Givology	4
Discovertheforest	4
Nikon	3
CRM Korea	2
Glamoursales	2
VIPABC	2
Volvo	1

Figure 5. Mainstream advertising

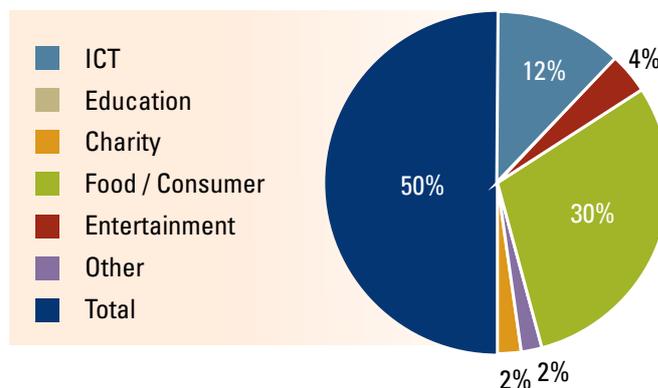
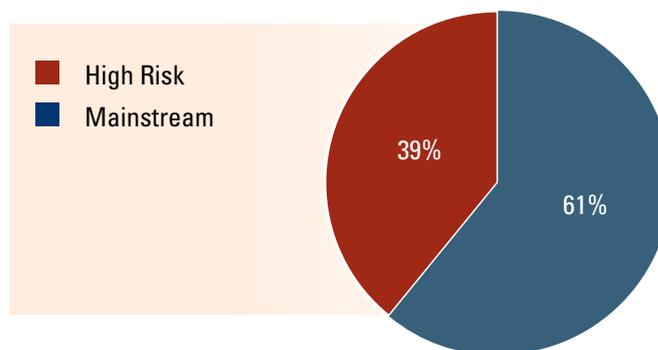


Figure 6. Mainstream versus High-Risk Advertising



Mainstream Advertising – Top 10 Ads (Hong Kong Movies/TV)

Table 10 shows the frequency distribution for the ten most frequently detected Mainstream ads.

Table 10 – Mainstream Advertisers Detected (Top 10)

Advertiser	Ad Network	Example Site Where Displayed	Frequency	%
Dramastyle	dramastory.net	dramastory.net	65	17.96%
Baidu	UNKNOWN	tvbzhibo.com	53	14.64%
Taowebrongoi	phimtvb.biz	phimtvb.biz	31	8.56%
Galaxy Macau	Google Ads	thegiophim.com	27	7.46%
Xdeal	hayghe.com	hayghe.com	20	5.52%
Camel shoes	chanet.com.cn	tom365.me	20	5.52%
Red Cross	Yahoo! Ads	azndrama.info	11	3.03%
Ambient Digital	ambientdigitalgroup.com	phimbo24h.com	10	2.76%
Dropbox	db.tt	online.hk	10	2.76%
51fanli.com	360.cn	www.360.cn	10	2.76%

CONCLUSION

The hypothesis that local content would attract more mainstream advertising than Hollywood content was supported by this study. It would seem that efforts by the online advertising community to avoid placement of mainstream ads on pirate websites have not been extended to Asian-language internet sites.

While the Taiwan study found that this was true for local Chinese content, this study provided the first direct comparison of advertising for local content versus Hollywood, and the results are startling: while the levels of mainstream advertising on rogue sites promoting Hollywood content were comparable with other countries including Australia, New Zealand, Canada, and Singapore, the levels of mainstream ads increased by a factor 20x when local content was promoted. This may be a result of brand protection strategies being more focused on sites which are listed in Google's DMCA Transparency Report; in a sense, the local content sites in Hong Kong appear to be escaping the attention of the global community, and drawing increased profits from mainstream advertisers who are shunning international content sites. This implies greater damage to Asian content industries from ad-supported piracy. Also, while some local content sites were carrying mainstream ads by local networks, Yahoo! and Google ads were also detected promoting mainstream content.

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

- Pirate websites featuring Asian (Chinese) content are much more likely to be supported by mainstream advertising than are international content sites.
- For local (Chinese-language) 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.
- For Hollywood content, high-risk advertising primarily comprised malware, while for local content, the greatest risk was from gambling.
- The top ad networks serving ads to Hong Kong residents for local content included chanet.com.cn, Yahoo! ads and Google ads, while xemphimso.com, bp.blogspot.com, adnetwork.vn and bidvertiser.com were found for Hollywood sites.
- A number of major household name brands in Hong Kong are choosing to advertise on pirate sites and their pages which are promoting the distribution of local and Hollywood infringing content (movies

and TV shows). This is causing greater damage to local (Chinese-language) content producers. 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 on pages which contain no obvious textual references to piracy, yet these pages do contain links to infringing content, and their referring pages have been verified by Google as DMCA-infringing.

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

- Hong Kong residents have a higher chance of viewing Mainstream ads on local movie/TV sites compared to Hollywood sites; the converse is also true. Hong Kong "eyeballs" are thus contributing to the demise of the local content industry.
- Hong Kong ads do not appear to be filtered. Hong Kong's advertising industry and regulators 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 be 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 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 that company 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://buyecstasyonline.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. The International Chamber of Commerce recently began promoting a global initiative to self-regulatory codes. However, this approach needs to be followed up by the Asian advertising industries.
- With respect to "high risk" advertising, 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 UK code of conduct¹⁷.
- Finally, and perhaps most importantly, 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. Ads promoting gambling were particularly prevalent on sites promoting local content. There are absolutely no age warnings on these pages, and no attempt is made by the Pirate Bay (for example) to verify if users are adults. Parents need to be aware that this is the type of content that will be served 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).

17

<http://www.bbc.co.uk/news/technology-23325627>

REFERENCES

- 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/
- 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/researchandconomics/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.
- 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 (CCCCF). 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. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2405281

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>

Appendices

(Appendices to this report can be found in the online version: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2468700)



