Mainstream Advertising Support for Online Piracy in Taiwan.

Dr. Paul A. Watters, Massey University

MARCH 2014

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

CONTENTS

EXECUTIVE SUMMARY | 2

INTRODUCTION | 3

METHODOLOGY | 5

RESULTS | 7

CONCLUSION | 15

REFERENCES | .... |

EXECUTIVE SUMMARY

Taiwan has a vibrant digital economy, characterised by high-speed broadband and the rise of e-commerce sites presented in Traditional Chinese script, providing everything from online auctions, payments and consumer retail.

Yet rogue websites are also widely supported in Taiwan, by millions of users whose browsing and clicking activity with ads generates revenue and profit for these rogue sites. The revenue models underpinning rogue websites have only recently received research attention (Taplin, 2013; Watters, 2013a; Watters, 2103b). The accumulation of wealth through advertising on rogue websites diverts revenues from rightsholders, who have invested in creative industries, and threatens the viability of such industries by eroding the earnings base. Yet the greatest risks from advertising on these sites are not primarily financial; instead, these sites represent a clear and present danger to their users, who are often children.

While users are often exposed to "mainstream" advertising – juxtaposing household company names with hardcore pornography and other illicit material – "high risk" advertising has been found to comprise the overwhelming majority of ads targeting Australians (Watters, 2013a) and Singaporeans (Watters, 2013b). In this study, we use the methodology developed by Watters (2013a) to assess the prevalence of mainstream and High Risk advertising. A total of 1,000 webpages were sampled from known rogue sites in Taiwan and were downloaded there.

Each ad banner was categorised as being High Risk or Mainstream, after each page was verified as being in breach of DMCA for movies and TV from major international studios. 61% of ads were Mainstream, while 39% were High Risk.

The prevalence of Mainstream ads being served to Taiwan is very high compared to similar advertising being shown to Australians, Singaporeans and Canadians. Even the Taiwanese government was observed to be advertising on these sites. The policy implications of this result and future research directions, including methodology enhancements, are discussed.

Key Findings

- 39% of the ads were High-Risk; 61% were Mainstream.
- In the High-Risk ads, the overwhelming majority were for gambling (83%), followed by the sex industry, malicious or suspected malicious code, and scams of various kinds, including premium rate SMS, investment and employment scams. These results were quite different to other countries; variations may be due to the implementation of ISP or government filtering regimes, and local customs and interests
- A significant number of household name brands in Taiwan are choosing to advertise on rogue sites, which are facilitating the distribution of infringing content (movies and TV shows). 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?

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. Taiwan's online advertising market was estimated to be worth NT$13.8 billion in 2013.[1]

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.

[1] http://www.dma.org.tw/upload/ResourceTrend/20130325035256917.pdf

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 an Australian case study (Herps et al, 2013). 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).

> *"Advertising revenues provide the commercial motivation for criminal syndicates to operate such 'rogue' websites."*

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 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 (Olsson, 2006; Kuprianko, 2009), with advertising revenues in the order of US$1.4m (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.

> *"Selling advertising on file locker and torrent search sites is the major source of revenue fir such sites.*
> *A key question for advertisers and ad networks is the extent to which they wish to be associated with this type of activity..."*

Tw116.com, for example, regularly features in the Top 150 sites accessed by Taiwanese (as computed by alexa.com[2]) , and so it is a potentially attractive space for advertisers and ad networks, since the number of potential "eyeballs" is very high. 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[3], and early indications are that most of the world's major web companies will participate[4].

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

Recently, the USC Annenberg Lab has begun producing a report that explores the relationship between piracy sites and online advertising networks (Taplin, 2013). The USC report 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, in the way that infringing sites are selected. The study uses a single source (the Google Transparency Report of domains with the most DMCA takedown requests), rather than using a consensus technique which combines the ranks of several different data sources to provide the most accurate ranking. This type of triangulation is commonly used in observational studies as a form of triangulation;
- Information bias, since only one technique for collecting data is used (HTML and JavaScript code scraping), where other techniques may be more accurate or representative of advertising behaviour. For example, persistent cookies have been strongly associated with behavioural advertising, and the frequency of tracking cookies being stored by ad networks could provide an alternative measure of presence of significance. Yet the USC report does not analyse cookies at all; and
- Recall bias, since the data analysed was only from English-language websites and advertising networks which may potentially have a higher level of visibility than networks which operate in other geographic zones, languages, encoding types etc.

Also, the lack of detail in how 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 researchers. By not providing this level of detail, the credibility of the USC report may be called into question by

[2] http://www.alexa.com/topsites/countries;4/TW

[3] http://2013ippractices.com/bestpracticesguidelinesforadnetworkstoaddresspiracyandcounterfeiting.html

[4] http://torrentfreak.com/tech-giants-sign-deal-to-ban-advertising-on-pirate-websites-130715/

the very vocal critics of any research in the anti-piracy field.

In this paper, we present a more rigorous and fully replicable methodology which should provide a much clearer view of advertising network behaviour in different countries, jurisdictions, languages etc. In this study, we specifically target Taiwanese content produced and distributed within the local market; the methodology itself is sufficiently general that it could be applied to any country and any category, including music, computer games, e-books etc.

Three previous studies using this methodology focused on the ads being served to Australians, Singaporeans and Canadians. In the Australian study, it was found that 99% of the ads from the "top 500" sites were High-Risk, while only 1% were Mainstream. In the Singaporean study, it was found that 90% of the ads from the "top 500" sites were High-Risk, while 10% were Mainstream (a similar rate was found for Canada; Watters, 2013c). It is predicted that a similar proportion in the range 1-10% Mainstream ads will be found in Taiwan.

METHODOLOGY

The main goal of the methodology is to identify the advertising networks and advertisers from a sample of known rogue websites in Taiwan, where recent release movie and TV titles are used to identify pages offering infringing content.

The methodology operates by downloading each page a number of times, with the advertisements on each page also being 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, advertisements are also categorised as belonging to the "mainstream" or "High-Risk" groups. Advertisers who may otherwise be unable to place their ads on a mainstream site can often 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

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, Taiwan was selected.
2. A list of known rogue sites for the target country is obtained from experts, and ten are selected at random.
3. For each site, a recent release TV show or movie is selected from the index page, since this indicates that it is available to download.
4. Each page is downloaded 100 times, giving a total of 1,000 web pages (the sample). Each sample page is downloaded, and a screenshot is taken, showing the ads being served. Note that 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[5](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.

[5] http://easylist.adblockplus.org/en/ - Adblock is a widely-used open-source ad-blocking system whose lists of acceptable ads or blocked ads are incorporated into major internet browsers, to allow consumers to block unacceptable advertising

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.

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

For all "mainstream" advertising networks identified as present on web page, a further 100 samples of advertising are downloaded and added to any unseen advertisers to the identified list.

For this project, ten popular Taiwanese rogue sites were selected for analysis, with a specific recent release Hollywood movie or TV show selected from the index page, as shown in Table 1. 100 sequential page impressions were then saved from each URL, and the ads for each were reviewed and categorised at belonging to the Mainstream or High Risk categories. For ads belonging to the Mainstream category, advertisers were identified and tabulated by industry category.

TABLE 1. URLs from rogue Taiwanese websites analysed

URL	TV/Movie
http://www.funshion.com/subject/107222/	Elysium
http://www.tw116.com/occident/iguokongbugushinvwujihuidisanji/	American Horror Story
http://www.yyets.com/resource/30859	White House Down
http://www.qiredy.com/occident/guotuanquandisanji/	Homeland
http://www.9tvb.com/vod/77136.html	World War Z
http://www.tw115.com/movie/20024.html	Pacific Rim
http://www.2000mov.com/vod-read-id-117.html	Walking Dead
http://dlkoo.com/down/2/2013/308443657.html	This Is The End
http://370kan.com/kanview/kanindex24828.html	Hangover Part III
http://www.fun698.com/vod-read-id-47324.html	Ted

RESULTS

From the 1,000 pages analysed in Step 4, a total of 2,887 visible ads were identified in Step 6[6]. These ads were then categorised, with 1,128 being High Risk (39.07%) and 60.93% being mainstream.

This contrasts greatly with the results obtained previously for Australia, for example, where only 1% of the ads were detected as mainstream.

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. Only three advertising networks were identified, as shown in Table 3. This means that only 13.85% of the actual ads displayed were identified through the Adblock list – this is quite unusual, as other countries analysed have always had their ads identified by Adblock.

It is extremely concerning that [Taiwanese viewers} would see almost 40 times more mainstream ads than Australians.

Further research is needed to compile a Taiwanese-specific Adblocklist that could be used to provide further intelligence on local advertising networks operating only within Taiwan, especially as they are serving up the lion's share of the ads on rogue sites. The fact that Adblock does not detect them makes Taiwanese users especially vulnerable to High Risk ads, even though these were fewer in number than the mainstream ads. It also indicates that the multinational advertisers (eg, Mini, Honda, Maybelline, Clarins etc) are engaging with these local ad networks.

Where advertisers are using Adblock-identified networks, they also use redirections to disguise their links to multinational advertising networks. For example, an ad element like http://www.9tvb.com/ad/A1.html contains two small ad images linked to software downloads, but the main banner ad is served up by Yahoo ads http://ads.yahoo.com/clk?3,eJytjd0KgkAQhZ-mOwN3XX-WpYtRt1hQU1kJu1NT8yeCEMGePrXwCfrgMGdmzjBIY1ZVFuWsrMp0syQ5QwRTg1TVjSBFZYwRFVNLQwYxFVgIyy4Ae2h4aq89cNioQQCkXx9a69KZQhpH35GxFw78BfcVxcnPi-Uv1K4FYkRh3wI7i5048SXHQdvfPZngq-SDL4-9P6EmaCPNuwgtfRf6WdZzjRs.2i4PikKHMd9hp3g-PmWKUk0=,).

TABLE 2. Frequency Analysis by Advertising Network – Top 3[7]

Advertising Network	Frequency	% of Ads^	Alexa Rank (Taiwan)
9tvb.com	200	6.9	1,539
baixing.tw	100	3.4	N/A
adsense.clicking.com.tw	100	3.4	N/A

These ads from 9tvb.com and Yahoo comprised 50% of the identified ad networks. An example is shown in Image1. Alexa.com reports that 25.2% of visitors to 9tvb.com were referred from a Yahoo page.

IMAGE 1. Sample, 9tvb.com

While it is positive that Taiwanese viewers are being exposed to fewer higher risk ads, it is also extremely concerning that they would see almost 40 times more mainstream ads than Australians visiting a similar number of sites. Not all sites were uniformly displaying either mainstream or High Risk ads; the breakdown is given in Table 3.

TABLE 3. Relative proportion of High-Risk vs Mainstream sites per site

Site	High Risk	Mainstream
funshion.com	80.0%	20.0%
tw116.com	0.0%	100.0%
yyets.com	37.4%	62.6%
qiredy.com	0.4%	99.6%
9tvb.com	3.0%	97.0%
tw115.com	70.9%	29.1%
2000mov.com	98.3%	1.7%
dlkoo.com	1.5%	98.5%
370kan.com	100.0%	0.0%
fun698.com	25.4%	74.6%

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

[7] 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.

^ Taken as a percentage of all networks, including identified and non-identified

High Risk Advertising

Table 4 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 gambling and the sex industry were the most popular distinct advertising types. The categories are summarised in Figure 1. An example of malware downloaded from a rogue website is http://www.funshion.com/subject/107222/ - once the user visits this site, and clicks on the "download" button, an executable file FunshionInstall2.8.6.75.exe is downloaded to the user's computer. The page indicates that "popular movies PC client" is required to view the movie or TV episode.

83% of high-risk advertisements were for gambling services.

Running this "PC client" file through the online scanner virscan.org – which analyses suspicious files using 36 different products – the file is verified as malware (DLOADER.Trojan) by DrWeb. A review of the other known filenames associated with similar malware on other rogue 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.

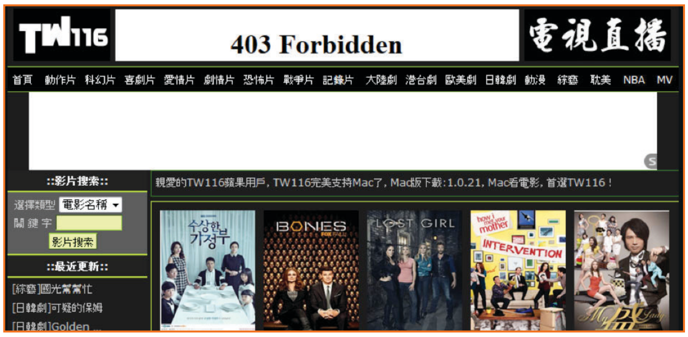
Mainstream Advertising

Results were obtained by visually inspecting every advertisement in the sample to identify whether they 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 many cases, if the browser presents with a non-Taiwanese IP address, localised advertisements will be blocked (see Image 2), 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.

Table 5 shows the results for all mainstream advertisers. The ten most prevalent ads came from Taobao, ibidian.com, Baixing, Pampers, the HKTDC Wine Fair, Maybelline, Anna Sui, Kijiji, Taishin Bank and jd.com. Although these ads were very frequent, a number of ads only appeared once or twice. However, as a prevalence figure, these are still significant, especially since the majority of ads viewed by Taiwanese users are mainstream.

IMAGE 2. Sample, *tw116.com*



403 Forbidden

親愛的TW116蘋果用戶，TW116完美支持Mac了，Mac版下載:1.0.21，Mac看電影，首選TW116！

[綜藝]國光幫幫忙
[日韓劇]可疑的保姆
[日韓劇]Golden ...

TABLE 4. Frequency by Ad Category – High Risk Ads

key: 🟥 Sex Industry 🟧 Scams 🟦 Gambling
🟩 Malware ⬛ Downloading Sites

	🟥	🟩	⬛	🟦	🟧
N	131	53	1	939	4
%	11.6%	4.7%	0.1%	83.2%	0.4%

FIGURE 1. High-Risk Advertising

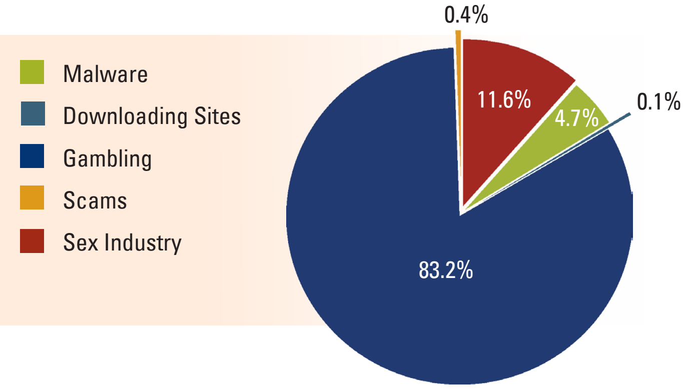


TABLE 5. List of mainstream advertisers

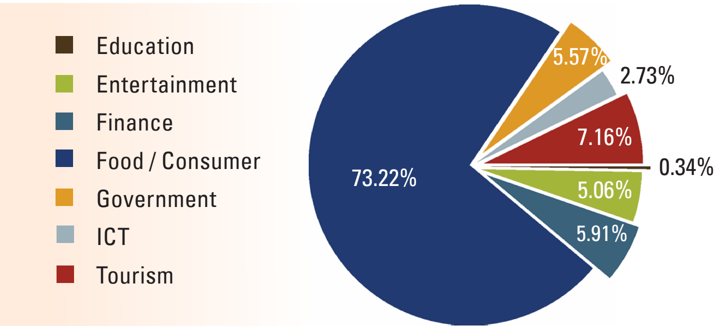
Advertising Network	Frequency	% of Ads
Taobao	300	17.06%
ibidian.com	100	5.69%
Baixing	100	5.69%
Pampers	92	5.23%
HKTDC wine fair	86	4.89%
maybelline	82	4.66%
anna sui	81	4.60%
kijiji	78	4.43%
taishin bank	66	3.75%
jd.com	64	3.64%
Garnier	57	3.24%
Neutrogena	56	3.18%
kbro	54	3.07%
KGCheck	41	2.33%
DBS Bank	36	2.05%
yes123	30	1.71%
Evergeen	29	1.65%
sxd.txwy.tw	28	1.59%
Nan Ji Ren	24	1.36%
Clarins	22	1.25%
Supor	20	1.14%
dettol	20	1.14%
lovebeauty.com.tw	17	0.97%
Michelin	17	0.97%
pazzo.com.tw	15	0.85%
Dept Labor Affairs	12	0.68%
Anixis	12	0.68%
Mini	10	0.57%
vipshop	10	0.57%
Lancome	9	0.51%
Vichy	8	0.45%
Honda	8	0.45%
104ehr	8	0.45%
P&G	7	0.40%
lifevc.com	7	0.40%
Corpo X	7	0.40%
surveycompare	7	0.40%
NBA	7	0.40%
Persil	7	0.40%
New Contemporary Dental Clinic	6	0.34%
LINE	6	0.34%
g1.com.tw	6	0.34%
pccenter	6	0.34%
acer	6	0.34%
Sony	6	0.34%
Ikea	6	0.34%
TutorABC	5	0.28%
Golden Ladies Photo	5	0.28%
poba.com.tw	5	0.28%
On Foot Travel	5	0.28%
BBQueen	5	0.28%
Huggies	5	0.28%
Beauty UP Clinic	5	0.28%
Mr ING	4	0.23%
oderbau	3	0.17%
GlaxoSmithKline	3	0.17%
FarGlory Hotel	3	0.17%
renrendai.com	2	0.11%
LV Shou	2	0.11%
hitachi	2	0.11%
ksmombaby-fair	2	0.11%
IBQ	2	0.11%
am pm skincare	2	0.11%
naruko	2	0.11%
einfach	2	0.11%
Calcisure	2	0.11%
qunar.com	1	0.06%
Toyota	1	0.06%
tanghouse	1	0.06%
wbiao.cn	1	0.06%
abtech.com.tw	1	0.06%
Lattice Semiconductor	1	0.06%
Cooper Tires	1	0.06%
Happi Go	1	0.06%
indexedu	1	0.06%
beddingworld	1	0.06%
sanlien	1	0.06%
sb-ele	1	0.06%
Agoda	1	0.06%
Skyview	1	0.06%
Audi	1	0.06%
Numanni	1	0.06%
Biochem	1	0.06%

Table 6 shows the overall industry category and prevalence rates. Interestingly, almost 4 out of 5 ads were for consumer purchases and food, especially cosmetics, motor vehicles, and sanitary products. Next most prevalent were tourism sites, followed by entertainment, finance and ICT. Disappointingly, the Taiwanese government was responsible for 0.68% of mainstream ads in the sample, and the Hong Kong government (through the HKTDC Wine Fair) was responsible for 4.89%. Figure 2 shows the breakdown by industry category.

TABLE 6. Mainstream advertisers by category

Industry	Percentage %
Education	0.34%
Entertainment	5.06%
Finance	5.91%
Food/Consumer	73.22%
Government	5.57%
ICT	2.73%
Tourism	7.16%

FIG.2 Mainstream ads by industry category



Validation Study

A small validation study was conducted to ensure that there was no bias in selecting Hollywood advertisements over local Taiwanese content. For each of the sites from which a Hollywood movie/TV show was sampled, one local Taiwanese language movie was selected from the home page of each site. In 80% cases, these pages contained mainstream ads, often for the same advertisers as per the Hollywood content, including Taobao and Baidu, as shown in Table 7.

TABLE 7. Validation Study

Site	Mainstream Advertisers
www.funshion.com	Taobao
www.tw116.com	Standard Chartered
www.yyets.com	Baidu
www.qiredy.com	Ci Yuan Ge Fortune Telling
www.9tvb.com	Martell Whisky
www.tw115.com	Rio Clinic (via Yahoo! Ads)
www.2000mov.com	-
www.dlkoo.com	Taobao
370kan.com	-

CONCLUSION

Unlike other countries examined, mainstream ads comprise the dominant form of advertising on rogue sites as viewed in Taiwan.

Taiwanese are at a high risk of being exposed to advertisements for the sex industry, malware, scams and gambling if they visit rogue websites. Such advertisements pose a real risk to the mental wellbeing of the Taiwanese people, and appear to be inconsistent with Taiwan's social policies. Yet mainstream ads were dominant, indicating the need for further regulation of online advertising in Taiwan as is the case for print and TV media. Allowing unscrupulous companies to build their brand by directly supporting rogue sites, who thereby make enormous profits, indicates the lack of effective regulation. Sadly, even the Taiwanese government was observed to be advertising on rogue sites.

The key findings from the analysis of the Taiwanese data set are discussed below:

- 39% of the ads were High-Risk;
  61% were Mainstream.
- In the High-Risk ads, the overwhelming majority were for gambling (83%), followed by the sex industry, malicious or suspected malicious code, and scams of various kinds, including premium rate SMS, investment and employment scams. These results were quite different to other countries; variations may be due to the implementation of ISP or government filtering regimes, and local customs and interests
- The top ad networks serving ads to Taiwanese users, as identified using Adblock, were 9tvb.com, baixing.tw and adsense.clicking.com.tw; these were different for the top advertising networks for Australian and Singaporean ads. Note that the majority of ads displayed were not served up by ad blocks identified through the Adblock list; further research is needed into identifying these local networks, and passing the details onto Adblock so that the list can be used for Taiwanese to more effectively block ads.
- A significant number of household name brands in Taiwan are choosing to advertise on rogue sites, which are facilitating the distribution of infringing content (movies and TV shows). 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?
- Household names from the top Mainstream advertisers included industries such as ICT, Finance, Tourism, Entertainment, Education, Government, and Food/Consumer. The latter comprised the majority of ads displayed, including brands such as Taobao, Anna Sui, Pampers, Neutrogena, Dettol, Maybelline, Taishin Bank, DBS Bank, Clarins, Honda and Mini. Taiwanese government ads had a prevalence rate of 0.68%.

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

- Taiwanese have a greater chance of viewing Mainstream ads compared to Australians or Singaporeans, but they are still at risk from ads served by rogue sites, which pose a real danger to viewers.
- No ads appeared to be filtered in Taiwan, with no "blank" ads appearing in screenshots where an ad should have been visible. Technical controls to block image and text ad content could be explored (eg, Ho & Watters, 2004), perhaps by enhancing Adblock.
- Advertisers need to have better mechanisms to control where there 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. ecent Sets of best practice guidelines for ad networks to address piracy and counterfeiting have recently been released in the US (Dredge, 2013) and UK (DTSG, 2013), with most of the world's major web companies participating.   Other sets of guidelines are under development in Europe, notably including Germany. Advertisers recently succeeded in pressuring Facebook, for example, to remove offensive advertisement by threatening to remove their ads as a group (Cellan-Jones, 2013).

- Other types of rogue content have been managed effectively by legal sanctions in the past. For example, search results for pharmaceuticals without prescriptions (O'Donnell, 2013) were removed by Google after they paid a very significant fine, given that illicit drug distribution is a growing problem online (Watters & Phair, 2012). A number of fake "Viagra" and "Cialis" ads were detected; what would be the consequences of citizens ordering and consuming such fake medicines?

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 who 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 code of conduct[8]. Also, no additional burden should be placed on rightsholders to police the internet for offensive material.

[8] 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/

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/

DTSG (2013). UK Good Practice Principles for the trading of Digital Display Advertising. Downloaded from http://www.jicwebs.org/digital-trading-standards-group-good-practice-principles/good-practice-principles

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

Google (2013). How Google Fights Piracy. Downloaded from https://docs.google.com/file/d/0BwxyRPFduTN2dVFqYml5UENUeUE/edit

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. (2013). Measuring Surveillance In Online Advertising: A Big Data Approach. 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 (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. (2013a). A Systematic Approach to Measuring Advertising Transparency Online:

An Australian Case Study. ICSL Technical Report, downloaded from http://www.icsl.com.au/files/ICSL_report_digital.pdf

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. (2013c). 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. (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

Acknowledgements:

4Oditiscideste quidunt aut ius vid utet am eaquam suntius, nesciat intium ex et eaquas autFicatum ute simagnam, sum faccae omniet adignis erist plitem ut fuga. Aquia corepuditas enim sincto verundu cimagnam, corepratio corum siniendamet et fugias a vollorro