



USC Annenberg Lab Ad Transparency Report

January 5, 2013

This is the first in a monthly series of reports from the Annenberg Innovation Lab at the University of Southern California that details support by Online Ad Networks of the major pirate movie and music sites around the world. The advertising business has always been a key part of the creative economy from the birth of radio. Advertising dollars have financed the production of television, music and even video games. The rise of ad-supported pirate networks is a relatively new phenomenon stemming from the birth of peer to peer (P2P) Internet sites in 2001. In the last five years, a large number of new advertising networks now service the seemingly infinite advertising inventory of the broadband era. Much of that inventory sits on more than 150,000 pirate entertainment sites. The top ten advertising networks placing the most advertisements to Illicit file sharing sites are:ⁱ

1. [Openx](#)
2. [Google](#) (including Double Click)
3. [Exoclick](#)
4. [Sumotorrent](#)
5. [Propellerads](#)
6. [Yahoo](#) (including Right Media)
7. [Media Shakers](#)
8. [Yesads](#)
9. [Infolinks](#)
10. [tribalfusion.com](#)

The list of top infringing sites was compiled using the [Google Transparency Report](#) of domains with the most Digital Millennium Copyright Act (DMCA) takedown requests.

A recent report, "[The Six Business Models of Copyright Infringement](#)", funded by Google and the firm PRS for Music on Brands, investigated advertising networks and their support of the major pirate movie and music sites around the world. The report found that advertising financed 86% of the P2P search sites that feature illegally distributed content. This finding clearly indicates that many major brands are not aware that they are, in fact, the key source of funds for the piracy industry. It is the goal of this "transparency report" to aid in helping these brands steer their advertising dollars away from sites that exploit film, television and music artists for what appears to be criminal gain.

The report, which will name the top offenders on a monthly basis, was compiled from multiple sources and will identify the top ten advertising networks that place the most advertisements on illicit file sharing sites. It makes use of an Internet bot that scrapes the Ad Network HTML identifier contained in each advertisement.

The screenshot shows a web browser window with the address bar displaying 'search.4shared.com/q/1/Call+Me+Maybe'. The page content includes a large advertisement for a 'Jeep 2013 GRAND CHEROKEE LAREDO 4X2' with a price of '\$299 MONTH FOR 36 MONTHS'. Below the ad, there are search results for 'call me maybe.mp3', 'carly rae jepsen call me maybe.jpg', and 'fun. - call me maybe (cover).mp3'. The browser's developer tools are open at the bottom, showing the 'HTML' tab. A red arrow points to a specific line of code in the HTML structure: '

A list of the top ten brands appearing on illicit file sharing sites will appear in the coming months.

Methodology

Advertising network code appearing in the HTML and Javascript URL locations where alleged infringement were occurring was collected from third party sources and from data contained in the Google Transparency report on an ongoing basis for the past twelve months. This data included a vast number of URL locations where alleged infringement were occurring, including youtube.com and blogspot.com data that is not included in the Google Transparency report. This data contains the precise URL location of the alleged infringements. Using server software the HTML and Javascript contained in the precise URL location was captured. Where possible a screen shot of the page was captured. Using the server software, the captured HTML and Javascript was analyzed to determine if there was Advertising Networks Code appearing at those URL locations. The Domain Name represented in the Advertising Network Code associated with the captured HTML and Javascript was then used to determine, to the best of our knowledge, the Advertising Network distributing the advertisements, if any, appearing at the URL location where the alleged infringement was occurring. It is our observation that advertisements flowing from Advertising Exchanges are essentially fungible and the Advertising Network originating the Advertisement may not be the Advertising Network delivering the Advertisement on the page where the Advertising Network Code was detected. In some instances, we used network protocol analyzers to determine the actual server locations where the Advertisements were originating.

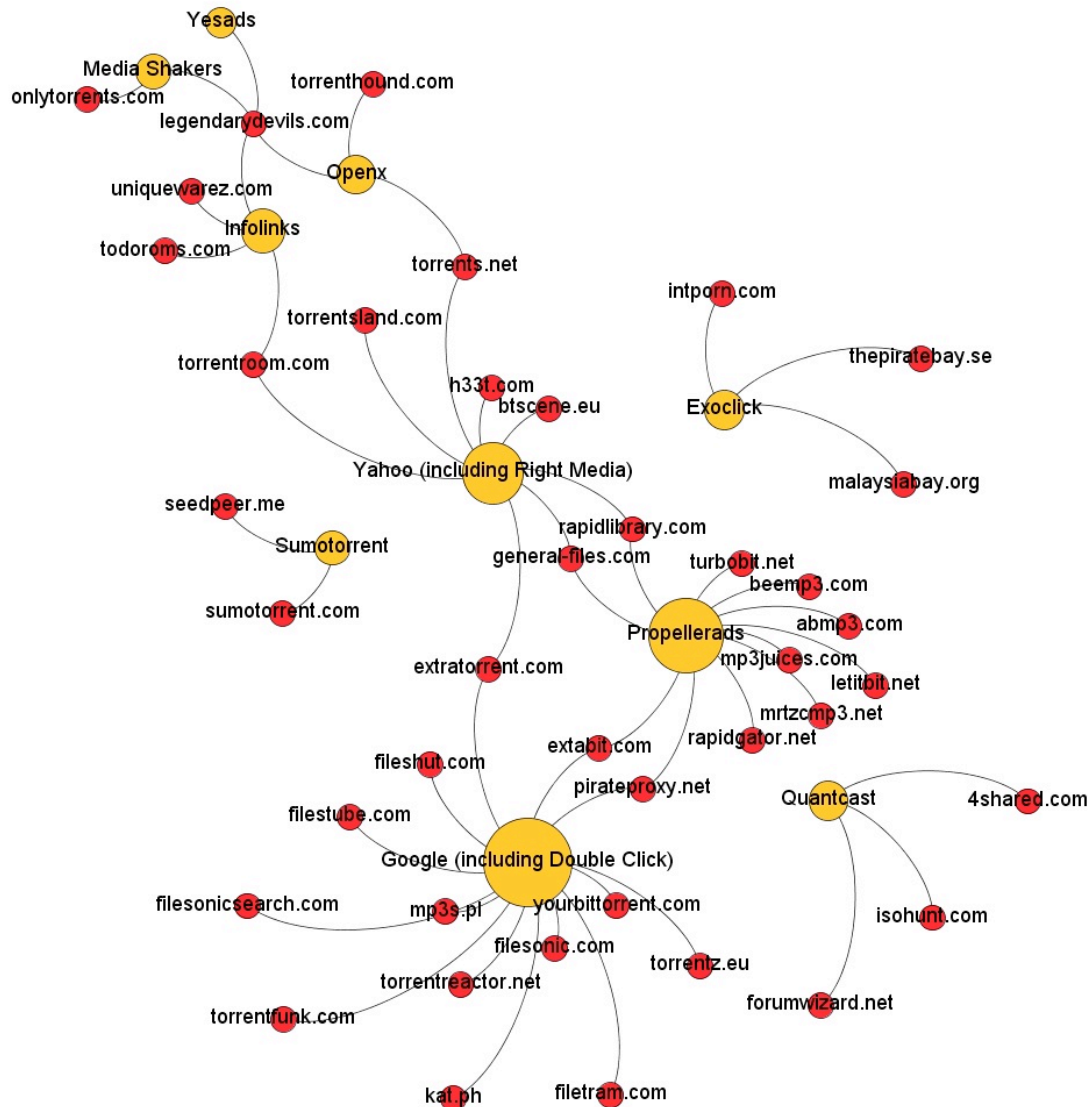
In late December 2012, domains appearing in the alleged infringing sites in the Google Transparency Report were then compared to determine the frequency of Advertising Networks that appeared on these domains for the preceding twelve months. We do not dispute that the landscape may be different today but the focus of this study was to take a historical snapshot now that the Google Transparency Report provides more direction to determine what websites are alleged to be trafficking in stolen intellectual property. We should also note that Google has chosen not to include alleged infringement data for youtube.com and blogspot.com in the Google Transparency report. In our estimation, the inclusion of that data would provide a more accurate view of the landscape.

Going Forward

With the limited tools available and the lack of accessible information to make absolutely definitive determinations, the study relied upon techniques that produced a result that in our best estimation represents the pattern of observed activity. We look forward to working with industry participants to improve our methodology. The data produced by the Google Transparency report contains a vast number of indications of alleged infringement. In development is technology to ingest the entire report as it is updated for real-time processing and reporting.

We are aware of the critique made by Google upon the original release of our findings. They wrote, "In addition to sites that participate in our network, millions of advertisers and publishers use our DoubleClick technology to manage their digital advertising, not just on our network but across the whole web. Advertisers and publishers ultimately decide how to use this technology and we cannot "see" where all these ads appear (nor do we have a revenue share). However, when we do become aware of DoubleClick technology powering ads on copyright-violating sites, we contact the affected advertisers and publishers to take action." However, we need to point out that the instances of Double Click ad code in our data was less than 9% of the total instances of Google advertising code found on the illicit sites. Most of the Google code was from the domains googleadservices.com, googletag.pubads and googlesyndication.com. We are however encouraged that Google has recently increased it's efforts to let advertisers know when their ads are appearing on illicit sites. We assume that effort will be reflected in the January data that we will publish in early February.

Data Visualization



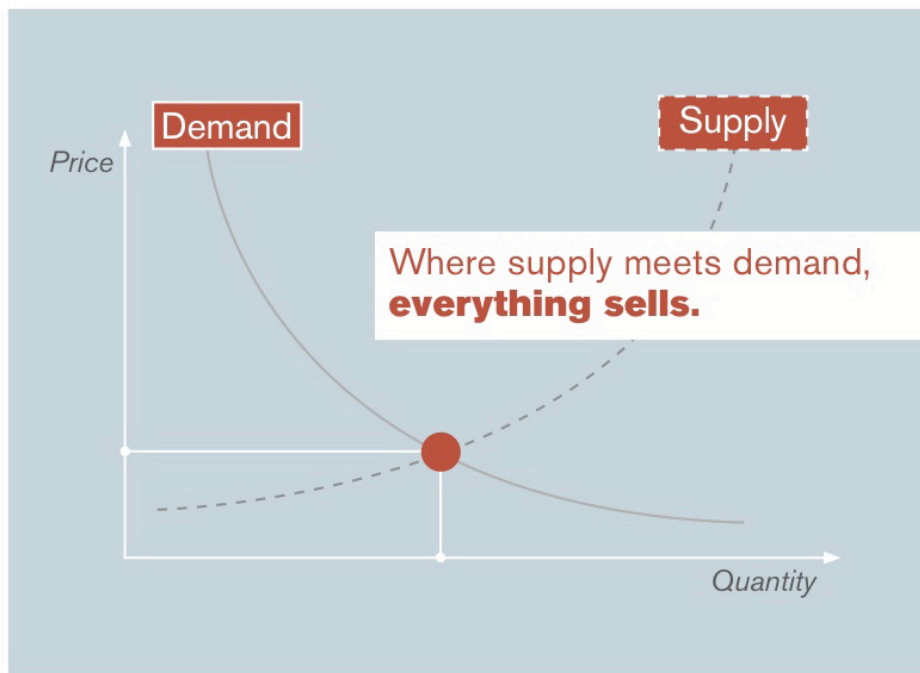
This network graph demonstrates the connections between the top ten advertising networks and some of the larger illicit file sharing sites included in the sample presented in the Annenberg Innovation Lab Transparency Report. It serves as an alternative way to conceptualize the data at the heart of the examination.

Advertising networks are presented in orange and illicit file sharing sites in red. Connections between the nodes represent the presence of ads on the associated sites. Node size for each advertising network is scaled to the number of illicit file sharing sites

each is linked to, the more connections the bigger the circle. More connected advertising networks are situated closer to the center of the graph.

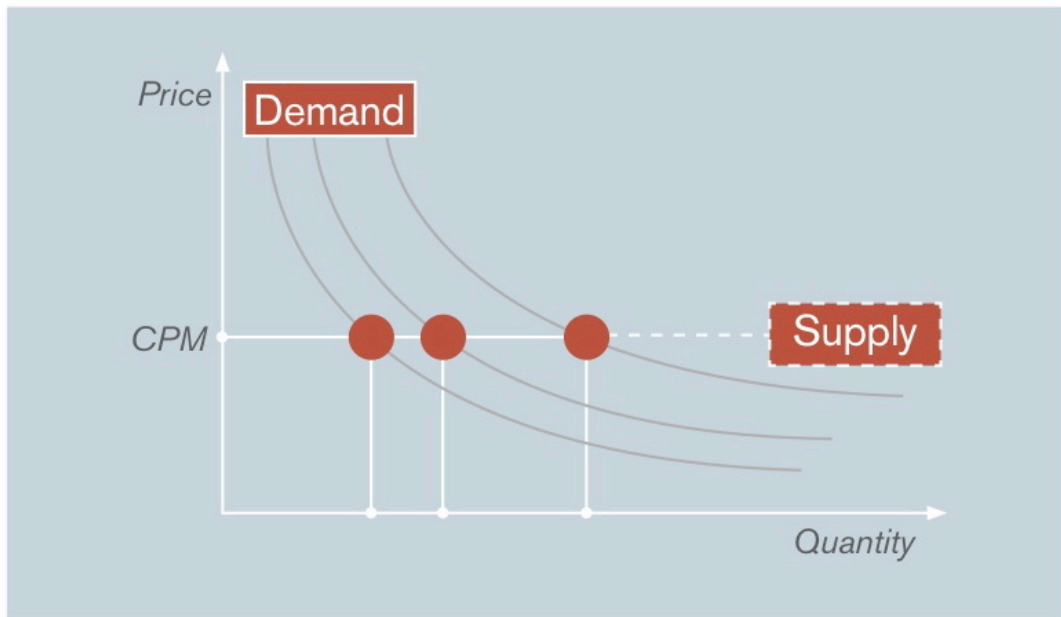
Findings

In the most recent Google Transparency Report, Filestube.com had more than 2,300,000 specific URLs cited for copyright infringement. It appears that large illicit file sharing sites distribute illegal content and continue to steal trademarked, copyrighted content and siphon millions of dollars away from the creative community, making it much harder for artists to make a living. Unlike the legitimate ad supported content sites like Spotify, Pandora or Hulu, the illicit file sharing sites do not return capital to create new content or sustain the creative economy. It appears that many brand advertisers are either unaware that their ad dollars are financing a parasitic business or have decided that the advantages to be gained by targeting a young demographic on illicit file sharing sites trumps their corporate social responsibility. But in a deeper way, the almost infinite ad inventory sitting on the 150,000 illicit file sharing sites named by Google harms the entire media industry by devaluing the worth of legitimate ad inventory. An August 2012 report from Comscore, “The Economics of On Line Advertising” delineates the problem.



In Adam Smith’s classic economy (above) supply is limited and at a certain clearing price everything sells. But as Comscore points out: “In digital media, the marginal cost of adding

new programming or advertising inventory is very low, and has been trending lower for years, as the costs of computer processing and storage have plummeted.” This produces a new supply-demand equation.



In this world, the unlimited supply of ad units on illicit file sharing sites devalues all advertising inventory on the Internet.

The Annenberg Innovation Lab does not propose any particular solution to this dilemma. We do believe that some sort of best practices code of conduct by brands, agencies and advertising networks would be productive. As a prominent Google executive has told us, “following the money is exactly the right focus”. As we have been tracking more than 100 advertising networks over the course of this research, we have come to believe that the lack of transparency in the advertising network industry is a classic case of market failure. It is perhaps a good place for academic research. We believe that both the prominent advertising networks and the major brands do not want to be perceived “to be evil.” In that regard, any mutual assistance we might provide to the industry is freely offered.

Contact: USC Annenberg Innovation Lab

3502 Watt Way

Los Angeles, CA 90089

213-740-0476

annenberglab.org

ⁱ In late February we have had productive talks with Quantcast about our January and February Ad Reports. They have given us access to code, which differentiates between their free, publicly available, self-serve measurement product and their actual revenue-generating advertising product.

We now believe that Quantcast was incorrectly identified as being among the top ten Ad Networks placing ads on infringing piracy sites. Quantcast is undertaking extensive efforts to maintain a global block list of infringing sites, and in the course of our discussions, we were able to help them find and block two additional sites.

Quantcast and USC Annenberg Innovation Lab intend to continue working together to ensure that ads are not placed on infringing sites.