Joint Industry Committee for Web Standards

JICWEBS

UK Traffic Taxonomy for Digital Display Advertising

Version 1 Issued October 2015

CONTENTS

1. Introduction

JICWEBS has one clear aim - to promote trusted digital ad trading. A key element of this is improved transparency (and risk reduction) in the area of digital ad fraud. This taxonomy document is a key step in support of that.

The JICWEBS Cross-Industry Anti-Fraud Working Groups were established in December 2014 with the purpose of creating cross-industry guidelines and principles to educate the wider market to reduce the risk of exposure to ad fraud, creating a safer, more transparent supply chain.

JICWEBS' steps to "anti-fraud" in digital advertising are:

1. Industry guidance on risk reduction (issued June 2015)

2. Taxonomy of types of traffic (this document)

3. Good Practice Principles for risk reduction

4. Independent review (and resultant "seal") for organisations demonstrating application of these Good Practice Principles

This taxonomy was originally created by the US Trustworthy Accountability Group (TAG), and has been revised by the UK Anti-Fraud Commercial and UK Anti-Fraud Technical Working Groups[1].

This document serves to identify sources of non-human traffic, which includes both malicious (or fraudulent) and non-malicious sources of traffic. Non-malicious sources of traffic include spiders and bots such as the Googlebot that help to index the web and enable Search Engines. A comprehensive list of these accepted spider and bots is maintained and updated by the IAB and ABC[2]. It is a detailed but not exhaustive list and should be used as a guide for those involved in the buying, selling and serving of digital display advertising. The intention will be to update this list on an ongoing basis when required.

[1] See Appendix 1

[2] See Appendix 2

2. Non-human Traffic Taxonomy

For the purposes of this guide the term 'fraudulent' is not intended to represent fraud as defined in UK law or as conventionally used in U.K. Court or other legal proceedings, but rather a custom definition strictly for advertising measurement purposes.

Note: the following taxonomy is not necessarily mutually exclusive.

ILLEGITIMATE AND NON-HUMAN TRAFFIC SOURCES

Description		Fraudulent / Potentially Fraudulent / Not fraudulent
1. Hijacked device – any user's device (browser, phone, app or other system) that has been modified to call html or make ad requests that is not under the control of a user and made without the user's consent. This includes:	**Hijacked device with a fully automated browser –** a hijacked device where the device is a browser and the modification is that the browser is hidden from user view and engaged in making html or ad calls.	Fraudulent
	Hijacked device with session hijacking – a hijacked device where a user is present and additional html or ad calls are made independently of the content being requested by the user. Ads and redirections are inserted into the user experience by the program running on the device.	Fraudulent
2. Crawler masquerading as a legitimate user – a browser, server or app that makes page load calls automatically without declaring themselves as a robot, instead declaring a valid regular browser or app user agent where there is no real human user**.**	**Advanced –** declares a user agent string normally associated with human activity, and also renders the page.	Fraudulent
	Basic – only declares a user agent string normally associated with human activity, does not render the page.	Fraudulent
3. Data-centre traffic – traffic originating from servers in data-centres, rather than residential or corporate networks, where the ad is not rendered in a user's device (there is no real human user).	-	Potentially fraudulent. However legitimate uses do exist.

NON-TRADITIONAL / OTHER TRAFFIC

Description		Fraudulent / Potentially Fraudulent / Not fraudulent
4. AdWare traffic – a device where a user is present and additional html or ad calls are made by the AdWare independently of the content being requested by the user.	-	Potentially fraudulent. Could be legitimate human traffic under certain strict criteria
5. Proxy traffic – traffic that is routed through an intermediary proxy device or network where the ad is rendered in a user's device where there is a real human user. This includes:	**Proxy traffic that is anonymized** – Proxy traffic where the call is anonymized. (e.g., Tor)	Potentially fraudulent
	Proxy traffic that is not anonymized – Proxy traffic where the call is not anonymized.	Potentially fraudulent
6. Non-browser User-Agent header – a device that declares a User-Agent header not normally associated with human activity.	**Non-browser User-Agent header App traffic** – a device that declares a non-standard or invalid User-Agent header that is sold as app traffic	Fraudulent
	Non-browser User-Agent header Non-app traffic – a device that declares a non-standard or invalid User-Agent header that is not sold as app traffic.	Fraudulent
7. Browser pre-rendering – a device that makes html or ad calls prior to the rendering of the resulting assets or web-page to an end user.	**Browser pre-rendering, un-rendered –** Browser pre-rendering calls where the page never exits the pre-rendering state. For example, the process by which the Safari browser creates thumbnails for its new tab page.	Potentially fraudulent if device is hijacked. Not fraudulent if user initiated.

HIJACKED TAGS

Description		Fraudulent / Potentially Fraudulent / Not fraudulent
8. Ad Tag Hijacking - Taking ad tags from a publisher's site and putting them on to another site without the publisher knowledge.	-	Fraudulent
9. Creative Hijacking - Copying the creative tags from a legitimately served ad so they can be rendered at a later time, without the consent of the advertiser or their contracted service provider.	-	Fraudulent

SITE OR IMPRESSION ATTRIBUTES

Description		Fraudulent / Potentially Fraudulent / Not fraudulent
10. Auto-refresh – a page or ad unit that calls for a new rendered asset more than once.	**Declared minimum interval –** Auto-refresh where the minimum time interval between calls is declared explicitly.	Potentially fraudulent
	Declared minimum interval with user interaction – Auto-refresh where the minimum time interval between calls is declared explicitly and user interaction with the page is detected at the time of refresh.	Potentially fraudulent
	Undeclared – Auto-refresh without any declaration of time or user interaction.	Potentially fraudulent
11. Ad Density – the number of ads or percentage of the page / app covered by ads	**Number of ads –** the ad density where the number of ads is declared	Potentially fraudulent
	Percentage of page – the ad density where the percentage of the page /	Potentially fraudulent

	app covered by ads is declared.	
	Undeclared – the number of ads or percentage of the page / app covered by ads is not declared	Potentially fraudulent
12. **Hidden Ads** – ads placed in such a manner that they cannot ever be viewable e.g., stacked ads, ads clipped by iframes, zero opacity ads.	-	Fraudulent
13. **Misappropriated Content**	**Links –** site contains links to copyrighted content but does not have the content itself	Potentially fraudulent
	Content – site contains copyrighted content (from another, unaffiliated entity) without the rights to monetize such content	Potentially fraudulent
14. **Falsely represented** – sites or impressions represented as one thing that are another, including	**Context** – HTML or ad calls that attempt to represent another site or device or other attribute, other than the actual placement e.g., referrer spoofing	Fraudulent
	Intention – a human user that is offered payment or benefits to view or interact with ads who is represented as not being offered payment or benefit.	Fraudulent
15. **Contains malware –** malware is found on the site, or the app contains malware.	-	Potentially fraudulent

AD CREATIVE / OTHER

Description		Fraudulent / Potentially Fraudulent / Not fraudulent
16. Cookie-stuffing – The process by which a client is provided with cookies from other domains as if the user had visited those other domains.	-	Fraudulent

3. Contact

Any questions or queries should be directed to JICWEBS (info@jicwebs.org) or to your representative trade body. For more information visit www.jicwebs.org.

JICWEBS

Created by the UK media industry to ensure independent development of standards for benchmarking best practice for online ad trading

The Voice of British Advertisers ISBA

IPA

iab Internet Advertising Bureau UK

aop Association of Online Publishers

newsmedia association

Debate Define Deliver

JICWEBS

Appendix 1

The UK JICWEBS Anti-Fraud Commercial Working Group is a committee of industry experts across the industry from the following companies:

ABC
Association of Online Publishers (AOP)
Adloox
Forensiq
FT
Google
GroupM
Incorporated Society of British Advertisers (ISBA)
Institute of Practitioners in Advertising (IPA)
Integral Ad Science
Internet Advertising Bureau UK (IAB UK)
Internet Advertising Bureau UK (IAB UK)
PHD Rocket
Reckitt Benckiser
Santander
Shell
Telemetry
Unilever
Videology
VivaKi
Yahoo

The UK JICWEBS Anti-Fraud Technical Working Group is a committee of technical experts predominantly from fraud detection vendors and from companies with a leading interest in fraud detection:

ABC
Adloox
AppNexus
Association of Online Publishers (AOP)
comScore
Crimtan
Forensiq
Forensiq
FT
Google
GroupM
Incorporated Society of British Advertisers (ISBA)
Institute of Practitioners in Advertising (IPA)
Integral Ad Science
Internet Advertising Bureau (IAB)
Meetrics
Microsoft
Moat
PHD Rocket
Pixalate
Reckitt Benckiser
Rocket Fuel
Santander
Shell
Sizmek
Telemetry
Videology
VivaKi
WhiteOps
Yahoo
YuMe

Appendix 2

About the International IAB/ABC Spiders & Bots List

Implementation of the International IAB/ABC Spiders & Robots List is a key step to prevent non-human traffic being counted in web analytics.

This comprehensive list, updated monthly, enables filtering of non-human activity that can significantly inflate ad impression and site traffic counts. The end result is a more transparent and accurate measurement for ad impressions and site traffic claims.

This service is available to any business, whether you are an ABC member or not.

Further Enquiries

For further enquiries about how to access this list please contact ABC at enquiries@abc.org.uk

Spiders & Robots Policy Board

A Spider & Robot Policy Board has been formed to oversee and approve list modifications (see list below).

Albert Roux, Microsoft
Andreas Piras, Microsoft
Brendan Riordan-Butterworth, IAB
Charlie Stafford, Weather.com
Derek O'Loughlin, Microsoft
George Ivie, Media Rating Council
Greg Taylor, IMServices Group
Mark Kourey, 24/7 Media
Mark Thielen, Adtech
Martin Liljenback, Adobe
Michael McNally, Google
Mike Swope, WebTrends.com
Per Bjorke, Google
Philippe Rivard, Google
Piotr Piwowarczyk, Yahoo!
Richard Bennett, ImServices Group
Richard Foan, ABC
Richard Thurman, AOL
Ron Calhoun, Yahoo!
Sherwette Mansour, Yahoo
Slawomir Krysiak, CBS Interactive
Steve Guenther, ImServices Group
Umesh Chandwani, AdMarvel