



Realmatch Inc. Traffic Quality Declaration

Table of Contents

1. Background

The problem is worth solving

About this document

Goal of this Declaration

Traffic Quality Determinants

2. The Declaration

Guidelines for Publishers

HUMAN vs. NON-HUMAN

GEOGRAPHY

DUPLICATION & FREQUENCY

3. Remediation

4. Education

Background

As recruitment marketing evolves toward interconnectivity and interdependence of advertisers, publishers, and technology solution providers, so grows the importance of establishing standards and transparency in how job advertising traffic quality is defined and valued.

In the absence of standards, measuring the effectiveness of job advertising media spend across publishers and analytics suites - whether purchased under duration, subscription or a pay-for-performance (CPM, CPC, CPA) payment model – is difficult if not impossible.

The problem is worth solving

1. Almost half – 48.5% - of all online traffic is not human (Imperva Incapsula Bot Traffic Report, 2015) - do bot clicks count?
2. For the U.S. market and for European markets, does foreign traffic count? For example, 16.5% of Austrians want a job outside of Austria (Indeed EU15 Movers Index, 2015), but for many employers, they would not be considered viable candidates for jobs advertised in U.S. or other European markets.

These are just two examples of how measurement differences affect the understanding of traffic quality.

About this document

The TAtech Traffic Quality Declaration is a voluntary publisher self-evaluation. It was developed by an international Working Group established by TAtech: The Association for Talent Acquisition Solutions to bring standards and transparency to traffic measurement for the job advertising industry. The Declaration is endorsed by TAtech as its official position on the importance of establishing standards for defining and measuring traffic quality.

Employers (Advertisers) deserve

1. **Transparency** - understanding what traffic counts and what doesn't is the foundation for evaluation.
2. **Honest and fair reporting** - reports that accurately describe the value that has been delivered.

Publishers Deserve

1. **A fair playing field** - empowering employers to make “apples to apples” value assessments through transparency.
2. **Industry standards** - a benchmark to be compared with and against that is endorsed and which can be implemented across the industry by publishing peers.

Goal of this Declaration

This Declaration aims to build understanding and trust amongst recruitment advertising buyers and sellers by:

1. Establishing a common set of terms and terminology for discussions about traffic quality, see Traffic Quality Determinants below.
2. Providing a framework for self-disclosure of traffic quality practices employed by publishers and technology solution providers (analytics, ad engines, etc)

3. Leveling the 'playing field', to facilitate equitable measurement of all job advertising media regardless of whether the transaction is unit (duration), time (subscription) or performance (CPC, CPA)-based.

See the *Education* section of this document below for additional background information and useful resources for further understanding.

Traffic Quality Determinants

Technographic signals - involve inspecting the unique technical aspects of each traffic event

1. **Human vs. Non-Human** - is the advertising action driven by a person or a machine?
2. **Geography**- is the physical location of the user likely relevant based on geographic proximity to the advertised need?

Behavioral signals - involve looking at the behavior of a given user over time

1. **Duplication and Frequency** - are repeated actions driving value? Is quality associated with the amount of time between actions?

The Declaration

Guidelines for Publishers

What you do vs. How you do it - In order to provide transparency without revealing information that could enable bad actors to avoid detection, describe your methods in general terms without providing vendor or tactical details. For example, say that you use software to evaluate non-human User Agent signals, but don't name the vendor or libraries used.

Flag - To provide consistency across publisher declarations, use the term 'flag' to mean the process of identifying the particular activity.

Transaction Model - Unit (duration) and time-based (subscription) job advertising providers should generally answer N/A for the 'Billable' portion of the response. 'Viewable on Reports' applies to all transaction models.

PUBLISHER NAME:

Realmatch, Inc

HUMAN vs. NON-HUMAN

Do you differentiate human vs. non-human advertising activity?

1. **User Agent** - do you analyze user agent contents?

Parsing user agent strings for declarative terms such as 'bot', 'crawler', 'spider', etc. and indicative terms such as 'phantomjs' are useful in the identification of traffic that is not from humans.

Realmatch is utilizing big data analysis to identify bots in real-time. Realmatch analyzes all user agent strings and compares them to a blacklist containing patterns that were identified in the past or to common user agent strings. (e.g. 'crawler').

New and suspicious string patterns are surfaced for a manual review process. The manual process updates the blacklist.

Billable : N/A

Viewable on Reports : No

2. **IP Address Filters** - Do you employ manual or automated traffic filtering based on the number of times that individual IP addresses drive advertising actions?

Filtering IPs that execute 100 clicks without any conversions would indicate a likelihood that the IP is not acting like a human.

Realmatch analyses all IP Addresses and Subnet addresses using a pattern recognition technology to identify non-human originating traffic.

The technology constantly observes these two factors:

1. Frequency of clicks
2. Transaction (i.e. funnel) click trough.

All bot related IP Addresses and Subnet Addresses are stored in a form of a blacklist. Any traffic that matches the blacklist is removed in real time.

Billable : N/A

Viewable on Reports : No

3. **IP Hosts** - Do you employ manual or automated filtering based on characteristics of the hosting service provider?

Certain hosts have low likelihood of being used by people who are interested in particular job advertisements. Examples include Amazon AWS, Digital Ocean, Cloud Sigma, etc.

Realmatch analyses all IP Host names using a pattern recognition technology to identify non-human originating traffic. The technology uses a blacklist to look for known IP Hosts such as cloud services (i.e. Amazon AWS). New IP Host names will be observed based on these two factors:

1. Frequency of clicks
2. Transaction (i.e. funnel) click trough.

All bot related IP Host names are stored in a a blacklist.
Any traffic that matches the blacklist is removed in real time.

Billable : N/A

Viewable on Reports: Yes

4. **IP Geolocation** - Do you employ manual or automated filtering based the IP proximity of repeated actions?

Sophisticated non-human activity can involve the use of IP Proxies that individually do not signal potential non-human activity, but when looked at as a group of actions from a single location, a more indicative pattern emerges.

Realmatch does not implicitly employ a filtering based on IP Geolocations rules.
Realmatch can report to customers' the traffic source location.

Billable : N/A

Viewable on Reports: Yes

5. **Explicit Validation** - Do you employ reCAPTCHA or other automated validation solutions such as mouse movement to confirm human activity?

[Google reCAPTCHA](#)

[Wikipedia](#)

Realmatch does not employ reCAPTCHA or other automated validation solutions.
The nature of the job application process requires several steps of user engagement, including upload of a resume file that is very hard for a non-human to accomplish.

Billable : N/A

Viewable on Reports: No

GEOGRAPHY

How do you distinguish traffic by geographic proximity?

6. **Location Relevance** - do you employ detection practices to determine geographic relevance?

The physical location of the user viewing and taking action on the ad makes a difference. For example, is the user clicking on a U.S. job listing from Russia or known to reside in another location (distant from the job posting location) based on profile/resume info?

Realmatch does not implicitly employ detection practices to determine geography relevance.
Realmatch can report to customers' the traffic source location.

Billable : N/A

Viewable on Reports: Yes

DUPLICATION & FREQUENCY

How do you distinguish traffic by frequency and duplicity?

7. **Duplicity / Frequency** - do you distinguish traffic by the number of times a user interacts with an advertisement and/or how much time elapses between interactions?

It's important to measure the number of times that a single user interacts with an ad. For example, if the same user clicks on the same job 10 times. or if a single user interacts with 10,000 ads, it's unlikely that that is legitimate job seeking activity.

It's also important to measure the time span between ad interactions. Is it different if a user clicks on the same job 5 times in a minute compared with once per week over the span of 5 weeks? What if a user clicks on the same ad twice in 2 seconds?

Realmatch distinguishes traffic based on the number of times a user interacts with the web site.
Realmatch uses a set of thresholds that "flag" suspicious interactions.
These thresholds are adjusted periodically based on going data analysis.

Billable : N/A

Viewable on Reports: No

Remediation

Publishers that voluntarily complete the Traffic Quality Declaration will be upholding honest, fair and transparent standards of traffic quality reporting. Should advertiser questions or concerns arise, however, remediation can be achieved by:

1. **Communication** - trust and transparency can generally be achieved through open conversation between buyer and seller.
2. **Policy adjustment** - buyer and seller may decide to alter the terms of applicable agreements in order to more closely align interests.
3. **Term adjustment** - buyer and seller may decide to define the procedures for the adjustment of payment terms in the event of measurement variance outside an agreed range. Generally, a 5-10% delta between measuring systems would be considered acceptable.

Education

Who cares?

Why is it important to pay attention to this issue even if you don't buy/sell on a pay-for-performance model? See the following presentation deck for answers - [Traffic Quality - What is it, who determines it and how?](#)

What is the "Traffic Quality Problem"?

Internet traffic comes in all shapes and sizes, and can be measured in a variety of ways. However, in advertising specifically, Traffic Quality refers to the identification, measurement, and handling of both valid and invalid user traffic. Valid user traffic is the result of genuine user interest; a real human viewing and potentially clicking on ads; traffic we honestly feel an advertiser should pay for.

Invalid traffic, on the other hand, provides no tangible value to a legitimate advertiser.

Examples of invalid traffic include:

1. GoogleBot, a legit and well-known crawler, crawling a website to add content to the Google search index
2. An errant "double click" on a link or ad
3. A malicious click bot designed to inflate a publisher's click volume
4. A malicious click bot designed to deplete a competing advertiser's budget
5. And so on and so forth

Other resources on this topic include:

http://www.google.com/intl/en_ALL/ads/adtrafficquality/index.html

<http://advertise.bingads.microsoft.com/en-us/traffic-quality#tab=1>

<http://www.bloomberg.com/features/2015-click-fraud/>

<http://www.wsj.com/articles/SB10001424052702304026304579453253860786362>