

# Analyzing Content-Level Properties of the Web Adversphere

Yong Wang<sup>\*</sup>

University of Electronic Science  
and Technology of China  
Chengdu, Sichuan, China  
ywang@uestc.edu.cn

Aleksandar Kuzmanovic

Northwestern University  
Evanston, IL, USA  
akuzma@cs.northwestern.edu

Daniel Burgener

Northwestern University  
Evanston, IL, USA  
d-burgener@northwestern.edu

Gabriel Maciá-Fernández<sup>†</sup>

University of Granada  
Granada, Spain  
gmacia@ugr.es

## ABSTRACT

Advertising has become an integral and inseparable part of the World Wide Web. However, neither public auditing nor monitoring mechanisms still exist in this emerging area. In this paper, we present our initial efforts on building a content-level auditing service for web-based ad networks. Our content-level measurements – understanding the ad distribution mechanisms and evaluating location-based and behavioral targeting approaches – bring useful auditing information to all entities involved in the online advertising business. We extensively evaluate Google’s, AOL’s, and Adblade’s ad networks and demonstrate how their different design philosophies dominantly affect their performance at the content level.

**Categories and Subject Descriptors:** H.3.5 [INFORMATION STORAGE AND RETRIEVAL]: On-line Information Services

**General Terms:** Measurement, Performance

**Keywords:** Behavioral Targeting, Location-based Advertising, Web-based Ad Network

## 1. INTRODUCTION

Advertising is thriving on the Web [1]. Understanding which ads are served at given publisher sites is important for evaluating a commissioner’s effectiveness in bringing the right content to the right audience. In addition, gaining insights about the prevalence and effectiveness of location-based and behavioral ad targeting applied by various commissioners can provide useful auditing information to end-users. Finally, such insights can provide invaluable information to advertisers and publishers when choosing which commissioners to work with. Our key contribution lies in developing an ad monitoring system and in demonstrating that it is capable of effectively screening ad networks at scale.

<sup>\*</sup>Yong Wang is a visiting student at Northwestern University.

<sup>†</sup>This work is supported by China Scholarship Council, NSF CAREER Award no. 0746360, and Spanish MEC project TEC2008-06663-C03-02 (70% FEDER funds).

The system and methodologies we present in this paper are the rudiments of the content-level Web auditing service that we intend to design for the Internet.

## 2. DISTRIBUTION MECHANISMS

We recruit 282 servers from PlanetLab, which are geographically distributed in 36 different countries, to analyze the mechanisms of distributing ads. For each vantage point, we calculate the ‘local’ similarity between itself and any other vantage point in terms of the percentage of identical ads observed in both vantage points. In addition, we also compute the ‘global’ similarity of each commissioner by averaging the ‘local’ similarities for all vantage points when accessing this commissioner’s ads. We evaluate Google, AOL-Adsonar, and Adblade ad networks, since they support *text-based* ads that can be feasibly retrieved from the Web.

Table 1 shows that Adblade’s global similarity is higher than that of AOL-Adsonar, which is in turn higher than Google. A probable cause for such sequence is that Adblade uses a single machine (or a cluster) to serve the requests; as this machine serves the whole pool of ads, it is expected that, after enough requests to publisher sites, all destinations (vantage points) receive all ads independently of their location (high similarity). AOL-Adsonar uses the Akamai CDN network to distribute ads. As the essence of a CDN network is to share the same content over CDN servers, all vantage points share a relatively high similarity. In the Google case, Google has its own private network and, hence, the full control on the methods of distributing ads. Consequently, different vantage points that fetch ads from different ad servers experience a low similarity.

Figure 1 shows the ‘local’ similarity among all vantage points. For each vantage point on the x-axis, a vertical ‘stripe’ corresponding to a value on the x-axis shows the similarities between this vantage point and others. The darker the color in a given  $(x,y)$  box is, the larger the similarity between  $x$  and  $y$  is. Coherently with the results shown in Table 1, the figure shows that the ‘local’ similarities in the Google case are relatively low. This implies that Google has a large pool of ads and distributes different ads into different servers. On the other extreme, the similarities in the Ad-

**Table 1: Global Similarity of each commissioner**

Commissioner	Global similarity (%)
Google	13.16
AOL-Adsonar	59.31
Adblade	72.62

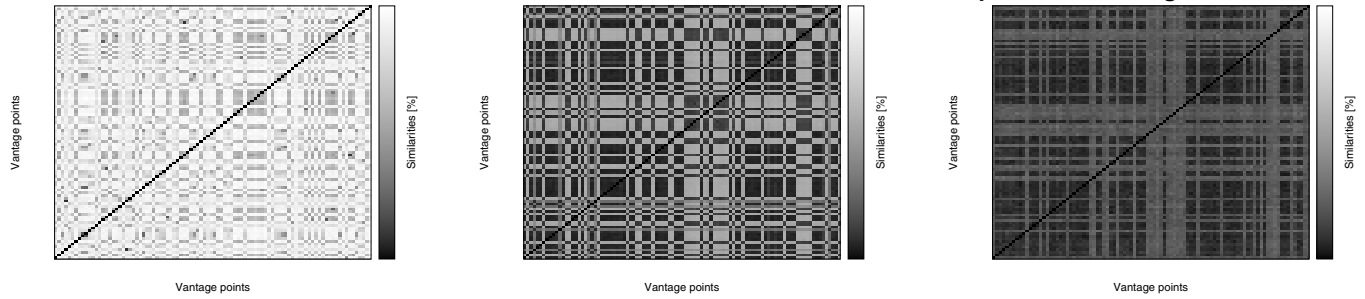


Figure 1: Local similarities among vantage points

blade case are quite high. We consider that this is because Adblade has a smaller pool of ads and puts all of them in the same machine (a cluster of machines).

### 3. LOCATION-BASED ADVERTISING

Commissioners could send ads containing information related to the geographical location of the Web users [2]. These 'localized' ads are more likely to target potential customers. Here, we quantify the percentage of vantage points in which the use of location-based advertising is observed (Table 2). In particular, we count once for the first column if we could observe a match at a *vantage point* at city-level, or the second column if the match is at the precision of state-level, or the third column if no association exists between the location of a vantage point and the texts of all of its ads. Thus, we compute the percentage of vantage points at which location-based advertising is observed.

Table 2 shows that the coverage of location-based advertising in the Google case (31.38% + 21.93% = 53.31%) is wider than in the Adblade case (37.31%), which is in turn larger than in the AOL-Adsonar case (8.00% + 12.00% = 20.00%). This sequence makes sense because Adblade and AOL-Adsonar only apply location-based advertising in most areas of U.S.. In Google case, since Google has the advertising business all over the world, exploiting the same location-based technology is quite feasible.

### 4. BEHAVIORAL TARGETING

Many commissioners claim to be able to more effectively reach users with behaviorally targeted ads [3]. We want to examine the extent to which commissioners participate in behavioral targeting. We decide to use the interest category "sports" in our tests of behavioral targeting because many websites in this category work for each commissioner. We first disable cookies on our PlanetLab nodes in order to prevent behavioral targeting, and then visit websites known to work with the commissioner that fit in the category "sports". After visiting the websites to establish a browsing pattern, we then retrieve the text-based ads from the previous list of websites, which may or may not be related to sports, and scan them for sports-related keywords. We use about 30 keywords here to classify the "sports" category, *e.g.*, sport, cycling, *etc.*

After establishing this baseline, we then repeat this experiment with cookies enabled, in order to determine the difference when behavioral targeting is used. If a company uses behavioral targeting, then we should receive a higher occurrence of sports related ads after establishing our inter-

**Table 2: % of vantage points observing location-based ads**

Commissioners	City (%)	State (%)	No association (%)
Google	31.58	21.93	46.49
AOL-Adsonar	8.00	12.00	80.00
Adblade	37.31	0.00	62.69

est in sports. This is referred to in Table 3 as "Local cookie" as the cookie is established locally for each computer.

We finally repeat this experiment by browsing the same path from a local computer, copying the cookies from that computer to all PlanetLab nodes, and then retrieving the ads again. The purpose of this experiment is to give us an understanding of whether user profile information is geographically distributed or merely stored on the closest ad server to a user. This is referred to in Table 3 as "Uniform cookie" as the same cookie is distributed to all computers.

Table 3 shows the percentage increase of sports related ads over the experiment without cookies. Our results show that both Google and AOL-Adsonar use behavioral targeting for the "sports" interest category, whereas Adblade does not. Google shows a 25% increase when cookies are enabled, and AOL-Adsonar shows a 13% increase. The increases when a uniform cookie is distributed are fairly negligible (3% and 5% for Google and AOL-Adsonar respectively). Apparently, both Google and AOL-Adsonar associate a user profile with interest categories only on a local machine, as the uniform cookie case shows very little increase over the situation without any behavioral targeting at all.

**Table 3: Incremental percentage of observed 'sport' related ads when behavioral targeting is enabled ('local/uniform cookie') comparing with disabled ('no cookie')**

	Local cookie (%)	Uniform cookie (%)
Google	25	3
AOL-Adsonar	13	5
Adblade	0	0

### 5. CONCLUSIONS

In this paper, we deployed a web ad auditing system that can be universally applied to arbitrary commissioners' networks to effectively monitor and help regulate Web-based ad industry. Using this system, we performed a content-level analysis of three representative ad networks with divergent design philosophies, which range from distributing a large number of data centers (Google), to using CDN services (AOL), to standing up servers at a single location (Adblade). We find that the explored commissioners deploy location-based and behavioral ad targeting at various levels of granularity.

### 6. REFERENCES

- [1] Google attracting 25% of online ad revenue. <http://www.searchenginejournal.com/google-attracting-25-of-online-ad-revenue/3910/>.
- [2] H. Xu, L. Oh, and H. Teo. Perceived effectiveness of text vs. multimedia location-based advertising messaging. *Int. J. Mob. Commun.*, 7(2), 2009.
- [3] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen. How much can behavioral targeting help online advertising? In *WWW '09*.