

Tracking the Trackers

An Exploratory Analysis of the Advertising Ecosystem

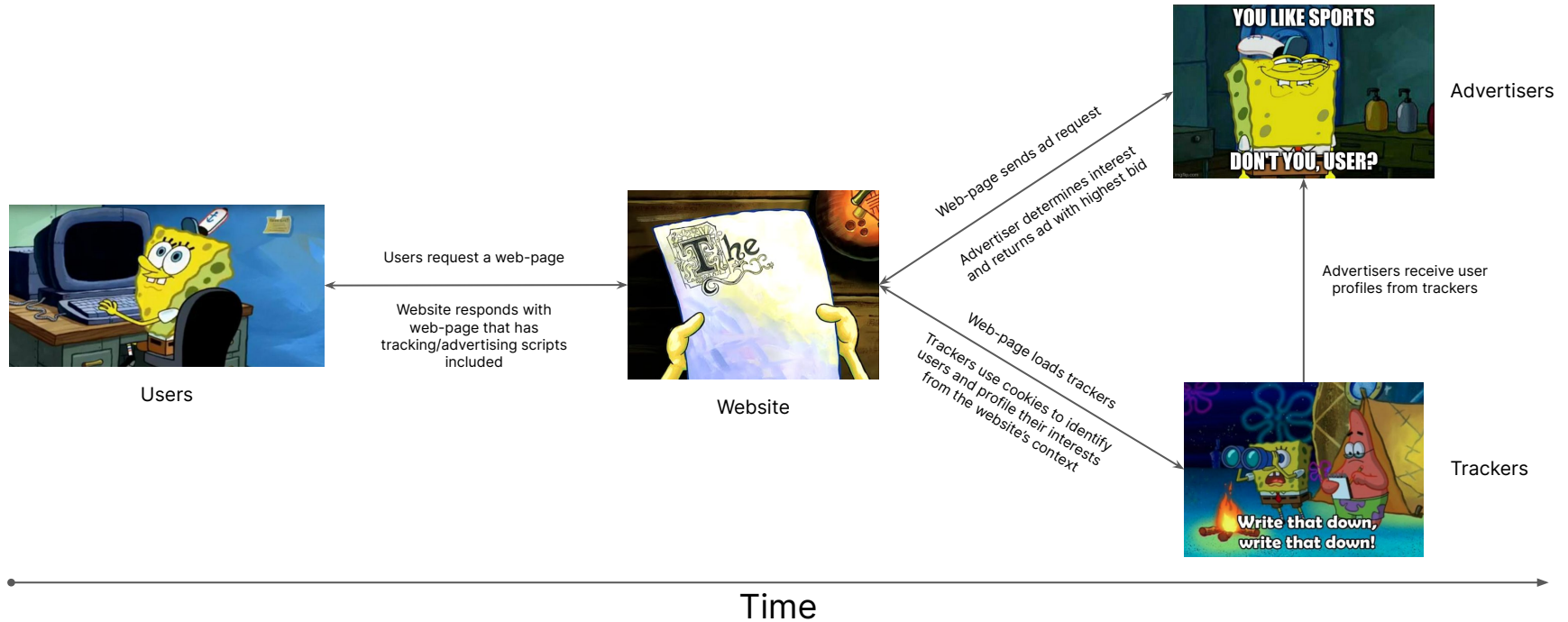
Andrei Blebea

Muhammad Haroon

Parichaya Chatterji

Introduction

- The advertising ecosystem consists of trackers, advertisers, and users



Terminology

- User Persona
 - Refers to a particular type of user with known interests
- Interest Segment
 - Advertiser-defined interest label that is associated with a user persona
- Prebid.js
 - Commonly used header-bidding implementation
 - Exposes auction data to client-side
- CPM (cost-per-mille)
 - Ad placement cost per thousand impressions

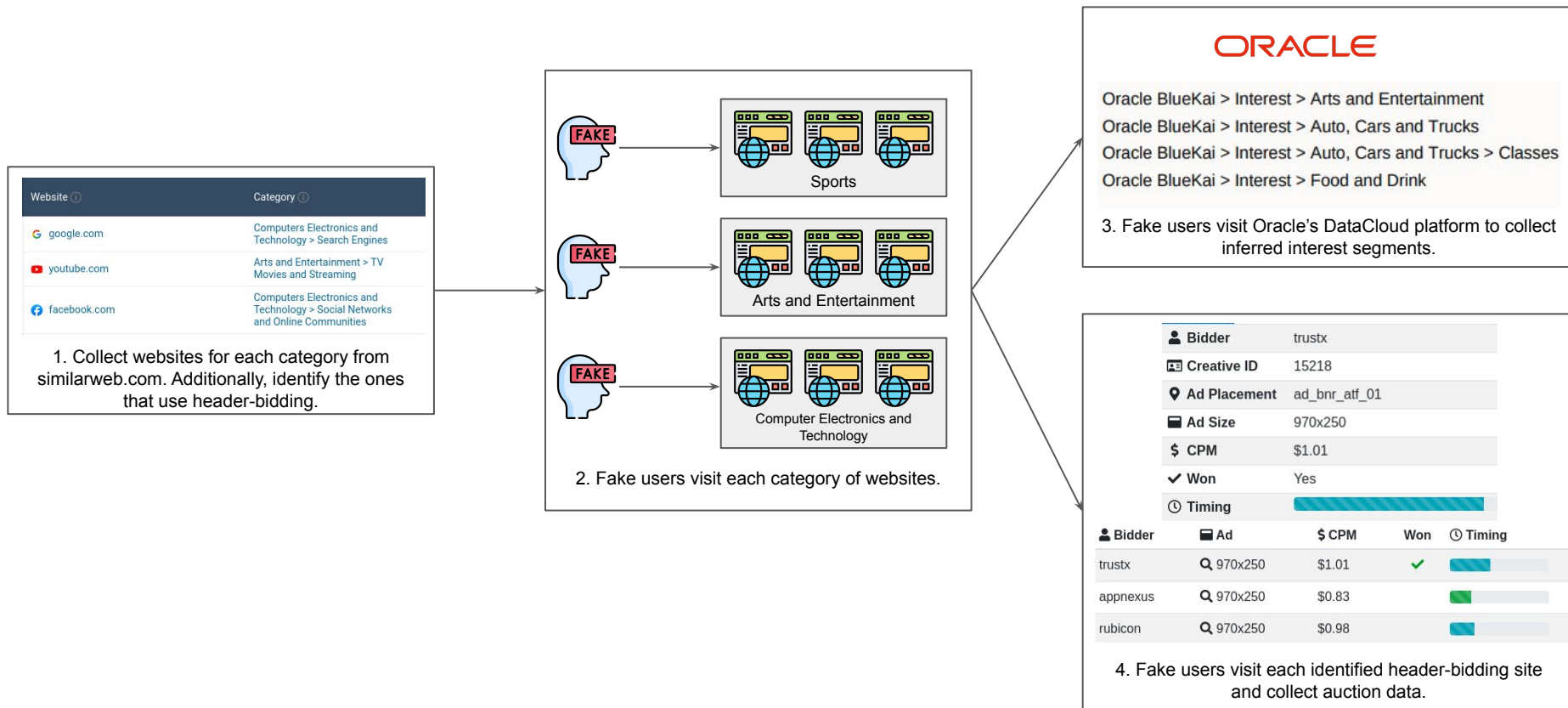
Objectives

- Determine:
 - key players in the advertising ecosystem
 - bidding patterns based on user personas
 - bidding patterns based on interest segments
 - bidding patterns based on website category
- Develop a system to:
 - determine which user profiles are currently lucrative for advertisers (*Cook et al.*)
 - demonstrate user spoofing by combining user profiles (*persona morphing*)

Data Requirements




- User personas with specific interests
- Interest segments for each persona
- Advertiser's bids for each persona

Methodology Overview



Methodology Details

1. Create fake user personas for different types of users
 - a. similarweb.com has a taxonomy of websites and their categories
 - b. Individual browser instances (selenium) that crawl through each site
 - c. Save browsing profiles (cookies, localStorage, etc.)
2. Collect predicted interest segments
3. Collect advertiser bids
 - a. Identify sites with header-bidding enabled
 - b. Visit those sites and collect bids









Website ①	Category ①
 google.com	Computers Electronics and Technology > Search Engines
 youtube.com	Arts and Entertainment > TV Movies and Streaming
 facebook.com	Computers Electronics and Technology > Social Networks and Online Communities




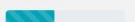

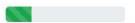

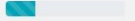
similarweb.com rankings

Oracle BlueKai > Interest > Arts and Entertainment
Oracle BlueKai > Interest > Auto, Cars and Trucks
Oracle BlueKai > Interest > Auto, Cars and Trucks > Classes
Oracle BlueKai > Interest > Food and Drink

Interest segments by Oracle

Header-bidding responses

 Bidder	trustx
 Creative ID	15218
 Ad Placement	ad_bnr_atf_01
 Ad Size	970x250
 \$ CPM	\$1.01
 ✓ Won	Yes
 ⌚ Timing	

 Bidder	 Ad	\$ CPM	Won	⌚ Timing
trustx	 970x250	\$1.01	✓	
appnexus	 970x250	\$0.83		
rubicon	 970x250	\$0.98		

Visualization and Analysis

- Live demo!

Data and Analysis (1) - Source Data

- Similar Web

Category	URL	Rank	SubCategory
Arts and Entertainment	youtube.com	1	TV Movies and Streaming
Arts and Entertainment	netflix.com	2	TV Movies and Streaming
Arts and Entertainment	bilibili.com	3	Animation and Comics
Arts and Entertainment	fandom.com	4	Arts and Entertainment
Arts and Entertainment	imdb.com	5	TV Movies and Streaming
Arts and Entertainment	spotify.com	6	Music
Arts and Entertainment	archiveofourown.org	7	Books and Literature
Arts and Entertainment	disneyplus.com	8	TV Movies and Streaming
Arts and Entertainment	worldstar.com	9	Music
Arts and Entertainment	pixiv.net	10	Visual Arts and Design
Arts and Entertainment	hulu.com	11	TV Movies and Streaming
Arts and Entertainment	hotstar.com	12	TV Movies and Streaming
Arts and Entertainment	syosetu.com	13	Arts and Entertainment
Arts and Entertainment	wattpad.com	14	Books and Literature
Arts and Entertainment	primevideo.com	15	TV Movies and Streaming
Arts and Entertainment	nicovideo.jp	16	TV Movies and Streaming

- URLs, Categories, Subcategories
- 24 unique Categories
- 119 unique SubCategories
- 1200 unique URLs

- Crawl Data

Category	Filename
Adult	Adult-20210513125113.zip
Adult	Adult-20210514002653.zip
Arts and Entertainment	Arts and Entertainment-20210513125213.zip
Arts and Entertainment	Arts and Entertainment-20210514003137.zip
Business and Consumer Services	Business and Consumer Services-20210513130229.zip
Business and Consumer Services	Business and Consumer Services-20210514003850.zip
Community and Society	Community and Society-20210513131412.zip
Community and Society	Community and Society-20210514004553.zip
Computers Electronics and Technology	Computers Electronics and Technology-20210513130036.zip
Computers Electronics and Technology	Computers Electronics and Technology-20210514003450.zip
E commerce and Shopping	E commerce and Shopping-20210512122453.zip

- Unique personas based on Category
- Persona state saved for future work
- Multiple persona for each Category
- 53 unique persona created

Data and Analysis (2) - Ad Interest Category

- Advertiser Segment

Segment	Profile	Category
Audiences by Oracle	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Hobbies and Interests (Affinity)	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Hobbies and Interests (Affinity) > Public In...	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Media and Entertainment	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Media and Entertainment > Movies	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Media and Entertainment > Movies > Enthusias...	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Media and Entertainment > Television (TV)	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
Audiences by Oracle > Media and Entertainment > Television (TV) > ...	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
BlueKai Internal - Private > Data Center	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
BlueKai Internal - Private > Data Center > Phoenix	Arts and Entertainment-20210513125213.zip	Arts and Entertainment
BlueKai Internal - Private > Overlap > Pool: 19505 (Oath - Yahoo/B...	Arts and Entertainment-20210513125213.zip	Arts and Entertainment

- Advertiser-identified Segment for each Persona
- We can cross reference to Category from Similarweb

- Header Bidding Sites

URL
wattpad.com
dailymotion.com
grid.id
9gag.com
genius.com
realtor.com
17track.net
biblegateway.com
horoscope.com
astrology.com
bibliaonline.com.br
biblestudytools.com
leboncoin.fr
ebay-kleinanzeigen.de
olx.pl
coingecko.com

Data and Analysis (3) - Header Bidding Responses

- Header Bidding Responses

```
getBidResponses
{
  "/21797273031/bo_web_mid": {
    "bids": [
      {
        "ad": "<span class=\"PubAPIAd\" id=\"9E026FFC\"",
        "cpm": 0.1,
        "ttl": 300,
        "adId": "160f1bd59ce2b392",
        "meta": {
          "buyerId": 1171826,
          "clickUrl": "aspercreme.com",
          "networkId": 452
        },
        "pbAg": "0.10",
        "pbCg": "",
        "pbDg": "0.10",
        "pbHg": "0.10"
      }
    ]
  }
}
```

Profile	HB_URL
Jobs and Career-20210514003537.zip	vagas.com.br
Jobs and Career-20210514003537.zip	computrabajo.com.co
Jobs and Career-20210514003537.zip	thebalancecareers.com
Jobs and Career-20210514003537.zip	tntsports.com.br
Jobs and Career-20210514003537.zip	grid.id
Jobs and Career-20210514003537.zip	biblegateway.com
Jobs and Career-20210514003537.zip	famousbirthdays.com
Jobs and Career-20210514003537.zip	bibliaonline.com.br
Jobs and Career-20210514003537.zip	biblestudytools.com
Jobs and Career-20210514003537.zip	ebay-kleinanzeigen.de
Jobs and Career-20210514003537.zip	olx.pl
Jobs and Career-20210514003537.zip	finviz.com
Jobs and Career-20210514003537.zip	money.pl
Jobs and Career-20210514003537.zip	marmiton.org
Jobs and Career-20210514003537.zip	nefisemektarifleri.com
Jobs and Career-20210514003537.zip	oceans-nadia.com

- Important fields (like 'getBidResponses' etc) contains very interesting information, such as bidder name, currency, CPM
- This includes multiple bids from a single website, for each persona across multiple categories

Data and Analysis (4) - Additional and Future Work

- Explore “Persona Morphing” based Advertiser Obfuscation
 - Create custom profiles based on the data collected and multiple personas created and saved
 - Explore advertiser classification (segment) for custom profile
 - Compare updated advertiser segment against expected classification based on similarweb category and expected segment
 - If possible, qualify and quantify the delta
 - Create a small dataset of custom profiles and record how it effects segments
 - This can be referenced in future work and validated if it might be possible to obfuscate advertiser segments

Data and Analysis (5) - Additional and Future Work

- Additional Visualization(s): new heatmap to show pairwise similarities between advertisers
 - Based on prominence, occurrences, bidding behavior, etc
- Explore relationship between categories, profiles, persona and segments in more detail
 - Search for correlations and patterns
 - Map to CPM values, bids (*Header Bidding*), etc and explore if any relationship or impact exists

Thank you

- Q & A