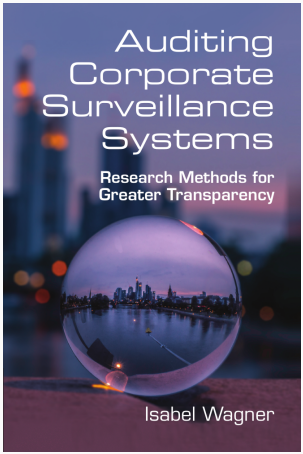
RESULTS FROM TRANSPARENCY RESEARCH

TRACKING, PROFILING, ANALYTICS, ADVERTISING

Isabel Wagner

De Montfort University

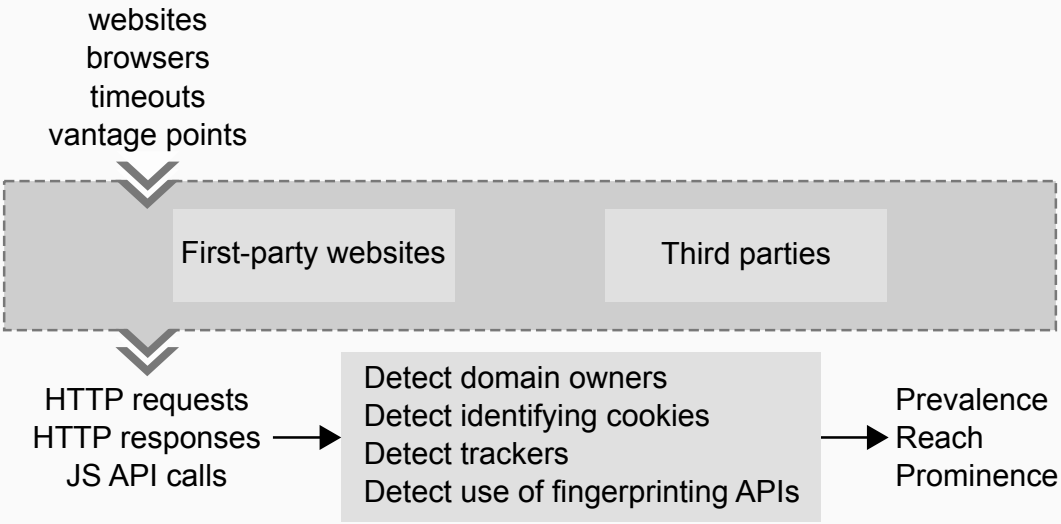


OVERVIEW

- Tracking
  - Stateful (-> chapter 3)
  - Stateless
  - On mobile devices
  - Cross-device (-> chapter 3)
  - Influence of regulations (-> chapter 3)
- Profiling
  - Contents of profiles
  - Anonymity vs identifiability
- Analytics
- Advertising
  - Ecosystem
  - Ad targeting
  - Bidding and pricing
  - Transparency mechanisms
  - Mobile ads
  - Adversarial advertising

Tracking

REMINDER: STUDY DESIGN FOR STATEFUL TRACKING



PREVALENCE AND REACH OF STATEFUL TRACKING

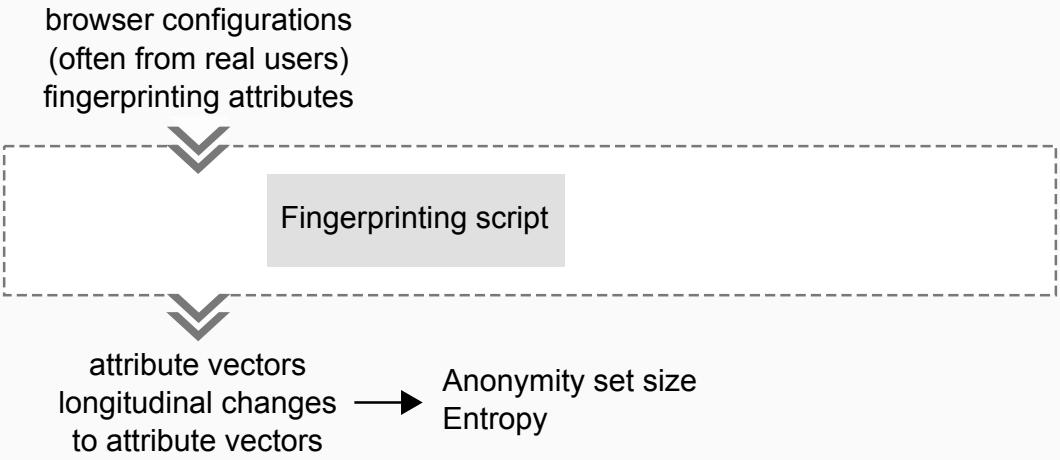
Year	Websites	Third parties or trackers?	Prevalence	Reach
2008	Alexa 1,000	Third parties	n/a	∼35%
2012	Alexa 500	Trackers	91%	39%
2012	Common Crawl	Trackers	89%	25%, 51%*
2015	Alexa 10,000	Trackers	46%	n/a
2015	Alexa 1 million	Third parties	87%	46%
2015	350,000	Trackers	n/a	42%*
2016	Alexa 500	Trackers	n/a	∼40%
2016	Alexa 1 million	Third parties	n/a	∼65%
2017	Alexa 200,000	Third parties	n/a	∼70%
2018	Alexa 1 million	Third parties	n/a	82%*
2019	68,000	Trackers	n/a	44%

- Reach marked with *: organization-based reach, generally larger than domain-based reach
- Tracking is *very common*

INFLUENCE OF VANTAGE POINT, WEBSITE CATEGORY

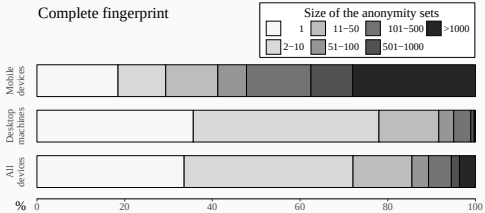
- Vantage point
  - Google has largest reach, except for vantage points in China, Russia, Iran
  - UK users see more trackers than other countries
  - Trackers on 10 most popular sites in UK and China are different
- Website category
  - News websites contain most trackers
  - Unique tracker ecosystem for some categories (e.g., pornographic websites)
  - Captive portals of Wi-Fi hotspots track before user accepts terms of use, 7.4 trackers on average
  - Subsites see more cookies, more trackers, and more cookie synchronization

DESIGN FOR STUDYING STATELESS TRACKING (FINGERPRINTING)



FINGERPRINTABILITY OF WEB BROWSERS

- Collections of browser fingerprints
  - panopticlick.eff.org: 10 browser features, ~470k fingerprints, 2010
  - amiunique.org: 17 features, ~110k fingerprints, 2014/15
  - G'omez-Boix[1]: 17 features, ~2m fingerprints from visitors to a top-15 French website, 2016/17
- Unique fingerprints (anonymity set size = 1): 83%, 89%, 33%
  - Fingerprinting less feasible on very large scale?
- Decline of Flash decreased fingerprintability
  - But new attributes available in HTML5



[1] A. Gómez-Boix, P. Laperdrix, and B. Baudry, "Hiding in the Crowd: An Analysis of the Effectiveness of Browser Fingerprinting at Large Scale," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18, Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 309–318. DOI: 10.1145/3178876.3186097

PREVALENCE OF FINGERPRINTING

- 2014[2]
  - Canvas fingerprinting present on 5.5% of top 100k websites
  - 20 providers of canvas fingerprinting scripts, 95% of instances served by addthis.com
  - addthis.com script uses more features than proposed in research
- 2016[3]
  - Canvas fingerprinting on ~3% of top 100k, 5% on top 1k
    - Possible impact from 2014 paper: reduction, shift from tracking to fraud detection
  - New fingerprinting features: canvas font (2.5%, top 1k), WebRTC (0.6%), AudioContext (<1%), BatteryAPI (<1%)

[2] G. Acar, C. Eubank, S. Englehardt, *et al.*, "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '14, Scottsdale, Arizona, USA: ACM, 2014, pp. 674–689. DOI: *10.1145/2660267.2660347*.

[3] S. Englehardt and A. Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16, Vienna, Austria: ACM, 2016, pp. 1388–1401. DOI: *10.1145/2976749.2978313*.

NEW FINGERPRINTING ATTRIBUTES

- Installed browser extensions
  - By querying web accessible resources
  - By observing changes to website DOMs: 9% of extensions make observable changes, 5.7% make non-functional changes
  - Uniqueness of extension list increases with number of installed extensions
  - 4 may be enough to re-identify most users
- WebGL rendering
  - Included in HTML5 *canvas* element
  - Can link different browsers on same machine because output depends on hardware and graphics driver, not browser

LINKABILITY OF FINGERPRINTS

- Fingerprints change over time: browser/operating system updates, other software updates, change of time zone, cookie deletion, etc.
- Tracking ability can be lost when fingerprint changes
- Linking subsequent versions of fingerprints via heuristics or machine learning
- Linking possible across 20 versions[4]

[4] A. Vastel, P. Laperdrix, W. Rudametkin, *et al.*, "FP-STALKER: Tracking Browser Fingerprint Evolutions," in *2018 IEEE Symposium on Security and Privacy (S&P)*, San Francisco, CA, USA: IEEE, May 2018, pp. 728–741. DOI: 10.1109/SP.2018.00008.

TRACKING ON MOBILE DEVICES

- Tracking on iOS: not well studied
- Tracking on Android: more prevalent, higher reach than on the web
- Alphabet has near 100% coverage
- 90% of apps have at least one tracker[5]
- 60% of paid apps have trackers[6]
- Persistent unique identifiers allow tracking across apps, across mobile/desktop versions[7]
- Ecosystem: 30% of trackers are desktop-only, 13% mobile-only[8]

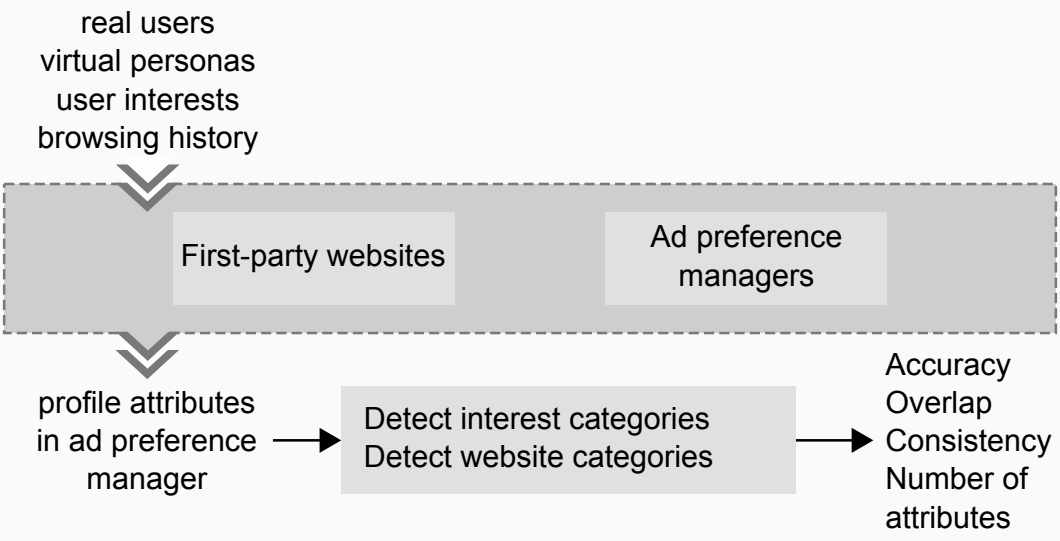
[5] R. Binns, U. Lyngs, M. Van Kleek, *et al.*, "Third Party Tracking in the Mobile Ecosystem," in *Proceedings of the 10th ACM Conference on Web Science*, ser. WebSci '18, Amsterdam, Netherlands: ACM, 2018, pp. 23–31. DOI: *10.1145/3201064.3201089*.

[6] S. Seneviratne, H. Kolamunna, and A. Seneviratne, "A Measurement Study of Tracking in Paid Mobile Applications," in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, ser. WiSec '15, New York, NY, USA: ACM, 2015, 7:1–7:6. DOI: *10.1145/2766498.2766523*.

[7] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, *et al.*, "Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem," in *Proceedings 2018 Network and Distributed System Security Symposium*, San Diego, CA: Internet Society, 2018. DOI: *10.14722/ndss.2018.23353*.

[8] Z. Yang and C. Yue, "A Comparative Measurement Study of Web Tracking on Mobile and Desktop Environments," *Proceedings on Privacy Enhancing Technologies*, vol. 2020, no. 2, pp. 24–44, Apr. 2020. DOI: *10.2478/popets-2020-0016*.

Profiling


real users
virtual personas
user interests
browsing history
First-party websites
Ad preference managers
profile attributes in ad preference manager
Detect interest categories
Detect website categories
Accuracy
Overlap
Consistency
Number of attributes

EXTENT OF PROFILING

- Facebook: each user has median of 310 profiling attributes
- Dataset with 114 users: 17,000 unique attributes[9]
- If every user had their own unique attributes, dataset would have 35,340 unique attributes
- Very small overlap between user profiles, makes users more identifiable

[9] A. Andreou, M. Silva, F. Benevenuto, *et al.*, "Measuring the Facebook Advertising Ecosystem," in *NDSS 2019 - Proceedings of the Network and Distributed System Security Symposium*, San Diego, United States: Internet Society, Feb. 2019. DOI: 10.14722/ndss.2019.23280.

HOW ACCURATE ARE THE USER PROFILES COLLECTED BY TRACKERS?

- 220 crowdsourced participants[10]
- Collect profiles from four ad preference managers (average # of interests): Google (42), Facebook (523), Oracle BlueKai (422), and Neilsen eXelate (9)
- Let participants rate accuracy of interests in their profile
- Overlap between APMs: median <25%
- Half of participants say less than half of APM interests are “relevant” (i.e. rated 3-5 on a 5-point scale)
- Recent browsing history does not explain interests (e.g. Facebook only 9% overlap)

[10] M. A. Bashir, U. Farooq, M. Shahid, *et al.*, “Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers,” in *Proceedings 2019 Network and Distributed System Security Symposium*, San Diego, California, USA: Internet Society, Feb. 2019, p. 15.

ANONYMITY VS IDENTIFIABILITY

- Uniqueness of Facebook profiles[11]
  - Profile is unique (distinguishable from all others) if entropy reaches 29 bits
  - Entropy of public profile attributes: user's current city (13 bits), age (10.5 bits), gender (1.4 bits), relationship status (4.4 bits)
  - Profilers who know these four attributes can look up a Facebook profile
- Uniqueness of click traces[12]
  - Timestamped series of website visits
  - Partial browsing history is sufficient to re-identify users
  - E.g., observable through shoulder surfing, through timestamps of links shared on Twitter
  - Unicity of click traces means they should be interpreted as pseudonyms -> legal implications (GDPR requires explicit informed consent)

[11] T. Chen, A. Chaabane, P. U. Tournoux, *et al.*, "How Much Is Too Much? Leveraging Ads Audience Estimation to Evaluate Public Profile Uniqueness," in *Privacy Enhancing Technologies*, ser. Lecture Notes in Computer Science 7981, E. D. Cristofaro and M. Wright, Eds., Springer Berlin Heidelberg, Jan. 2013, pp. 225–244.

[12] C. Deußer, S. Passmann, and T. Strufe, "Browsing Unicity: On the Limits of Anonymizing Web Tracking Data," in *2020 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA: IEEE, May 2020, pp. 279–292. DOI: *10.1109/SP.2020.00018*.

Analytics

ANALYTICS

- Analytics services rarely studied separately from third-party tracking
- On mobile devices:[13]
  - Analytics library embedded by app developer
  - Library receives unique device identifiers + user information
- Example:
  - App developer wants to analyze app performance by gender
  - Analytics library receives unique identifier + gender for all users
  - Can combine this with user data from all other apps that embed the analytics library
  - Popular analytics libraries can build *very* comprehensive user profiles

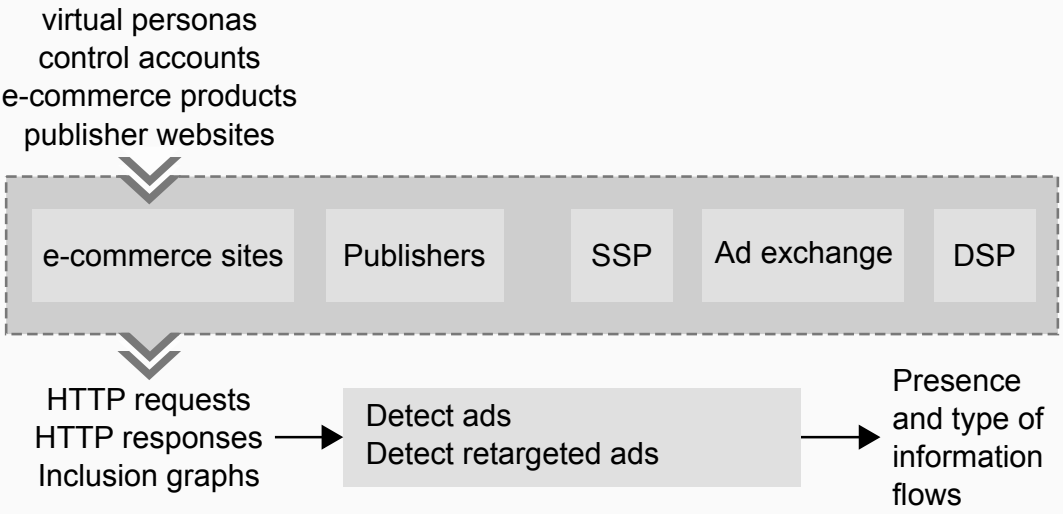
[13] X. Liu, J. Liu, S. Zhu, *et al.*, "Privacy Risk Analysis and Mitigation of Analytics Libraries in the Android Ecosystem," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019. DOI: 10.1109/TMC.2019.2903186.

Advertising

- 2015: 17% of all HTTP requests and 1.1% of all bytes related to advertising[14]
- Ad industry relies on cloud resources and CDNs:
  - Top-10 autonomous systems deliver 50+% of ads
  - Indicates that same back-end infrastructure delivers both content and ads
- Ad traffic has longer delay between HTTP request and response than content traffic – real-time bidding is time-consuming
- 1% of ads are malvertising[15]
- 82% of malicious ads are served by top 10,000 websites, compared to 76% of all ads

[14] E. Pujol, O. Hohlfeld, and A. Feldmann, "Annoyed Users: Ads and Ad-Block Usage in the Wild," in *Proceedings of the 2015 Internet Measurement Conference*, ser. IMC '15, Tokyo, Japan: ACM, 2015, pp. 93–106. DOI: *10.1145/2815675.2815705*.

[15] A. Zarras, A. Kapravelos, G. Stringhini, *et al.*, "The Dark Alleys of Madison Avenue: Understanding Malicious Advertisements," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14, Vancouver, BC, Canada: ACM, 2014, pp. 373–380. DOI: *10.1145/2663716.2663719*.


virtual personas
control accounts
e-commerce products
publisher websites
e-commerce sites
Publishers
SSP
Ad exchange
DSP
HTTP requests
HTTP responses
Inclusion graphs
Detect ads
Detect retargeted ads
Presence and type of information flows

TRACKING CAPABILITIES OF AD NETWORKS

- Analyze traffic from browsing ~890 e-commerce and publisher sites[16]
- Inclusion graph (1900 nodes) is almost fully connected, small-world graph
  - Short path lengths, high clustering between advertising domains
  - Users data can spread rapidly to all participants in the ecosystem
- Information about user browsing history can diffuse through
  - Cookie matching
  - Participation in RTB auctions
- Top 10 ad nodes observe 89-99% of user traffic (depending on RTB assumptions)
- Even with ad blocking, top 10 ad nodes can observe 40-59%

[16] M. A. Bashir and C. Wilson, "Diffusion of User Tracking Data in the Online Advertising Ecosystem," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 4, pp. 85–103, Oct. 2018. DOI: *10.1515/popets-2018-0033*.

ENTITIES IN THE ADVERTISING ECOSYSTEM

- Many interdependent entities[17]
  - Top ad entities by betweenness centrality: Google, Facebook, AppNexus, Integral Ad Science
  - DoubleClick observes 90% of all page impressions – most of each user's browsing history
- Ad entities do not like to be studied
  - ProPublica browser extension recognized Facebook ads by searching for the word "sponsored"
  - Facebook changed the text to "Sp*S*on*S*so*S*red*S*" (invisible *S*, separate *span* elements for 1-2 letters), then "SpSpSononSsososSredredSSS"
  - Facebook disabled automated clicks on ad explanations button[18]

[17] M. A. Bashir and C. Wilson, "Diffusion of User Tracking Data in the Online Advertising Ecosystem," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 4, pp. 85–103, Oct. 2018. DOI: *10.1515/popets-2018-0033*.

[18] J. B. Merrill and A. Tobin, "Facebook Moves to Block Ad Transparency Tools —...," *ProPublica*, Jan. 2019.

USER EXPOSURE TO ADS

- Analyze data from ~5k users collected via browser extension[19]
- Median Facebook user sees 70 ads per week from 33 advertisers
- Ads represent 10-15% of newsfeed content
- Only 23% of user "interests" relate to the ads they see
  - Use McAfee categorization service to match ad landing pages to user interests
  - Matching: compute semantic similarity using spacy.io
  - With similarity threshold 0.5, only 22% of user interests are related to an ad
  - Only 5% of user interests if threshold is 0.7
- Increasing portion of ads by 1% could increase Facebook's revenue by $8 million (weekly)

[19] A. A. Galán, J. G. Cabañas, Á Cuevas, *et al.*, "Large-Scale Analysis of User Exposure to Online Advertising on Facebook," *IEEE Access*, vol. 7, pp. 11 959–11 971, 2019. DOI: *10.1109/ACCESS.2019.2892237*.

AD TARGETING

- 2014: half of websites have targeted ads in at least 80% of ad slots[20]
- 2015: 5% of web ads and 14% of Gmail ads are targeted at user profile[21]
- 2019:[22]
  - 20% of targeting methods are invasive (PII-based) or opaque (look-alike)
  - 12% of ads are retargeted
  - 24% of advertisers use multiple targeting attributes
- 2020: users understand traditional targeting methods (demographics, keywords), but not newer methods (look-alike audiences, custom audiences based on PII lists)[23]

[20] P. Barford, I. Canadi, D. Krushevskaja, *et al.*, "Adscape: Harvesting and Analyzing Online Display Ads," in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14, Seoul, Korea: ACM, 2014, pp. 597–608. DOI: *10.1145/2566486.2567992*.

[21] M. Lecuyer, R. Spahn, Y. Spiliopolous, *et al.*, "Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence," in *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15, Denver, Colorado, USA: ACM, 2015, pp. 554–566. DOI: *10.1145/2810103.2813614*.

[22] A. Andreou, M. Silva, F. Benevenuto, *et al.*, "Measuring the Facebook Advertising Ecosystem," in *NDSS 2019 - Proceedings of the Network and Distributed System Security Symposium*, San Diego, United States: Internet Society, Feb. 2019. DOI: *10.14722/ndss.2019.23280*.

[23] M. Wei, M. Stamos, S. Veys, *et al.*, "What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data," in *29th USENIX Security Symposium (USENIX Security 20)*, USENIX, 2020, pp. 145–162.

SENSITIVE DATA USED FOR AD TARGETING (1)

- 2012: Google shows ads targeted at sensitive attributes, but does not allow users to change or remove these attributes (sensitive attributes did not show up in the ad preference manager)[24]
- 2015: Gmail ads targeted at emails with sensitive contents[25]
- 2015: Virtual personas that were trained only with sensitive keywords see behavioral advertising[26]

[24] C. E. Wills and C. Tatar, "Understanding What They Do with What They Know," in *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '12, Raleigh, NC, USA: ACM, 2012, pp. 13–18. DOI: *10.1145/2381966.2381969*.

[25] M. Lecuyer, R. Spahn, Y. Spiliopolous, *et al.*, "Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence," in *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15, Denver, Colorado, USA: ACM, 2015, pp. 554–566. DOI: *10.1145/2810103.2813614*.

[26] J. M. Carrascosa, J. Mikians, R. Cuevas, *et al.*, "I Always Feel Like Somebody's Watching Me: Measuring Online Behavioural Advertising," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '15, Heidelberg, Germany: ACM, 2015, 13:1–13:13. DOI: *10.1145/2716281.2836098*.

SENSITIVE DATA USED FOR AD TARGETING (2)[27]

- 73% of EU Facebook users (i.e., 40% of EU population!) are assigned at least one of 500 “sensitive” attributes (in the GDPR sense)
  - Collect list of Facebook ad preferences
  - Classify into sensitive/non-sensitive with NLP
  - Let experts verify sensitivity of 20 most common attributes
  - Quantify using Facebook ad audience size estimation
- Women and young people are assigned sensitive attributes at higher rate
- Pre-GDPR, Facebook allowed targeting based on sensitive attributes (paper does not study post-GDPR changes)

[27] J. G. Cabañas, Á. Cuevas, and R. Cuevas, “Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes,” in *27th USENIX Security Symposium (USENIX Security 18)*, Baltimore, MD, USA: USENIX Association, 2018, pp. 479–495.

PII USED FOR AD TARGETING[28]

- Targetable PII:
  - PII in user's Facebook profile
  - PII provided to Facebook Messenger
  - PII shared with Facebook when sharing a phone's contacts
  - PII added to user accounts for 2FA
  - PII added for login alerts
- Non-targetable PII:
  - PII provided to WhatsApp
  - PII uploaded by advertisers to target customers via custom audiences
- How? Reverse-engineer "potential reach" of custom audiences, then construct audiences differing in one entry that trigger change in reach if targetable

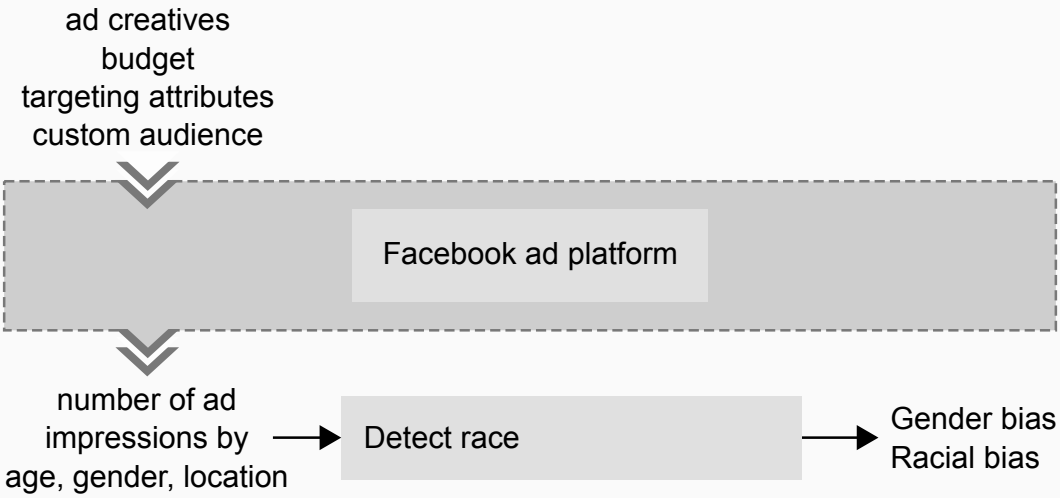
[28] G. Venkatadri, E. Lucherini, P. Sapiezynski, *et al.*, "Investigating sources of PII used in Facebook's targeted advertising," *Proceedings on Privacy Enhancing Technologies*, vol. 2019, no. 1, pp. 227–244, Jan. 2019. DOI: *10.2478/popets-2019-0013*.

DISCRIMINATORY TARGETING WITHOUT USING SENSITIVE ATTRIBUTES

- Custom audiences or proxy attributes allow creating discriminatory ad audiences[29]
- Custom audiences:
  - Advertiser needs list of PII from group of individuals with the desired sensitive attribute, e.g., white male adults under 30
  - US voter records may have this information! (or data brokers)
  - Facebook's lookalike audience algorithm will then find similar people to target
- Proxy attributes:
  - Attributes that strongly correlate with desired sensitive attribute
  - Facebook audience for "Marie Claire" is 90% female
  - Audience for "BlackNews.com" is 89% African American

[29] T. Speicher, M. Ali, G. Venkatadri, *et al.*, "Potential for Discrimination in Online Targeted Advertising," in *Conference on Fairness, Accountability and Transparency*, New York, NY, USA: ACM, Jan. 2018, pp. 5–19.

DESIGN FOR STUDYING BIASES IN AD DELIVERY



BIASES CAUSED BY AD DELIVERY PROCESS

- Gender-neutral STEM ad is shown to 20% more men than women[30]
  - Young women are expensive demographic to target
  - Low-budget ad campaigns reach 55% men on Facebook, high-budget campaigns 55% women[31]
  - Ad delivery that optimizes for cost-effectiveness can exacerbate existing discrimination
- Ad content influences ad delivery[32]
  - Ads for stereotypically male interests (e.g., bodybuilding) reach 80% men, even if targeted at all genders
  - Similar for other attributes: females (cosmetics), black users (hip-hop), white users (country music)
  - Effect occurs even with neutral ad headline, neutral ad text, and image with 98% alpha channel (i.e., invisible to humans)

[30] A. Lambrecht and C. Tucker, "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," *Management Science*, vol. 65, no. 7, pp. 2966–2981, Apr. 2019. DOI: *10.1287/mnsc.2018.3093*.

[31] M. Ali, P. Sapiezynski, M. Bogen, *et al.*, "Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes," *arXiv:1904.02095 [cs]*, Sep. 2019. DOI: *10.1145/3359301*. arXiv: *1904.02095 [cs]*.

REAL-TIME BIDDING

- Retargeted ads for studying cookie synchronization[33]
  - Retargeted ad: ad for product that user has previously viewed on another site
  - DSP needs to connect user on publisher site (seeing the ad) with user on e-commerce site (viewing the product)
  - Only possible through cookie synchronization
  - Observing retargeted ads: 31% more cases of cookie sync than with other heuristics (e.g., cookie identifiers in requests)
- Browsing history leak through real-time bidding[34]
  - Bid requests include visited site + cookie that identifies user
  - Bidders can record this information observe up to 11% of user's browsing history
  - Works even for bidders who do not win the auction

[33] M. A. Bashir, S. Arshad, C. Wilson, *et al.*, "Tracing Information Flows Between Ad Exchanges Using Retargeted Ads," in *25th USENIX Security Symposium*, Austin, TX, USA: USENIX, Aug. 2016, p. 17.

[34] Ł. Olejnik, M.-D. Tran, and C. Castelluccia, "Selling off Privacy at Auction," in *Proceedings 2014 Network and Distributed System Security Symposium*, San Diego, CA: Internet Society, 2014. DOI: *10.14722/ndss.2014.23270*.

ADS.TXT

- Industry standard introduced in 2017 to combat fraud in real-time bidding
- Adoption by 2019: 20% of top 100,000 websites[35]
- Compliance with standard: 70% of buyer-seller pairs comply
- Difficult to study with randomized controlled experiments: would need to be accepted as a *participant* by an ad exchange

[35] M. A. Bashir, S. Arshad, E. Kirda, *et al.*, "A Longitudinal Analysis of the ads.txt Standard," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19, Amsterdam, Netherlands: Association for Computing Machinery, Oct. 2019, pp. 294–307. DOI: 10.1145/3355369.3355603.

HEADER BIDDING[36]

- Most information flows are observable from user-side
- 2019: used by 14% of top 35,000 websites
- Half of websites use only one demand partner
- Google is demand partner for more than 80% of websites
- Latency caused by header bidding: increases with number of demand partners and ad slots
- Median latency: 600ms

[36] M. Pachilakis, P. Papadopoulos, E. P. Markatos, *et al.*, "No More Chasing Waterfalls: A Measurement Study of the Header Bidding Ad-Ecosystem," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19, Amsterdam, Netherlands: ACM, Oct. 2019, pp. 280–293. DOI: 10.1145/3355369.3355582.

AD PRICING

- 2014: Average price for 1,000 ad impressions (CPM): 0.36$ (based on clear-text RTB prices)[37]
- Compare with user valuation of their presence on a website: €7[38]
- 2017: 68% of ad prices encrypted, encrypted prices 1.7x higher[39]
- 2019: header bidding prices lower than RTB prices, median cost between $0.00084 CPM and $0.096 CPM[40]

[37] Ł. Olejnik, M.-D. Tran, and C. Castelluccia, "Selling off Privacy at Auction," in *Proceedings 2014 Network and Distributed System Security Symposium*, San Diego, CA: Internet Society, 2014. DOI: 10.14722/ndss.2014.23270.

[38] J. P. Carrascal, C. Riederer, V. Erramilli, *et al.*, "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online," in *Proceedings of the 22Nd International Conference on World Wide Web*, ser. WWW '13, Rio de Janeiro, Brazil: ACM, 2013, pp. 189–200. DOI: 10.1145/2488388.2488406.

[39] P. Papadopoulos, N. Kourtellis, P. R. Rodriguez, *et al.*, "If You Are Not Paying for It, You Are the Product: How Much Do Advertisers Pay to Reach You?" In *Proceedings of the 2017 Internet Measurement Conference*, ser. IMC '17, London, United Kingdom: ACM, 2017, pp. 142–156. DOI: 10.1145/3131365.3131397.

[40] M. Pachilakis, P. Papadopoulos, E. P. Markatos, *et al.*, "No More Chasing Waterfalls: A Measurement Study of the Header Bidding Ad-Ecosystem," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19, Amsterdam, Netherlands: ACM, Oct. 2019, pp. 280–293. DOI: 10.1145/3355369.3355582.

AD PREFERENCE MANAGERS

- Transparency mechanism: tool for users to view (and sometimes correct) their profile attributes
- User awareness: 90% of users unaware of BlueKai and eXelate, and 48%–68% unaware of Google's and Facebook's APMs[41]
- Completeness of information shown in APMs:
  - Some interest categories, e.g., sensitive interests, are not shown[42]
  - Prevents effectful choice
  - For example, persona only interested in substance abuse was shown significantly more ads on the topic, but APM did not display this category, persona could not manually remove the interest

[41] M. A. Bashir, U. Farooq, M. Shahid, *et al.*, "Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers," in *Proceedings 2019 Network and Distributed System Security Symposium*, San Diego, California, USA: Internet Society, Feb. 2019, p. 15.

[42] C. E. Wills and C. Tatar, "Understanding What They Do with What They Know," in *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '12, Raleigh, NC, USA: ACM, 2012, pp. 13–18. DOI: *10.1145/2381966.2381969*, A. Datta, M. C. Tschantz, and A. Datta, "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 1, pp. 92–112, Apr. 2015. DOI: *10.1515/popets-2015-0007*.

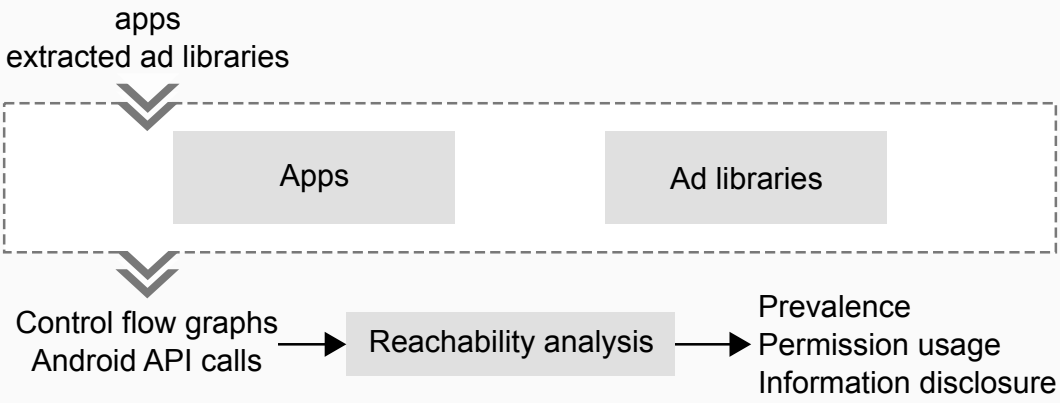
AD EXPLANATIONS ON FACEBOOK[43]

"One reason you're seeing this ad is that [advertiser] wants to reach people interested in Facebook, based on activity such as liking pages or clicking on ads. There may be other reasons you're seeing this ad, including that [advertiser] wants to reach people ages [age] and older who live in [location]. This is information based on your Facebook profile and where you've connected to the internet."

- Good ad explanations are: correct, personalized, complete, consistent, deterministic
- Run own ad campaigns and examine their ad explanations
- Facebook's ad explanations are:
  - Incomplete: do not specifically list attributes or PII, negated attributes never appear in explanation, display demographic attributes rather than behavioral attributes whenever possible
  - Misleading: list targeting attributes that were not used

[43] A. Andreou, G. Venkatadri, O. Goga, *et al.*, "Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations," in *Proceedings 2018 Network and Distributed System Security Symposium*, San Diego, CA: Internet Society, 2018. DOI: *10.14722/ndss.2018.23191*.

DESIGN FOR STUDYING MOBILE ADS WITH STATIC ANALYSIS



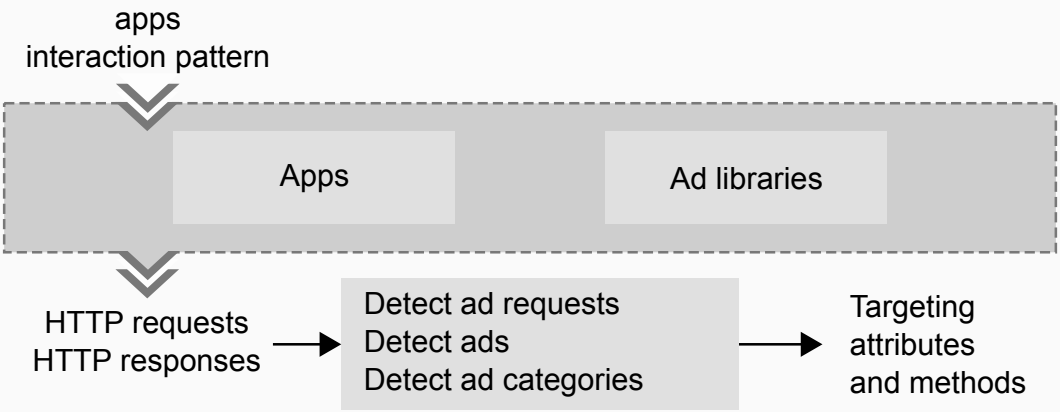
- More concentrated than web: 73% of ads served by Google[44]
- Use of ad libraries (2012):[45]
  - 33% of apps have 1 ad library
  - 10% have 2
  - 3% have 5 or more
- Permission use:
  - 2011: dangerous permissions available to 10% of ad libraries[46]
  - 2013: dangerous permissions available to 20% of ad libraries

[44] N. Vallina-Rodriguez, J. Shah, A. Finamore, *et al.*, "Breaking for commercials: Characterizing mobile advertising," in *Proceedings of the 2012 Internet Measurement Conference*, ser. IMC '12, Boston, MA, USA: ACM, Nov. 2012, pp. 343–356. DOI: *10.1145/2398776.2398812*.

[45] M. C. Grace, W. Zhou, X. Jiang, *et al.*, "Unsafe Exposure Analysis of Mobile In-app Advertisements," in *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WISEC '12, Tucson, Arizona, USA: ACM, 2012, pp. 101–112. DOI: *10.1145/2185448.2185464*.

[46] T. Book, A. Pridgen, and D. S. Wallach, "Longitudinal Analysis of Android Ad Library Permissions," in *Mobile Security Technologies (MoST)*, San Francisco, CA, USA, May 2013, p. 9.

DESIGN FOR STUDYING MOBILE ADS WITH DYNAMIC ANALYSIS



AD TARGETING ON MOBILE DEVICES

- 80% of ad requests have more than 10 attribute-value pairs[47]
  - Unique device identifiers: 66% of ad requests
  - Location: 28% of ad requests
- Behavior-based targeting:
  - Ad network identifiers user based on device identifier, targets based on its internal records
  - Used only by DoubleClick for 80% of their ads
- User data exposed to the advertiser[48]
  - Some DSPs deliver ad content + tracking pixels
  - Advertisers can insert macros into URL parameters of tracking pixels, learn user's device identifier, device model, operating system, ISP, GPS coordinates, gender

[47] S. Nath, "MAdScope: Characterizing Mobile In-App Targeted Ads," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '15, Florence, Italy: ACM, 2015, pp. 59–73. DOI: *10.1145/2742647.2742653*.

[48] M. D. Corner, B. N. Levine, O. Ismail, *et al.*, "Advertising-based Measurement: A Platform of 7 Billion Mobile Devices," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '17, Snowbird, Utah, USA: ACM, 2017, pp. 435–447. DOI: *10.1145/3117811.3117844*.

ADVERSARIAL ADVERTISING

- Idea: person outside the advertising ecosystem can use targeted ads to reidentify or track individuals
  - Target at user's advertising identifier + grid of geolocations to track user[49]
  - Deanonymize website visitors[50]
  - Infer Google profile of website visitors[51]
  - Infer social connections between users[52]
- Ad ecosystem enables not just corporate surveillance, but also surveillance by adversarial third parties

[49] P. Vines, F. Roesner, and T. Kohno, "Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob," in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, ser. WPES '17, Dallas, Texas, USA: ACM, 2017, pp. 153–164. DOI: 10.1145/3139550.3139567.

[50] G. Venkatadri, A. Andreou, Y. Liu, *et al.*, "Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface," in *2018 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA: IEEE, May 2018, pp. 89–107. DOI: 10.1109/SP.2018.00014.

[51] M. Conti, V. Cozza, M. Petrocchi, *et al.*, "TRAP: Using Targeted ads to unveil Google personal profiles," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, Nov. 2015, pp. 1–6. DOI: 10.1109/WIFS.2015.7368607.

[52] Y. Wang, Y. Chen, F. Ye, *et al.*, "Implications of smartphone user privacy leakage from the advertiser's perspective," *Pervasive and Mobile Computing*, vol. 53, pp. 13–32, Feb. 2019. DOI: 10.1016/j.pmcj.2018.12.006.

Summary

RESULTS FOR TRACKING, PROFILING, ANALYTICS, ADVERTISING

- Study designs for classes of research questions:
  - Stateful tracking
  - Stateless tracking
  - Profiling
  - Data flows in the ad ecosystem
  - Ad delivery process
  - Static and dynamic analysis for mobile apps
- Overview of research results
- New developments on web: need new studies, but often rely on/tweak existing study designs

ABOUT THIS SLIDE DECK

- These slides are designed to accompany a lecture based on the textbook “Auditing Corporate Surveillance Systems: Research Methods for Greater Transparency” by Isabel Wagner, published in 2022 by Cambridge University Press.
-
-
- Please leave this slide intact, but indicate modifications below.
  - Version 2022-04
  - Improved version for release on book website (Isabel Wagner)
- Updated versions of the original slide deck are available online: corporatesurveillance.org