



Universiteit
Leiden
The Netherlands

Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification

Eijk, R.J. van

Citation

Eijk, R. J. van. (2019, January 29). *Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification*. Retrieved from <https://hdl.handle.net/1887/68261>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/68261>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/68261>

Author: Eijk, R.J. van

Title: Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification

Issue Date: 2019-01-29

Acknowledgments

A special word of thanks goes to Dr. Adriaan in 't Groen and Dr. Johannes Magliano-Tromp, who have been supportive from the moment I was welcomed at the Dual PhD Centre, Faculty of Governance and Global Affairs of Leiden University. The Dual PhD Centre offered a wonderful environment that cultivates the scientific prerequisites and social values necessary for this undertaking. Moreover, I would like to recognize two persons in particular for their support: Joke Hellemons and Petra Evers.

It is a pleasure to thank Wilbert Tomesen and Jacob Kohnstamm for supporting me throughout my journey. The completion of the support by the Autoriteit Persoonsgegevens was performed by the new chairman Aleid Wolfsen who immediately showed enthusiasm for the topic.

Then, I would like to express my deep gratitude to Jeff Chester. I also thank Udo Oelen, who made the thesis possible. I thank Aldo Cortesi who made the data collection possible by his technical support and Dr. Frank Takes for insightful comments. I would like to say a word of appreciation to Hannes Tschofenig and Rigo Wenning for introducing me to the standards community. Not to be forgotten are David van Altveer, Dr. Alexander Dix, Dr. Balachander Krishnamurthy, Sjoera Nas, Mike O'Neill, Dr. Petros Papadopoulos, Cecile Schut, and Floor Terra.

I thank the members of the assessment committee for the time taken to read the thesis. My supervisors can be thanked more explicitly since they have no say in the final assessment. Prof. dr. H. Jaap van den Herik and Dr. Mark Dechesne made true what they had promised and that was much.

I am grateful to my parents and my family. Last, and therefore most important I thank Delia Chin-Ten-San. I dedicate this thesis to you.

Publications and presentations

ORCID: <https://orcid.org/0000-0002-2203-6819>.

- IWGDPT (2014). Working Paper on Web Tracking and Privacy: Respect for context, transparency and control remain essential. In Dr. Alexander Dix (Ed.), *International Documents on Data Protection in Telecommunications and Media* (Vols. 1983–2013, pp. 492–514). Berlin: Berliner Beauftragter für Datenschutz und Informationsfreiheit. (Working paper adopted in 2013. First official draft by Van Eijk, R.)
- O'Neill, M., & Van Eijk, R. (2016). *Do Not Track 1.1 (Unofficial Draft)*. [Position Paper]. Retrieved 10 Januari 2017, from <https://github.com/w3c/dnt/blob/master/DNTBugs.html>
- Tschofenig, H., Leiba, B., Cook, B., & Van Eijk, R. (2011). Browser support for the Open Authorization (OAuth) Protocol. In *W3C workshop Identity in the Browser*. Mountain View, CA: W3C. [Position paper]. Retrieved 11 January 2014, from http://www.w3.org/2011/identity-ws/papers/idbrowser2011_submission_32.pdf
- Tschofenig, H., & Van Eijk, R. (2011). Do Not Track, An Attempt to Frame the Debate. In *W3C workshop Web Tracking and User Privacy*. Princeton, NJ: W3C. [Position paper]. Retrieved 11 January 2014, from <http://www.w3.org/2011/track-privacy/papers/Tschofenig.pdf>
- Van den Herik, H. J., & Van Eijk, R. (2013). Social Innovation & Big Data. In *IT Innovation Day*. Amersfoort: IDG Nederland. [Presentation].
- Van den Herik, H. J., & Van Eijk, R. (2014). The all-pervasive impact of Big Data in daily life: from Causality to Correlation. In *Launch of the Future Center - Smart Systems*. Leiden: Leiden Institute of Advanced Computer Science, Leiden University. [Presentation].
- Van den Herik, H. J., & Van Eijk, R. (2015). Big Data: een Introductie. In *Ondernemen en Big Data*. The Hague: VNO-NCW WEST. [Presentation].

- Van den Herik, H. J., & Van Eijk, R. (2016a). Big Data and Judicial Decisions. In *LERU Summer School 2016: Data Stewardship for Scientific Discovery and Innovation*. Leiden: League of European Research Universities. [Presentation]. doi: <http://dx.doi.org/10.2139/ssrn.2809612>
- Van den Herik, H. J., & Van Eijk, R. (2016b). De haakjes van Big Data. In *(Big) Data festival*. The Hague: Ministry of Infrastructure and Environment. [Presentation].
- Van Eijk, R. (2011). *Web Tracking Detection System (TDS): An effective strategy to reduce systematic monitoring and profiling of user habits across websites*. MSc thesis. Leiden: Leiden University, Leiden Institute of Advanced Computer Science. Retrieved 5 March 2018, from <http://liacs.leidenuniv.nl/assets/Masterscripties/ICTiB/Rob-van-Eijk-non-confidential.pdf>
- Van Eijk, R. (2012). *International Perspectives on Cloud Computing*. 36th Datenschutzfachtagung. [Presentation]. Köln: DAFTA.
- Van Eijk, R. (2013a). *Big Data: the Big Disconnect*. Workshop Formal methods for the informal world. [Presentation]. Leiden: Lorentz Center.
- Van Eijk, R. (2013b). *Forensic Aspects in Digital Investigations*. Colloquium. [Presentation]. Worcester, MA: Worcester Institute of Technology.
- Van Eijk, R. (2015). Towards a Delicate Balance in Strategic Innovations. In *Leadership challenges with big data. Module 10: Ethical, Legal and Privacy Challenges*. Rotterdam: Rotterdam School of Management, Erasmus University. [Presentation]. doi: <http://dx.doi.org/10.2139/ssrn.2696674>
- Van Eijk, R. (2016a). *A brief introduction to Real Time Bidding (RTB)*. [Presentation]. Brussels: Article 29 Working Party. doi: <http://dx.doi.org/10.2139/ssrn.2856941>
- Van Eijk, R. (2016b). Informatie: Leidt meer informatie tot minder vrijheid? In A. in 't Groen, C. De Roen, & P. Slaman (Eds.), *Leids-Haags Allegorieënlexicon* (pp. 151–160). Delft: Uitgeverij Eburon.
- Van Eijk, R. (2016c). Update on Privacy Legislation. In *Leadership challenges with big data. Module 10: Ethical, Legal and Privacy Challenges*. Rotterdam: Rotterdam School of Management, Erasmus University. [Presentation]. doi: <http://dx.doi.org/10.2139/ssrn.2804720>
- Van Eijk, R. (2017a). *On browser settings, cookies, and (not) being tracked by digital advertisements*. Working Party on Telecommuni-

- cations and Information Society. [Presentation]. Brussels: Council of the European Union. doi: <http://dx.doi.org/10.2139/ssrn.3003152>
- Van Eijk, R. (2017b). Tracking classification: a cluster analysis approach to identify missing values. In N. Bielova, N. Laoutaris, A. Narayanan, & N. Nikiforakis (Eds.), (Vol. 7, p. 103). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik. [Peer Reviewed Report]. doi: <http://dx.doi.org/10.4230/DagRep.7.4.83>
- Van Eijk, R. (2018a). Processing personal data: privacy requirements and security measures. In *Leadership challenges with big data & analytics*. Rotterdam: Rotterdam School of Management, Erasmus University. [Presentation]. doi: <http://dx.doi.org/10.2139/ssrn.3285144>
- Van Eijk, R. (2018b). *Technology for Privacy: Protecting Against Online Tracking and Profiling*. [Presentation]. Brussels: Computers, Privacy, and Data Protection congress. doi: <http://dx.doi.org/10.2139/ssrn.3106582>
- Van Eijk, R., & Toubiana, V. (2017). *Browser settings and consent: A brief introduction to Do Not Track (DNT)*. [Presentation]. Brussels: Committee on Civil Liberties, Justice and Home Affairs.
- Van Wissen, A., Kamphorst, B., & Van Eijk, R. (2013). A Constraint-Based Approach to Context. In P. Brézillon, P. Blackburn, & R. Dapoigny (Eds.), *Modeling and Using Context* (Vol. 8175, pp. 171–184). Heidelberg: Springer. [Peer Reviewed Journal].
- Wefers Bettink, W., Van Eijk, R., & Wagner, F. (2012). *Strictly Speaking: Cookies, Consent and Compliance*. Europe Data Protection congress. [Presentation]. Brussels: IAPP.

Curriculum Vitae

Robbert J. van Eijk (1969, Rosmalen) completed secondary school (1988, Gymnasium- β) at the Sint-Janslyceum in 's-Hertogenbosch. He received an undergraduate degree in electrical engineering (1990) from Eindhoven University of Technology, a BSc in electrical engineering (1994) from Fontys University of Applied Sciences in Eindhoven, and a MSc in ICT and business (2011) from the Leiden Institute of Advanced Computer Science, Faculty of Science of Leiden University.

Robbert works as senior supervision officer at the Autoriteit Persoonsgegevens, the Dutch Data Protection Authority (DPA). He represented the Dutch DPA in (inter)national meetings and as technical expert in court. He also represented the Article 29 Data Protection Working Party in the multi-stakeholder negotiations of the World Wide Web Consortium (W3C) on Do Not Track.

He started his professional career in the automotive industry. As an onsite consultant he specialized in dealer-network planning using geographical information systems. Then, he founded Blaeu Business Intelligence, a company with a focus on office automation for small-sized enterprises. The company was acquired after he had been running it successfully for nine years.

He worked together with SBV Forensics, a company specialized in financial forensic investigation. Together they assisted the official receiver after the bankruptcy of KPNQwest. They also collaborated in preparation of a court case after the bankruptcy of Ceteco.

Recently, he served as an external supervisor of the concluding Master's thesis projects of Scheepers [2018] and Turcios Rodríguez [2018]. He is an invited expert in the W3C Tracking Protection Working Group and the W3C Privacy Interest Group. He also contributes to the Inria PrivaWEB project on web tracking, e-Privacy Regulation, and user protection.

Index

NOTE: This index contains keywords representing concepts commonly used in Web Privacy Measurement (WPM) and Real-Time Bidding (RTB) technology. The keywords are listed alphabetically. In some cases they are extended to closely related (self-contained) key concepts.

B

bidding-as-a-service, 145

big data, 31–34

and internet platforms,
35

application of big data,
34

meta definition

analysis, 33

exchange, 33

interpretation, 33

maintenance, 33

real-time, 33

retention, 33

search, 33

variety, 32

velocity, 32

visualization, 33

volume, 32

narrative science, 33

bridging node, 187

C

C-steps, 90–92

central node, 187

D

data, 30

big data, 31–34

context, 5, 21, 28, 31, 63,
134, 146, 153, 179,
185, 193, 196,
203–204, 206

information, 31

knowledge, 31

metadata, 38–40

missing values, 124

small data, 36–38

Data Management Platform,
145

deal ID, 149

Demand Side Platform, 144

dendrogram, 161, 164

branches, 164

clades, 164

leaves, 164

phylogenetic tree, 165

display advertising

ad-slots, 142

publisher, 143

remnant inventory, 203

E

empirical view, 133–140,
153–171, 195

entropy, 63, 68, 85, 105, 111,
113–118
 high-entropy identifier,
 114
 Shannon entropy, 63,
 114
event tracking, 49–53

F

fingerprinting
 active, 60, 65–67
 browser extensions, 99
 cookie-like, 61
 cross-device, 61
 mitigations, 67
 passive, 60
 risks, 62
 stateful, 60
 stateless, 60
FourthParty, 72, 74, 83
FPDetective, 73, 83

G

graph, 106
 cluster-edge
 betweenness, 183
 Cypher, 107, 111, 122
 eigenvector centrality,
 190
 induced subgraph, 110
 isomorphism, 119
 node betweenness,
 184–185
 refactoring, 127, 157,
 184, 186–188, 266
 self-referencing loop,
 157, 166, 190
 subgraph, 108–110
 subgraph isomorphism,
 119
Graph-Based
 Methodological

Approach, 42, 79,
81, 102, 105, 129,
132, 180, 196, 197,
199, 200, 205

H

header bidding, 144
heatmap, 161
HyperText Transfer Protocol,
68
 header field
 registration, 70
 redirection, 121, 157
 request header, 69
 response header, 70
 standard, 71

I

identifier, 40
 cookie, 37
 false negative, 117
 false positive, 116
 Youtube-nocookie, 181,
 184
identity, 41
information, 31
interconnecting node, 188
inventory source, 142

K

knowledge, 31

L

legal view, 171–179, 195

M

M-steps, 129, 157–158
metadata, 38–40, 47
 difference between data
 and metadata, 38
 importance of metadata,
 39
 twenty common types,
 63

Mitmproxy, 73, 74, 86–87, 89,
96, 97, 103–104, 127
multidisciplinary
 legal-science
 paradigm, 55

N

narrative science, 33
Netograph, 77, 89, 154, 157
network of partners, 133,
195

O

OpenWPM, 74, 81, 83, 86,
117

P

privacy
 acquisitions, 76,
 133–134
 and webtracking, 21–30
 and WPM themes, 46
 commercial tracking, 48
 consent
 implementation
 differences, 175
 implied, 175
 strict, 174
 enforceability, 183, 204
 fundamental right, 22,
 24, 27, 202–205
 information asymmetry,
 53
 joint controller, 202
 mergers, 76, 133, 135
 partnerships, 76, 133,
 135–137, 195
 political advertising
 online, 48–49
 principles, 22–23
 state tracking, 48
 user control, 204

Westin, 22
profiling, 29, 98
prototyping, 54, 205

R

R-steps, 103–104
Real-Time Bidding
 bidding-as-a-service,
 145
 deal ID, 149
 definition, 20–21
 header bidding, 144
 missing values, 124
 node classification,
 124–128, 128
 node types, 185–190
 private deal, 148
 process, 150–152
 protocol, 146–148
 sequence diagram, 151
 system, 140
RTB-steps, 150–151, 154, 174,
190

S

satellite node, 187
small data, 36–38
societal view, 180–192, 196
Supply Side Platform, 143

T

Tag-Based Integration, 149
theoretical view, 140–152,
195
top-100 edges, 163, 167
top-20 edges, 161, 178

W

web privacy measurement,
45–49
 deep crawl, 91, 92,
 153–156

- long tail phenomenon,
 - 94–95
- research methodologies,
 - 54
- research themes, 46
- shallow crawl, 92, 117,
 - 154, 171
- visualization tools, 77
- web tracking, 25, 27–28
 - cross border, 160,
 - 168–170, 193
 - definition, 28, 106
 - deterministic
 - identification, 58
 - event tracking, 49–53
 - extended model,
 - 111–113, 118
 - generic model, 105–107,
 - 123, 124, 156
 - probabilistic
 - identification, 58
 - profiling, 29, 98
 - stateful, 61
 - stateless, 61
- WebXray, 76