



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

5

Conclusions

In this chapter, we provide a graph-based answer to our research questions RQ1, RQ2, and to the problem statement. In Section 5.1 we summarize the two research questions and the answers provided in Chapter 3 and Chapter 4. In Section 5.2 we formulate our answer to the PS and draw a general conclusion from our research work. In Subsection 5.3 we discuss Step 8 of our research methodology, viz. evaluation of the framework. Finally, in Section 5.4 we conclude the chapter by indicating directions for further research.

5.1 ANSWERS TO THE RESEARCH QUESTIONS

RESEARCH QUESTION 1: How do we move from data collection to graph analysis?

The answer to our first research question is derived from Chapter 3. Addressing RQ1 we presented a graph-based approach with small-data models based on metadata from the HTTP header. The properties of the nodes and edges in the small-data model correspond with the properties of the metadata as defined in our research environment consisting of data stored on websites.

The main improvement of the research methods and the capabilities for analysis is the fact that a small-data model for web tracking allows us to apply a variety of network science algorithms to the data. The Graph-Based Methodological Approach (GBMA) allows us to deepen our understanding of what actually happens when we visit websites containing RTB advertisements.

RESEARCH QUESTION 2: What are the emerging characteristics of the graph that is fit for graph analysis?

The answer to our second research question is derived from Chapter 4. Addressing this question, we presented four views on RTB systems. We started with an *empirical* view on the network of partners (View 1) and we may conclude that (1) each network of partners relies on its own targeted-advertising framework and (2) has its unique way of tracking end-users for various purposes.

Next, we presented a *theoretical* view of RTB systems (View 2) consisting of eight key building blocks. We may conclude that partnerships enable the real-time integration of RTB systems. Moreover, the interconnection of RTB systems and their partners with standardized RTB protocols creates the infrastructure for RTB.

We developed a *legal* view by linking our theoretical view to an empirical view on partner networks (View 3). We described the landscape in August 2016. We based our legal view on the analysis of a contextual dataset comprised of news items from news websites from 27 European countries. Furthermore, we provided the Graph-Based Methodological Approach (GBMA) to gauge the situation of differences in implementations of consent mechanisms for (tracking) cookies.

Finally, we showed our *societal* view on graph analysis of partner networks in RTB systems (View 4). The graphical representation of the interconnectedness (1) inside a partner network and (2) between partner networks helped us to precisely differentiate between clusters that are ad related and other non-ad related.

In summary, we may conclude that RTB as a technology relies on dynamic networks of partners to enable an ad. The interconnection in a betweenness cluster is caused by the data flows of the companies themselves due to their specializations in ad technology. As a result of the specializations, the dataflow is functional and contains structural metadata which we use to build and annotate our picture.

5.2 ANSWER TO THE PROBLEM STATEMENT

Below we provide an answer to the problem statement based on the flow of arguments throughout the thesis and the eight steps of our research methodology. The primary focus of our investigation was: to address on the gap between (1) what end-users think is

going on and (2) what is actually happening? The two issues led us to formulate a problem statement connecting privacy and RTB.

PROBLEM STATEMENT: *To what extent can we measure the privacy component connected to Real-Time Bidding?*

The answer to the problem statement is as follows. The network of RTB partners measured from the perspective of an end-user is the start of an extensive (say disruptive) RTB dataflow. Based on our research findings we arrive at four conclusions. They are different and address the following topics: (1) specializations, (2) visualization and annotation, (3) limitations, and (4) privacy impact.

CONCLUSION 1: SPECIALIZATIONS

We may conclude that the distinction between RTB systems is based on the specializations in ad technology of the partner companies. As a result of the specializations, the dataflow is functional and contains structural metadata.

CONCLUSION 2: VISUALIZATION AND ANNOTATION

We may conclude that we can visualize and annotate the interconnection in a partner network with a betweenness cluster algorithm.

CONCLUSION 3: LIMITATIONS

We may conclude that we cannot measure the full dataflow since most of the RTB functions are performed through sharing data outside the sphere of the browser.

CONCLUSION 4: PRIVACY IMPACT

We may conclude that due to a lack of metadata in current RTB implementations, e. g., about the identity of a partner and the purpose of a UID, we are rather limited in differentiating between UIDs. There are UIDs with a high impact on the privacy of an end-user (relating, e. g., targeting or retargeting cookies) and those with a smaller impact (e. g., UID cookies for campaign measurement).

5.3 FUNDAMENTAL DATA PROTECTION RIGHTS

As briefly addressed in Section 1.2, the privacy of an end-user should be regarded in conjunction with fundamental rights. In Recitals 1–13 GDPR, the European legislator explains the relation between the principles, the legal norms in the GDPR, and the fundamental right to the protection of personal data.

Furthermore, the European legislator explains the boundaries to the concept of protecting personal data in Recital 4 GDPR:

„The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality.“

MULTIPLE (JOINT) CONTROLLERS

Our answer to the problem statement leads us - with the fundamental rights in mind - to two intriguing questions for future research: (1) is RTB as a technology disproportionate to its intended use? and (2) should RTB be prohibited if (a) data leakage to unknown (joint) controllers cannot be prevented and (b) leaked data is subsequently processed for unknown purposes? Whatever the answer, we take the stance that, in principle, end-users should be able to exercise their fundamental data-protection rights.³⁵⁸

Art. 29 WP provided recently a guidance on the information requirement in a situation of multiple (joint) controllers (as defined in Article 4(7) GDPR).³⁵⁹

(...) „Art. 29 WP notes that in a case where the consent sought is to be relied upon by multiple (joint) *controllers* or if the data is to be transferred to or processed by other controllers who wish to rely on the original

³⁵⁸ We remark Leenes and Kosta [2015], who also questioned whether the renewed informed consent framework of the GDPR can change the current status quo in web tracking.

³⁵⁹ Article 4(7) GDPR reads as follows: „controller means the natural or legal person, public authority, agency or other body which, *alone or jointly with others, determines the purposes and means of the processing of personal data*; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law.“ (emphasis added)

consent, these organizations should all be named. *Processors* do not need to be named as part of the consent requirements, although to comply with Articles 13 and 14 of the GDPR, controllers will need to provide a full list of recipients or categories of recipients including processors. To conclude, Art. 29 WP notes that depending on the circumstances and context of a case, more information may be needed to allow the *data subject* to genuinely understand the processing operations at hand." [Article 29 Working Party, 2018, WP259 rev.01] (emphasis added)

Below we provide an example on remnant inventory on a publisher's website. We aim to illustrate the (joint) responsibility with respect to the GDPR of an *advertiser* when resorting to RTB for prospects (potential new customers). The new customers have expressed an interest by visiting a product page on the corresponding website. We remark that the use of RTB technology for this purpose is a (relatively) new phenomenon in some private sectors that are known for processing data of a sensitive nature (e.g., insurance companies and financial service providers).

EXAMPLE: REMNANT INVENTORY

We already know that publishers largely depend on RTB advertising based on remnant inventory. Visiting a product page of a *non-premium advertiser* (Context A) may lead to an ad on a publisher's website (Context B). The publisher shows ads to its readers of its *premium advertisers* wishing to market them to the newspaper's audience. However, a visit to the publisher's website may also lead to an ad shown to the end-user based on remnant inventory. In other words, a (re)targeted ad from a *non-premium advertiser* based on data about the user collected in Context A may be shown to the end-user in the publisher's Context B. We may conclude that due to a dependency on remnant inventory in the RTB landscape, RTB ads can follow prospects when they navigate from one website to another website.

COLLECTION OF RTB DATA

The interesting point we would like to convey with the above example is, that the non-premium advertiser's product page - where

the RTB data originates from - does not have to include RTB *ad space* itself when resorting to RTB technology. In absence of RTB-ad space, the advertiser may restrict the use of RTB technology to partners that specialize in *collecting* data about the end-user for the purpose of marketing them to other websites on the web. Such a restriction is fully understandable since the advertiser does not want to run the risk of ads from competitors on its own website.

LACK OF USER CONTROL

The example above leads to our view that an advertiser is (most likely) a (joint) data controller with respect to the GDPR for the (re)targeted advertisement (in Context A).³⁶⁰ As a result the burden is - among other (joint) data controllers - on the publisher as well as on the advertiser to demonstrate the legal basis for the ad. Moreover, the advertiser is accountable for the collection of the end-user's data for the purpose of the (re)targeted RTB campaign in Context A and the (subsequent) processing of his data by tens, possibly hundreds of RTB parties. Moreover, the end-user should be able to exercise his fundamental rights, e. g., revoking consent to prevent further data collection and use.

We remark that current RTB frameworks tend to operate in the interest of the *ad exchange* and (in most cases) provide for information about the origin of the (behavioral) ad to the end-user when the ad is rendered. Conversely, we may conclude that current RTB frameworks do not (tend to) operate in the interest of the *advertiser* and the *end-user*, i. e., by providing a mechanism to provide for revocation of consent (in Context A).

ENFORCEABILITY OF DO NOT TRACK MECHANISMS

The legislative intervention by the European legislator with the GDPR is promising, and simultaneously complex. At the time of writing it is unclear whether consent expressed by using the appropriate technical settings of a software will be enforceable by DPAs. Since the GDPR-model of transparency and consent raises technical and legal questions when applied to the dynamic nature of RTB networks, we advocate to include a provision in the EPR on the enforceability by DPAs of Do-Not-Track expressions through

³⁶⁰ To the extent that the collected RTB data qualifies as personal data as defined in Article 4(1) GDPR.

modern browsers. The EPR will further clarify the GDPR by translating its principles in specific rules. Such a provision would not only strengthen the transparency, user control, and trust (see Recital 66 EPD, Article 21(5) GDPR, and the article on consent in the (draft) EPR), but the provision is - we advocate - also needed to catalyze the change to bring back the principles of lawfulness, fairness, and transparency to RTB (Article 5(1) a-f and 5(2) GDPR).

5.4 FURTHER RESEARCH

Further research in the field of WPM may focus on legal and technical aspects of RTB. Below, we propose three new directions for further research.

1. PROTOTYPE A REPORTING PLUG-IN

Now that we have virtually developed a new technology (the GBMA) to differentiate between RTB networks, a next step could be to develop a prototype of a plug-in for GDPR compliance reporting. Prototyping could help publishers to better understand the RTB technology which they embed in their webpages. In this thesis we have learned that data leakage to unknown parties who may process the RTB data for unknown purposes has become a risk for the publisher. Especially because accountability is a key element in the GDPR. This means that the burden of proof with regards to the lawfulness of processing of online identifiers for RTB purposes rests (first and foremost) upon the publisher.

2. DEVELOPMENT OF NEW A MEASURES

Further research into partner networks should include the development of new a measures. For instance, new a measure may be derived from the node types, e. g., the interconnecting node (Definition 4.19). In fact, the proportion of edges connected to a partner network of an interconnecting node can be compared with the size of a partner network (cf. Blondel, Guillaume, Lambiotte, and Lefebvre [2008, Figure 4, p. 7]). Such a measure would allow us to zoom in (1) on interconnecting nodes in large(r) referrer graphs and (2) on the purpose(s) of the processing of the data flowing through the nodes.

3. INVESTIGATE PRIVACY-FRIENDLY SSP TECHNOLOGY

Further research should focus on Supply Side Platforms, in particular on privacy friendly SSP header bidding. For instance, one may encourage research in SSP bidding strategies for privacy-friendly *contextual* advertising in contrast to privacy-intrusive *(re)targeting* business models. Such research could improve RTB bidding strategies by taking notice of expressions of an end-user (1) exercising his right to object, (2) granting consent, or (3) revoking his consent.

We already know that (a) the end-user, (b) the publisher, (c) the SSP, and (d) the advertisers form a network. The publisher is in a unique position due to the direct connection to an end-user and the SSP. The SSP is in a unique position due to the direct connection to a publisher and the advertisers. We mention two possibilities.

First, special attention could be given to embedding the end-user's granular consent expressions by using technical settings of a browser square in the center of the design of SSP header-bidding protocols. Second, special attention could be given to (further) exploration of the combination of granular consent expressions with centralized tag-management solutions. For instance, firing a pixel can be made conditional on the W3C DNT settings of a browser (cf. Fed [2018, pp. 1–2]). Such use of Tag-Based Integration is, already today, available for publishers and SSPs.

In closing, we remark that the publisher can increase competition by connecting to multiple SSPs (cf. Jauvion et al. [2018]). Together they can guard the rights and freedoms of the end-user, among others, by seeking consent to be relied upon by multiple controllers (as defined in Article 4(7) GDPR).