Enhancing user transparency in online ads ecosystem with site self-disclosures

Lukasz Olejnik

Email: me@lukaszolejnik.com.com

Site: lukaszolejnik.com

Abstract—**Transparency is a key ingredient in building user awareness. It is not only a regulatory requirement, but also a set of concrete technical rules of how systems should be implemented and operated. The changing environment in Europe in relation to the General Data Protection Regulation renews interest in designing systems with built-in transparent capabilities. In this paper, we perform a longitudinal study of a new standard gaining importance in the programmatic ads ecosystem, Ads.txt. We show that the adoption of Ads.txt has rapidly increased (178-fold, during the study period) and is now included on over 18,1% most popular 100k sites. We study the patterns seen in the configuration of ads.txt files as defined by websites. We also identify over 940 unique hostnames involved in real-time bidding auctions.**

Although the standard was originally intended to increase transparency in the interaction between players in Real-Time Bidding (RTB) auctions, we explore a different avenue. We use ads.txt to construct a proof-of-concept transparency enhancing technology, informing the user when a website self-discloses participation in the Real-Time Bidding ecosystem. Our analysis and demonstrations provide insight into increasing user transparency in the modern programmatic ads ecosystem.

I. INTRODUCTION

Transparency is one of the key ingredients building user awareness and trust. In a system with support of transparency, information about the implications of using a particular service should be easily available to the users. At the very least, this information could describe the ways that user data is or may be used. Transparency is the core assumption of many privacy regulations, particularly the European General Data Protection Regulation (GDPR) [1]. The 2017 Opinion of European Data Protection Authorities hihlights these relevant requirements [2].

The main goal of transparency is building of user awareness, it should therefore be understood as a core property of systems designed with privacy in mind, and using privacy engineering methods. In this sense, transparency can be understood as privacy awareness [3], and can be technically enabled by transparency-enhancing technologies (TETs) [4], as opposed to privacy-enhancing technologies (PETs). Transparency may be perceived as one of the fundamental properties of systems built with privacy engineering, and may require building of specialized transparency services [5]. Systems implementing privacy-enhancing technologies can also incorporate transparency properties [6].

In this work, we address one of the issues related with the current lack of standardised sources for automatic transparency-aiding systems on the web. We analyse the recent addition to the web advertisement landscape, called Ads.txt [7], a standard designed to enhance transparency in programmatic advertising ecosystems that takes advantage of ad exchanges working running real-time bidding auctions for ad slots [8], [9]. Creation of Ads.txt specification is motivated by the growing problem of fraudulent advertisements disseminated at the ad exchange auctions. This root issue lies in the design of RTB protocols, lacking ways to guarantee the identity of valid ad exchanges. A practical example scenario includes when buyers of advertising space on a publisher's site are not certain if an ad exchange is authorized to sell this particular publisher's ad space. Ads.txt offers a simple public record hosted at publishers' sites. This approach allows buyers (bidders) to cross-verify incoming bid requests with public ads.txt record. This is one of the proposed solutions to the so-called spoofing attacks, where sellers are falsely claiming to sell ad space on websites, and bidders have limited options of detecting these kind of frauds [10]. The ads.txt standard is rapidly gaining popularity, and some ad exchanges are even disabling options of buying ads inventory from networks not included on particular publishers' sites [11]. It is highly likely that the standard will become an industry requirement [12].

The standard already helped in the uncovering of one of the largest fraud botnets, HyphBot [13]. Real-time Bidding technology can be understood as an interaction between the bidders (working on behalf of advertisers), competing in auctions managed by ad exchanges for ad space on publisher's websites or applications. In RTB, bidders usually receive various information about the user's profile, typically including private data such as visited websites, the background behind the user such as the demographic description, and sometimes even location. Real-Time Bidding is not offering transparency to users; for example, users are unable to find out that RTB is taking place, or how many bidders receive the data concerning them. At the core of the Ads.txt standard lies an increase of transparency in online ads for publishers and bidders, rather than for the users. However, in this paper we show that Ads.txt provides an interesting transparency signal for the users as well. This is because until now there were no reliable signals flagging websites which include the use of Real-Time Bidding ads. Due to the open nature of Ads.txt specification as a self-imposed regulation in the online ads world, this information now becomes easily available for interpretation and research studies. In this paper, we introduce a simple transparency-

enhancing technology which takes advantage of ads.txt. This demonstration tool shows that it's possible to easily inform users about Real-Time Bidding scripts operating on websites. In this sense, the use of the word *users* should be understood as **data subjects** with respect to the visited websites.

Specifically, the main contributions of this work are:

- Proof-of-concept Chrome extension detecting websites taking advantage of Real-Time Bidding to increase transparency of user web browsing – we show that it is possible.
- Validation of Ads.txt prevalence on the Web and its evolution during the early days of Ads.txt.
- Study of Real-Time bidding ad exchanges.

Our work conforms to the INFORM privacy by design pattern described in [14]

A. Related Work

Transparency-enhancing technologies are standard tools intended to increase the privacy awareness of users, as reviewed by Janic [4]. The work of Angulo is one example of building a user interface of a transparency-enhancing tool [15]. Murmann and Fischer-Huebner offer an overview of transparency tools, focusing on usability factors [16].

Olejnik et al. provide a privacy and transparency study and analysis of Real-Time Bidding, the prevalence of this technology on most popular sites, and the insight behind the value of user data [8], [9]. This study was recently complemented by [17], and both studies introduce a browser extension able to analyze Real-Time Bidding. We note that these tools, which provide information to the users, are in fact transparency-enhancing technologies. In Real-Time Bidding, information flows aren't transparent to the users.

Englehard et al. provide an in-depth overview and description of principles in privacy measurement domain [18]. Narayanan and Reismen described the OpenWPM complex and its role in building broad transparency and accountability census [19]. This work is establishing privacy measurement as standard approach of testing web transparency state.

In this paper, we study a new standard of the web advertising ecosystem, and construct a browser extension to use the new signal to build user awareness.

II. ANALYSIS

We first describe the outcome of our multi-week ads.txt measurement. Subsequently, we introduce a tool that takes advantage of ads.txt standard to increase user transparency. Web crawling in order to understand web transparency is a standard practice [19].

A. Ads.txt standard

In this section, we provide a brief technical overview of the specification. Ads.txt are files hosted under exammple-publisher.tld/ads.txt. Their simplest format is shown below:

```
example.com, 42, DIRECT
example.org, 4536, RESELLER
```
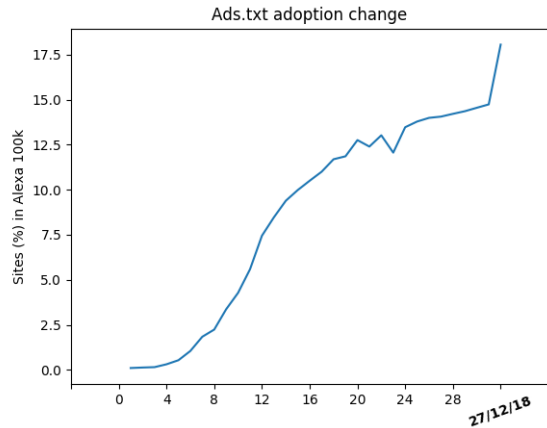

Fig. 1. Evolution of adoption of Ads.txt during first weeks of its introduction.

In the example above, the first column is the authorized ad exchange host name (i.e. example.com in the example above), the second column is the publisher's id (i.e. 42), and the third column describes their relationship (reseller means that space on the publisher's site may sold by another party, at the specified the ad exchange). By including this file, example-publisher.tld authorises its ad space to be sold at auctions of example.com and example.org. Although the specification also defines other columns, we are only interested in the host name of ad exchanges.

B. Measurement

To validate the feasibility of using Ads.txt as a transparency signal for users, we crawled 100.000 Alexas most popular sites. Our study took place over the course of 32 consecutive weeks, from August 2017 to March 2018. To enhance the longitudinal value of our work, we also perform a validation by making an additonal measurement in the end of December 2018; as such it is to our knowledge the longest longitudinal measurement of ads.txt standard use. For each crawled site, our crawler downloaded ads.txt files on a weekly basis. We used the structure of Ads.txt files conforming to the specification (i.e. the number of columns, the nature of values at each line) to process the downloaded files in order to identify valid data.

Although the specification is very simple, we discovered that some websites still had problems with the adoption of the standard. In many cases we saw ads.txt formatted incorrectly, for example typing ads.txt as a html file, or prefixing host names with a protocol (i.e. "http://"). We did not consider these kinds of files as correct Ads.txt and we excluded them from the results presented in this paper.

One of the contributions of this paper is a TET extension. In order to validate the idea of using Ads.txt as a transparency signal, we studied the evolution of Ads.txt adoption. Figure 1 shows the percentage (Y axis) of websites including Ads.txt each week (X axis) during the course of the study. We saw 101 (0.1% of the 100,000 site dataset) sites including ads.txt during the first crawl (06.08.17), this number growing to 14,741
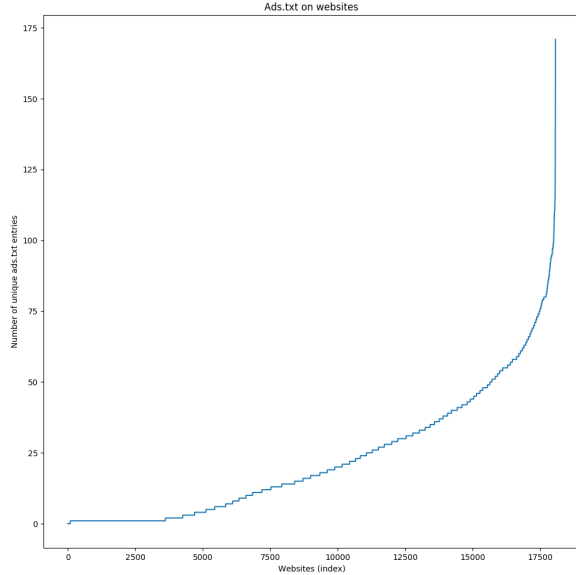
Fig. 2. Number of unique hostnames found in ads.txt (27.12.18)

google.com	0.96
appnexus.com	0.7
rubiconproject.com	0.68
openx.com	0.67
pubmatic.com	0.65
indexexchange.com	0.6
contextweb.com	0.56
spotxchange.com	0.54
spotx.tv	0.53
sovrn.com	0.49

TABLE I
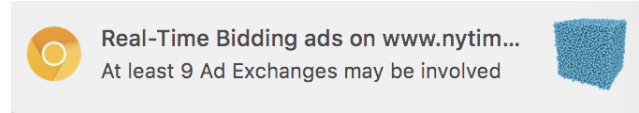FRACTION OF SITES INCLUDING HOST NAMES OF AD EXCHANGES
(27.12.18)

Fig. 3. A pop-up informing the user about existence and contents of ads.txt on the site of the New York Times

(14.7%, on 29.03.18), a 146-fold increase; and to 18,062 (18.6% on 27.12.2018' 178-fold increase), . Ads.txt adoption has been growing rapidly - it is in fact becoming an accepted industry standard.

It is interesting to analyse the latest crawl executed on 27 December 2018 in more detail. Figure 2 shows that for more than 50% of sites, ads.txt contain more than 20 unique entries; the average over whole dataset was 16.31 sites. Among the sites having the largest number of host names in ads.txt were Breitbart.com (US news tabloid; 174 entries), and in comparison Lemonde.fr, the French news site contains 54 entries.

The data also show which of the ad exchanges are included in the publisher's ads.txt most frequently (in other words, which are the most "popular"). It is important to realize that by this act of self-disclosure publishers infom about those ad exchanges likely processing data of website visitors. Table 1 displays the fraction of websites including hostname of a particular ad exchange in its ads.txt for most popular hosts of ad exchanges. We see that Google is by far the most prevalent ad exchange, included on 95% sites using RTB, followed by other key players in the ecosystem: AppNexus, Rubicon, OpenX and PubMatic. In total, we identified 941 unique hosts (an increase from 72 hosts found on 06.08.17, the date of the first crawl).

Finally, we also point out that our study is a reliable measurement of prevalence of RTB use on websites using publicly available data.

C. Extension

In previous section, we estimated the prevalence of Real-Time Bidding in 100.000 most popular Internet sites using the new method of voluntary self-disclosure by websites. We used crawl data to study transparency of online ads ecosystem, using ads.txt as a transparency vehicle. In this section, we introduce Transparency Enhancing Technology in form of a browser extension. Our tool parses ads.txt for each site visited by the user. If it finds a suitable file, the extension displays information about the site implementing Real-Time Bidding. Figure 3 shows the standard pop-up message displaying basic information contained in the ads.txt file, when the user visits a website including it.

We note that our extension can signal that a website is participating in the online ads ecosystem as a publisher even to the users that browse the web with ad blockers on. That is because ads.txt files are not blocked resources. In that, it provides a transparency signal even to the users who block ads. Although the information conveyed by the extension popup is very basic, we argue it is a proof-of-concept demonstration that a mechanism primarily intended to provide transparency in the online ads ecosystem between publishers and bidders can in fact be used as means of enhancing transparency for the users as well. We made the proof-of-concept extension available at *https://lukaszolejnik.com/adstxt/adstxt.crx*.

D. Discussion

The current debate around the role of online ads in supporting information of questionable nature underlines the need for transparency in the online ads ecosystem [20]. Ads.txt is an attempt to increase transparency and trust, but only between bidders and ad exchanges. There are currently no industry standard proposals that enable the user to easily discover the flow of data on the marketplace, or see the entities taking part in these transactions. It is difficult to predict if the current regulatory landscape, changing due to the

GDPR and ePrivacy regulations, will impact the motivation to offer more information to the user. In this situation, any available signals increasing understanding and clarity in real-time bidding transactions should be considered.

The concept of a tool we introduce in this paper stems from the fact that currently users have no option of seeing that a publisher is offering ad space at online auctions marketplaces. Yet, when a website self-discloses participation in the real-time bidding auctions this readily available information can be automatically parsed, interpreted and showed to user. We also answer the question of valuability of the data contained in ads.txt. We observe a stable growth in the adoption of ads.txt. We argue that this signal will not disappear (which is one of the issues in privacy measurements of this kind [18]), and that it provides reliable information, because websites have a monetary incentive to implement ads.txt. This, on the other hand, means that the signal can be considered stable - a source of information that should not suddenly disappear. This is a good premise for further study of Ads.txt and possible uses to enhance transparency of online ad ecosystems.

III. CONCLUSION

Ads.txt standard is a facility with a clear objective of decreasing fraud in programmatic ads. As such, user transparency is not among the intended goals. However, we have shown that it is possible to apply the mechanism to build tools enhancing user understanding and transparency on the Web. We note that even the simplest information channels in an otherwise closed ecosystem is valuable.

In this paper, we have shown that the ads.txt standard adoption is growing, and that it already provides useful insight about the role of real-time bidding in the web ads ecosystem. We indicate that over 18% of 100.000 most popular websites self-disclose participation in this ecosystem. We hope that the research community will find additional creative uses of ads.txt.

Finally, we expect that the regulatory changes related to GDPR and ePrivacy may motivate engineers and standardizers from the Real-Time Bidding domain to introduce new transparency features to Real-Time Bidding standards. We encourage academic researchers to actively pursue any new additions and streamlining them with web browser user interfaces.

REFERENCES

[1] "General Data Protection Regulation (Regulation 2016/679)," http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=en, 2016.

[2] WP29, "Guidelines on transparency under Regulation 2016/679 ," http://ec.europa.eu/newsroom/just/document.cfm?doc_id=48850, 2017.

[3] E. Kani-Zabihi and M. Helmhout, "Increasing service users privacy awareness by introducing on-line interactive privacy features," in *Nordic Conference on Secure IT Systems*. Springer, 2011, pp. 131–148.

[4] M. Janic, J. P. Wijbenga, and T. Veugen, "Transparency enhancing tools (tets): an overview," in *Socio-Technical Aspects in Security and Trust (STAST), 2013 Third Workshop on*. IEEE, 2013, pp. 18–25.

[5] M. Hansen, M. Jensen, and M. Rost, "Protection goals for privacy engineering," in *Security and Privacy Workshops (SPW), 2015 IEEE*. IEEE, 2015, pp. 159–166.

[6] O. Seneviratne and L. Kagal, "Enabling privacy through transparency," in *Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on*. IEEE, 2014, pp. 121–128.

[7] IAB, "ADS.TXT Authorized Digital Sellers," https://iabtechlab.com/ads-txt/, 2017.

[8] L. Olejnik, T. Minh-Dung, and C. Castelluccia, "Selling off privacy at auction," 2013.

[9] L. Olejnik and C. Castelluccia, "To bid or not to bid? measuring the value of privacy in rtb," 2015.

[10] S. Sluis, "Domain Spoofing Be Gone: Ads.Txt Will Filter Out Imposter Sites ," https://adexchanger.com/ad-exchange-news/domain-spoofing-gone-ads-txt-will-filter-imposter-sites/, 2017.

[11] The AppNexus Team, "An AppNexus Update on Ads.txt Implementation ," https://blog.appnexus.com/2017/appnexus-update-ads-txt-implementation/, 2017.

[12] G. Slefo, "Tag tells publishers to adopt ads.txt," http://adage.com/article/digital/tag-tells-publishers-adopt-ads-txt/311976/, 2018.

[13] Adform Fraud Protection, "How Adform Discovered HyphBot: One of the Largest Botnets to Ever Hit Digital Advertising," https://site.adform.com/media/85132/hyphbot_whitepaper_.pdf, 2017.

[14] "General Data Protection Regulation (Regulation 2016/679)," https://www.cs.ru.nl/~jhh/publications/pdp.pdf, 2016.

[15] J. Angulo, S. Fischer-Hübner, T. Pulls, and E. Wästlund, "Usable transparency with the data track: a tool for visualizing data disclosures," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2015, pp. 1803–1808.

[16] P. Murmann and S. Fischer-Hübner, "Usable transparency enhancing tools."

[17] P. Papadopoulos, P. R. Rodriguez, N. Kourtellis, and N. Laoutaris, "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*. ACM, 2017, pp. 142–156.

[18] S. Englehardt, C. Eubank, P. Zimmerman, D. Reisman, and A. Narayanan, "Web privacy measurement: Scientific principles, engineering platform, and new results," *Manuscript posted at http://randomwalker. info/publications/WebPrivacyMeasurement. pdf*, 2014.

[19] A. Narayanan and D. Reisman, "The princeton web transparency and accountability project," in *Transparent Data Mining for Big and Small Data*. Springer, 2017, pp. 45–67.

[20] C. Fuchs, "Fake News Is Not Just The Word Of The Year, But A Danger To Democracy That Can Only Be Overcome By Adequate Legal And Media Reforms ," http://www.huffingtonpost.co.uk/entry/fake-news-media-reform_uk_5a393e24e4b0cebf48e9f7be, 2016.