Web Transparency for Complex Targeting: Algorithms, Limits, and Tradeoffs

Guillaume Ducoffe, Mathias Lécuyer, Augustin Chaintreau, Roxana Geambasu
Inria & U. Nice Sophia Antipolis, Columbia U.
guillaume.ducoffe@inria.fr, {mathias,augustin,roxana}@cs.columbia.edu

Big Data promises important societal progress but exacerbates the need for due process and accountability. Companies and institutions can now discriminate between users at an individual level using collected data or past behavior. Worse, today they can do so in near perfect opacity. The nascent field of *web transparency* aims to develop the tools and methods necessary to reveal how information is used, however today it lacks robust tools that let users and investigators identify targeting using multiple inputs.

Here, we formalize for the first time the problem of detecting and identifying targeting on *combinations of inputs* and provide the first algorithm that is asymptotically exact. This algorithm is designed to serve as a theoretical foundational block to build future scalable and robust web transparency tools. It offers three key properties. First, our algorithm is *service agnostic* and applies to a variety of settings under a broad set of assumptions. Second, our algorithm's analysis delineates a *theoretical detection limit* that characterizes which forms of targeting can be distinguished from noise and which cannot. Third, our algorithm establishes *fundamental tradeoffs* that lead the way to new metrics for the science of web transparency. Understanding the tradeoff between effective targeting and targeting concealment lets us determine under which conditions predatory targeting can be made unprofitable by transparency tools.

1. QUICK OVERVIEW

A primer on web transparency tools.

To address the big-data web's untenable opacity, a new set of transparency tools have been proposed recently [7, 4, 5, 6, 3, 9]. Generally speaking, they assume no insider information about how the data-driven web service operates and instead rely on a specific form of *black box testing* [2] to detect data use. Briefly, a transparency tool works as follows. First, it collects the results of a series of tests in which *inputs* vary, *e.g.*, browsing history [7], search history [4], emails [6], locations [9], or explicit profile information [3]. Second, by examining the observed *outputs* – *e.g.*, search results [4, 9], ads seen [3, 6], recommendations [6, 5], or prices [7, 5] – the tool deduces how the system personalizes its behavior based on this input. Finally, the tool's deductions are used as *hypotheses* that are further analyzed for implications by the tool's users, such as end users, journalists, privacy watchdogs, or federal investigators.

To be valuable to their users, transparency tools strive to meet three requirements:

1. **Scalability:** Each test may involve multiple preliminary steps to open a new web account and populating its inputs. These steps cannot always be automated and may be expensive (e.g., creating a Google account requires buying a new phone number). It is also important to keep resources to a minimum as the size of outputs/inputs grows.

2. **Accuracy:** The deduction that the tool provides should be *sound*, which means that it can be trusted not to originate from noise or other limitations of the experiments. The tool should also be *complete* which specifies that it rarely misses an important deduction.

3. **Broad Applicability:** Ideally, the same tool should apply not only to many different services but also various forms of data usage within those services, with only minor and intellectually simple changes.

Perhaps unsurprisingly, the first two requirements are often in conflict. The third makes the problem extremely challenging, and has barely been considered to date. With few exceptions [6, 3], previous transparency tools were designed for a specific service or usage in order to detect a particularly sensitive topic: price discrimination [7], search results personalization [4], censorship [9]. Only recently has development of widely-applicable, generic tools begun to be considered to allow generic data collections [5] and service-agnostic detection methods [6, 3]. Despite appearences, however, we find that even the latest transparency tools are limited in the kind of data uses they can support accurately and scalably. We believe that the biggest roadblock is the lack of support for detecting complex, multi-input targeting, which we find mandatory for building scalable, accurate, and broadly applicable tools.

Our new findings.

We prove that targeting that uses one or several combinations of N inputs can be detected and identified with asymptotically perfect accuracy, and that this only requires

SIGMETRICS'15, June 15–19, 2015, Portland, OR, USA.
ACM 978-1-4503-3486-0/15/06.
http://dx.doi.org/10.1145/2745844.2745896.

a logarithmic order $O(\ln(N))$ of experimental observations. To place our contribution in respect to prior work, this shows that a web transparency tool can remain scalable and accurate, without being strictly restricted to single input targeting. However, and this is where our contribution contrasts the most with previous results, it comes at a cost: the intensity of targeting (defined below) needs to be sufficiently large. In other words, web transparency need not always be blind to combined targeting, however as the inherent complexity of targeting increases, targeting becomes easier to conceal. Our results open a new chapter in the understanding of Big Data: to determine sufficient and necessary conditions under which one can prevent its opacity.

2. PROBLEM FORMULATION & RESULTS

We formalize the following intuitive problem: given a set of N inputs representing possible information items present in a user's account (such as emails or searches), we wish to determine how they affect occurence of one particular output of interest (such as an ad or a recommendation).

Our main assumption is that the output is affected through an unknown targeting function f of the inputs, to be determined. The function f is defined separately for each output. The targeting function f is a mapping from the set of all combinations to $\{0;1\}$. By convention, $f(\mathcal{C}) = 1$ indicates that an account containing $\mathcal{C}$ is targeted, and we denote $f(.) = 0$ if the ad is untargeted.

Experiments and outcome properties.

Because in practice we have no access to the targeting function, we rely on experiments to observe its reaction to various inputs. Intuitively, these experiments collect outputs from a set of accounts that contain subsets of the inputs and produce a set of observations of f. For example, experiments could collect ads for accounts with different subsets of emails. More formally, the experimental infrastructure we assume is similar to an *oracle* from function learning theory [8, 1]. We assume that our experimental oracle satisfies the following axiom. There exist two probabilities p_{in}, p_{out} such that:

$$\mathbb{P}\left[\mathcal{O}(\mathcal{C}_i)=1 | f(\mathcal{C}_i)=1\right] \geq p_{\text{in}} > p_{\text{out}} \geq \mathbb{P}\left[\mathcal{O}(\mathcal{C}_j)=1 | f(\mathcal{C}_j)=0\right].$$

where p_{in} is a minimal bound on the probability that an account receives an output that is relevant for it and p_{out} is a maximal bound on the probability that an account receives an output that is not relevant for it. This axiom properly states that f is related to the outcome we study. It allows the variables to also depend on other factors: hidden inputs that are not in the set of N we study, external sources of randomness such as availability of ad-slot, competition. One experimental design used in practice [6, 3] and that fits this axiom is to populate each account randomly so that an input independently appears with probability α.

Under the assumptions above, we say that an algorithm using m observations solves the targeting detection problem if it can correctly decide whether $f(.) \neq 0$ and hence that the output is targeted using at most m queries to $\mathcal{O}$. Going further, an algorithm solves the targeting identification problem if it correctly returns the function f. Naturally, both problems rely on random observations and hence our goal is to design algorithms whose detection/identification error is arbitrarily small for large N.

Since one should distinguish (at least) between N inputs, it seems that a minimum of $\Omega(\ln(N))$ binary observations are absolutely necessary at least for the identification. This is hence what we assume and we aim at keeping it at this absolute minimum.

THEOREM 1. *Assuming that f is a monotone DNF with size at most s and width at most w, and that ratio p_{out}/p_{in} is below a predetermined bound, we provide a targeting detection algorithm that for any $\varepsilon > 0$ requires $O(\ln(N/\varepsilon))$ observations, $O(N^s \ln(N/\varepsilon))$ operations and is correct with probability $(1 - \varepsilon/N)$.*

THEOREM 2. *Under the same assumption, we provide a targeting identification algorithm that for any $\varepsilon > 0$ requires $O(\ln(N/\varepsilon))$ observations, $O(N^{s+w} \ln(N/\varepsilon))$ operations and is correct with probability $(1 - \varepsilon/N)$.*

3. ACKNOWLEDGEMENTS

This work was supported by DARPA Contract FA8650-11-C-7190, NSF CNS-1351089 and CNS-1254035, Google, and Microsoft.

4. REFERENCES

[1] D. Angluin. Queries and concept learning. *Machine Learning*, 2(4):319–342, Apr. 1988.

[2] B. Beizer. *Black-Box Testing*. Techniques for Functional Testing of Software and Systems. John Wiley & Sons, May 1995.

[3] A. Datta, M. C. Tschantz, and A. Datta. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *arXiv.org*, Aug. 2014.

[4] A. Hannak, P. Sapiezynski, A. M. Kakhki, B. Krishnamurthy, D. Lazer, A. Mislove, and C. Wilson. Measuring personalization of web search. In *WWW '13: Proceedings of the 22nd international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, May 2013.

[5] A. Hannak, G. Soeller, D. Lazer, A. Mislove, and C. Wilson. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *IMC '14: Proceedings of the 2014 Conference on Internet Measurement Conference*. ACM Request Permissions, Nov. 2014.

[6] M. Lecuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu. XRay: Enhancing the Web's Transparency with Differential Correlation . In *23rd USENIX Security Symposium (USENIX Security 14)*, San Diego, CA, 2014. USENIX Association.

[7] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting price and search discrimination on the internet. In *HotNets-XI: Proceedings of the 11th ACM Workshop on Hot Topics in Networks*. ACM Request Permissions, Oct. 2012.

[8] R. A. Servedio. On learning monotone DNF under product distributions. *Information and Computation*, 193(1):57–74, Aug. 2004.

[9] X. Xing, W. Meng, D. Doozan, N. Feamster, W. Lee, and A. C. Snoeren. Exposing Inconsistent Web Search Results with Bobble. *Passive and Active Measurements Conference*, 2014.