CS6290: Reading Summary 3

LING Shengchen
Dept. of Computer Science
ID: 57730900

I. SUMMARY OF PAPER [1]

A. Problem Statement

Current web services lack transparency on data, thus common users have little insight on how their data are been used, and investigators have little robust tool to track these data.

B. Problem Significance

We live in a world full of personal data: locations, addresses, emails, search histories, etc., and almost all privacy information are available by web services providers like Google, Amazon, and Facebook. Despite the data can help improving efficiency and customization, however, the exciting potential of these data may also threaten our data privacy and result in data abuses.

C. State of the Art

At present, tracking in a controlled environment has been achieved, e.g., in a modified operating system, which is often known as taint tracking systems [2][3][4][5]. However, tracking in an uncontrolled environment remains to be answered. Some researchers have proposed about "how much output is personalized" and "mainly in what type" [6][7][8][9], but not the concrete correlation between user data and corresponding output.

D. Contributions

1) *From the perspective of novelty of problem formulation,* the paper constructs four specific and practical scenerios to illustrate the problems and incompleteness of current approaches, and then leads to the goals that the new system should meet, which are 1) fine-grained and accurate data tracking, 2) scalability, and 3) extensibility, generality, and self-tuning.

2) *From the perspective of novelty of technical solution,* the paper proposes XRay to correlate designated inputs with outputs. The fundamental principle of XRay is the differential correlation mechanism that detects targeting by comparing outputs from various accounts that have similar subsets of data inputs but not exactly identical. In order to achieve this, XRay establishes several shadow accounts for every individual account, each of which comprising of distinct subsets of data, and have them ranked with a simple Bayesian model.

Particularly, XRay consists of three basic components: Browser Plugin, Shadow Account Manager, and Correlation Engine. See Fig.1 below.

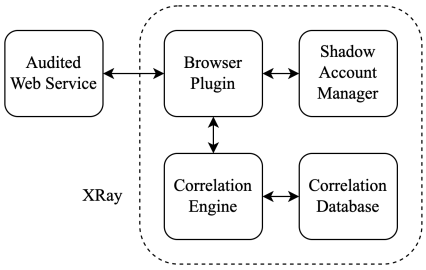

Fig. 1. XRay architecture

The Browser Plugin is to intercept inputs and outputs to/from the web services by recognizing DOM elements of web pages from web service providers. After receiving a new input, it forwards the input to the Audited Web Service and Shadow Account Manager. Shadow Account Manager is responsible for populating the Shadow Accounts with subsets of inputs and periodically collecting outputs from the Audited Web Service for each Shadow Account. After receiving a new output, it forwards to Correlation Engine to form a mapping correlation with its previous input and Correlation Engine will get it stored in the Correlation Database.

3) *From the perspective of theoretical analysis of algorithm.* The process that correlating outputs with inputs in Correlation Database is actually a time-consuming process, thus the paper uses a Correlation Algorithm in logarithmic time, asynchornously from user requests.

The simplest way is to use Naïve Non-Logarithmic Algorithm which traverses all possible combinations with unfortunately an exponential complexity. The paper defines a base algorithm of Threshold Set Intersection and extends it to a self-tuning Bayesian model, thus achieves automate adjustment of parameters via iterated inference for maximum correctness.

4) *From the perspective of positiveness of experimental evaluation result.* The paper evaluates XRay on Gmail ads, Amazon product recommendations, and YouTube video suggestions in four aspects: accuracy, scalability, managing overlaps, and practical usefulness. For all cases, XRay achieved over 80% precision as well as recall, which is in line with expectations.

E. Remaining Questions

1) In reality, the targeting processes may not be completely independent with each other, as well as the accounts, thus other covert factors should also be taken into consideration.

2) The paper just literally assumes that targeting noise is smaller than targeting singnal, which makes little sense. How to distinguish between these two still remains considerations.

II. SUMMARY OF PAPER [10]

A. Problem Statement

Web services typically collect users' data and leverage them in various ways; however, no one can exactly tell how the data are actually being used.

B. Problem Significance

Today's web services are growing rapidly, large amounts of privacy data are collected, stored, analyzed, and even mutually exchanged by service providers. The abuses of data leaving user privacy exposed under risks, thus there is urgent demand for targeting detection infrastructures, yet little progress has been found in this particular literature.

C. State of the Art

Previous research has studied the problem, however, there exists a significant gap in various aspects. Some of them are hard to scale, especially for those conducted controlled experiments by varying data inputs *one at a time* and observe the changes in outputs [11][12][6][7]. Most of them lack formal assessment of confidence, let alone providing statistical confidence except AdFisher [13]. Generic methodologies like XRay [1] and AdFisher [13] are scarce, and few have designed evaluation process; they use manual assessments instead [1][13]. In other words, the results prior works are not believable, persuasive, and at-scale enough.

D. Contributions

1) *From the perspective of novelty of problem formulation,* the paper constructs two scenerios with targeting detection problems, based on which, broadens the scope to show the basic requirements to design such a system: high confidence and interpretability.

2) *From the perspective of novelty of technical solution,* the paper proposes Sunlight, a system for targeting detection. The core innovation of Sunlight is to separate the operations into four phases, organize them in a pipeline, then further analyze and identify them by statistics and machine learning to reveal potential correlations between inputs and outputs.

Before the four phases, Sunlight randomly splits the data into two parts using a sample-splitting approach (also known as "holdout method" in machine learning), one for generating hypotheses called "training set", the other for verifying called "testing set".

The first phase, scalable hypothesis generation, is to create multiple plausible targeting hypotheses correlating inputs with outputs by using sparse linear regression method called Lasso [14] to estimate simultaneously. The second phase, interpretable hypothesis formation, is to transform the targeting hypotheses into an interpretable form that users can easily read and understand. The third phase, statistical significance testing, is to establish the statistical significance score (specifically a p-value) of the targeting hypotheses by testing veracity in another independent dataset. Finally, the fourth phase, multiple testing correction, is to test multiple hypotheses in one dataset to decrease the chance of misjudgment of individual hypotheses, both Holm-Bonferroni method [15] and Benjamini-Yekutieli procedure [16] are applicable in this case. See Fig.2 below for the procedures of Sunlight.

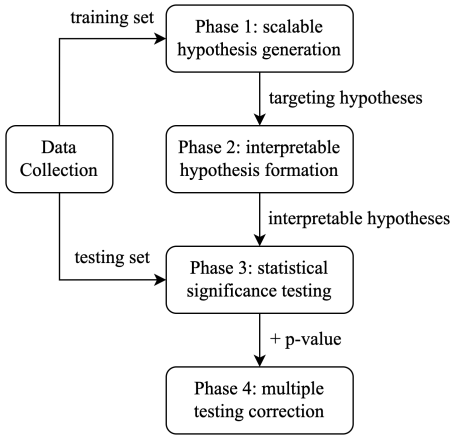

Fig. 2. Sunlight architecture

Sunlight also implements the four phases in a modular manner to enable instantiation and evaluation of each phase. In this way, users can flexibly change various mechanisms whenever needed, e.g., using either linear regression or Bayesian Algorithm in phase 1 will be fine, while choosing different mechanisms may lead to different terms of tradeoffs between scale and confidence of the results.

3) *From the perspective of positiveness of experimental evaluation result.* The paper also implements an evaluation of Sunlight compared with multiple prior arts on Gmail and on the web, showing that the high-confidence hypotheses are precise and well suited, but also revealing contradictions of statements from the claimed policies and FAQs.

Particularly, the evaluation makes use of the split of dataset and modularity to measure the effectiveness in different phases, and leverages a multitude of metrics of use cases that are originally designed for various testing situations and goals.

E. Remaining Questions

1) The statistical significance p-value can help in the process of hypotheses generation, thus forms a iteration to improve the accuracy of the hypotheses in the very beginning.

2) The data used in the experiments are intentionally generated instead of actual production data, with which it will be more rational.

REFERENCES

[1] M. Lécuyer, et al., "XRay: Enhancing the Web's Transparency with Differential Correlation", in *Proc. of the 23rd USENIX Security Symposium*, 2014.

[2] W. Cheng, Q. Zhao, B. Yu, and S. Hiroshige, "Tainttrace: Efficient flow tracing with dynamic binary rewriting", in *Proc. of the 11th IEEE Symposium on Computers and Communications*, 2006.

[3] W. Enck et al., "TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones", in *Proc. of the USENIX Symposium on Operating Systems Design and Implementation*, 2010.

[4] D. B. Giffin et al., "Hails: Protecting data privacy in untrusted web applications", in *Proc. of the 10th USENIX Conference on Operating Systems Design and Implementation*, 2012.

[5] Y. Zhu, J. Jung, D. Song, T. Kohno, and D. Wetherall, "Privacy scope: A precise information flow tracking system for finding application leaks", *Technical Report UCB/EECS-2009-145*, 2009.

[6] A. Hannak et al., "Measuring personalization of web search", in *Proc. of the 22nd International Conference on World Wide Web*, 2013.

[7] X. Xing et al., "Exposing Inconsistent Web Search Results with Bobble", *Passive and Active Measurements Conference*, 2014.

[8] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris, "Detecting price and search discrimination on the internet", in *Proc. of the 11th ACM Workshop on Hot Topics in Networks*, 2012.

[9] L. Sweeney, "Discrimination in online ad delivery", *Communications of the ACM*, 2013.

[10] M. Lecuyer, et al., "Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence", in *Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015.

[11] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan, "Adscape: harvesting and analyzing online display ads", in *Proc. of the 23rd international conference on World Wide Web*, 2014.

[12] T. Book and D. S. Wallach, "An Empirical Study of Mobile Ad Targeting", arXiv:1502.06577, 2015.

[13] A. Datta, M. C. Tschantz, and A. Datta, "Automated Experiments on Ad Privacy Settings", in *Proc. of Privacy Enhancing Technologies*, 2015.

[14] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society*, Series B 58 (1994), 267–288.

[15] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure", *Scandinavian Journal of Statistics 6*, 2 (1979), 65–70.

[16] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency", *Annals of statistics*, 2001.