Review on "The best privacy defense is a good privacy offense: obfuscating a search engine user's profile"

Joshua Fenech *, **and Omar Elsabrout** *

*Jean Monnet University

Submitted to the faculty of science and technology at Jean Monnet University

Since the growth rate of the Internet is almost exponential, it has a drastic effect on people's lives everyday. Such effect is accompanied with many issues such as user privacy. It is not settled until now how service providers are allowed to use stored information of the users and compromise their privacy. Jörg Wicker and Stefan Kramer introduce a tool that utilizes machine learning and data mining to confuse search engines to protect the user's privacy by obfuscating exploited personal information. Not only methods are introduced for such technique, but also an experiment to evaluate its results indicating whether this approach should be investigated further.

Introduction

The motivation behind providing this approach is that the current privacy protection have major flaws. On one hand, users rely on service providers to process their data, on the other hand, providers do not have any advantage in user privacy preserving technologies as the analysis of this data and sharing it with advertisers is the basis of their business model. No doubt that providers such as search engines must store information on users. Nevertheless, these data can be used to generate detailed profiles on a large scale and identify unidentified users. As mentioned, Jörg Wicker and Stefan Kramer target privacy from another perspective. Hence, they suggest a user tools to defend her or his privacy so the user does not have to rely on the other uncontrollable side for this issue. Data is conventionally stored in large scales and analyzed automatically using data mining technologies. As a result, the intuitive approach to protect the users private information would be to flood the data storage with random data and hope the users interest or identity would be obfuscated. On the contrary, data mining algorithms are designed to distinguish a signal from random noise. Consequently, this approach will fail in most settings where the data is analyzed with sophisticated data mining algorithms. The paper tackles this issue in a more highly developed manner as it gives a brief discussion of the user and search engine model in addition to the proposed method. Moreover, it shows the details of the experimental set-up and results. Taking into account, the paper should be considered as proof-of-concept and not a final product ready for the market. In this review, we present pointers for the given approach and criticize it by explaining its strong arguments and also its weak ones. This is not meant to be a summary of the tackled paper as we merely mention concepts and do not dive into details.[3].

Discussion

One of the first obstacles that is faced and discussed in the paper, is the user model. The authors admit a problem of major simplification for the user's interest categories as a result of reducing these categories to one at a time. The reason behind such simplification is to simplify the evaluation so they can compare the reaction of the search engine to one interest category. Although they stated that future work will address users with multiple interest categories and users with variably strong interest in multiple categories, we think that such simplification needs to be accompanied with assurances of the ability to upscale the model for multiple categories or at lease some pointers for such ability. These assurances are not provided which leaves the problem unattended. Nonetheless, another simplification is justified which is to only address query results that provided ads due to the problem's nature.

Another obstacle that might be rather unavoidable is that obfuscating a search engine user's profile, even though if it protects the user's privacy, reduces the user's experience quality. A user who utilizes this approach will lose most of personalized services and automated customizations provided by search engines if not all of them. We know that this is a decision that users must make. However, it is a hard one as these services are more practical and useful so the trade-off might not be fair. The author's assumption that user prefers privacy overlooks the idea that the number of those users can be low in comparison with the ones who prefer personalization. Despite that, we think that it is still perfect to execute the obfuscation idea in form of a tool that can be easily deactivated if the user wanted. This does not affect the user's profile if he or she wishes to return to using the search engines normally.

In addition, obfuscating search engines and service providers waste their resources and invalidates their research results and statistical studies that might lead to new applications based on the user's needs. These needs are usually investigated by online surveys and mining user data which will be useless because of mining fake or randomized data.

Method

Since the full knowledge of the states and information of the search engine are not available, the algorithm behaves according to only the feedback from the search engine in the form of ads.

Simulations.

Simulation 1

Vivamus magna enim, alique

Reserved for Publication Footnotes

Discussion

Nam fermentum.

Materials and Methods

Suspendisse viverra eleifend nulla at facilisis. Nullam eget tellus orci. Cras sit amet lorem velit. Maecenas rhoncus pellentesque orci eget vulputate. Phasellus massa nisi, mattis nec elementum accumsan, blandit non neque. In ac enim elit, sit amet luctus ante. Cras feugiat commodo lectus, vitae convallis dui sagittis id. In in tellus lacus, sed lobortis eros. Phasellus sit amet eleifend velit. Duis ornare dapibus porttitor. Maecenas eros velit, dignissim at egestas in, tincidunt lacinia erat. Proin elementum mi vel lectus suscipit fringilla. Mauris justo est, ullamcorper in rutrum interdum, accumsan eget mi. Maecenas ut massa aliquet purus eleifend vehicula in a nisi. Fusce molestie cursus lacinia.

Definition 1. *A bounded function θ is a weak solution of QG if for any $\phi \,\epsilon\, C_0^\infty(\mathbb{R}/_{\mathbb{Z}} \times \mathbb{R} \times [0,\varepsilon])$ we have*

$$\int_{\mathbb{R}^+ \times \mathbb{R}/_{\mathbb{Z}} \times \mathbb{R}} \theta(x,y,t)\, \partial_t \phi\,(x,y,t) dydxdt + \\ + \int_{\mathbb{R}^+ \times \mathbb{R}/_{\mathbb{Z}} \times \mathbb{R}} \theta\,(x,y,t) u(x,y,t) \cdot \nabla\phi\,(x,y,t) dydxdt = 0 \quad [1]$$

where u is determined previously.

Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Mauris eu sapien nunc, sit amet accumsan dui. Nulla ac diam ut nunc placerat semper eget et libero. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Cras hendrerit ullamcorper sapien vitae luctus. Quisque vel diam massa. Vestibulum dui nibh, facilisis vel vestibulum eu, viverra in quam.

Theorem 1. *If the active scalar θ satisfies the equation* **[1]**, *then φ satisfies the equation*

$$\begin{aligned} \frac{\partial\varphi}{\partial t}(x,t) &= \int_{\mathbb{R}/_{\mathbb{Z}}} \frac{\frac{\partial\varphi}{\partial x}(x,t) - \frac{\partial\varphi}{\partial u}(u,t)}{\left[(x-u)^2 + (\varphi(x,t) - \varphi(u,t))^2\right]^{\frac{1}{2}}} \\ & \quad \chi(x-u, \varphi(x,t) - \varphi(u,t)) du \; + \\ &+ \int_{\mathbb{R}/_{\mathbb{Z}}} \left[\frac{\partial\varphi}{\partial x}(x,t) - \frac{\partial\varphi}{\partial u}(u,t)\right] \\ & \quad \eta(x-u, \varphi(x,t) - \varphi(u,t)) du + Error \end{aligned} \quad [2]$$

with $|Error| \leq C\,\delta|log\delta|$ where C depends only on $\|\theta\|_{L^\infty}$ and $\|\nabla\varphi\|_{L^\infty}$.

Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Integer accumsan ornare tortor at varius. Phasellus ullamcorper blandit dolor sit amet tempus. Curabitur ligula urna, ultrices in iaculis eu, eleifend vel urna. Praesent ullamcorper imperdiet purus, ut interdum sem interdum dictum. Proin euismod volutpat eros ac mattis. Quisque sit amet massa ac tortor cursus malesuada at vitae nisi. Nam quis neque et nunc vehicula cursus sit amet at tellus.

Appendix

An appendix without a title.

Appendix: Appendix title

An appendix with a title.

ACKNOWLEDGMENTS. This work was part of an assignment for a research methodology course.

1. M. Belkin and P. Niyogi, Using manifold structure for partially labelled classification, Advances in NIPS, 15 (2003).
2. P. Bérard, G. Besson, and S. Gallot, Embedding Riemannian manifolds by their heat kernel, Geom. and Fun. Anal., 4 (1994), pp. 374–398.
3. R.R. Coifman and S. Lafon, Diffusion maps, Appl. Comp. Harm. Anal., 21 (2006), pp. 5–30.
4. R.R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part I: Diffusion maps, Proc. of Nat. Acad. Sci., (2005), pp. 7426–7431.
5. P. Das, M. Moll, H. Stamati, L. Kavraki, and C. Clementi, Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction, P.N.A.S., 103 (2006), pp. 9885–9890.
6. D. Donoho and C. Grimes, Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data, Proceedings of the National Academy of Sciences, 100 (2003), pp. 5591–5596.
7. D. L. Donoho and C. Grimes, When does isomap recover natural parameterization of families of articulated images?, Tech. Report Tech. Rep. 2002-27, Department of Statistics, Stanford University, August 2002.
8. M. Grüter and K.-O. Widman, The Green function for uniformly elliptic equations, Man. Math., 37 (1982), pp. 303–342.
9. R. Hempel, L. Seco, and B. Simon, The essential spectrum of neumann laplacians on some bounded singular domains, 1991.
10. Kadison, R. V. and Singer, I. M. (1959) Extensions of pure states, Amer. J. Math. 81, 383-400.
11. Anderson, J. (1981) A conjecture concerning the pure states of $B(H)$ and a related theorem. in Topics in Modern Operator Theory, Birkhaüser, pp. 27-43.
12. Anderson, J. (1979) Extreme points in sets of positive linear maps on $B(H)$. J. Funct. Anal. 31, 195-217.
13. Anderson, J. (1979) Pathology in the Calkin algebra. J. Operator Theory 2, 159-167.
14. Johnson, B. E. and Parrott, S. K. (1972) Operators commuting with a von Neumann algebra modulo the set of compact operators. J. Funct. Anal. 11, 39-61.
15. Akemann, C. and Weaver, N. (2004) Consistency of a counterexample to Naimark's problem. Proc. Nat. Acad. Sci. USA 101, 7522-7525.
16. J. Tenenbaum, V. de Silva, and J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science, 290 (2000), pp. 2319–2323.
17. Z. Zhang and H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignement, Tech. Report CSE-02-019, Department of computer science and engineering, Pennsylvania State University, 2002.

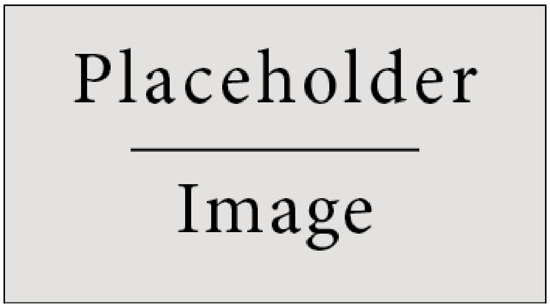

Fig. 1. Figure caption

Table 1. Table caption

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296