Preserving Privacy in Geo-Targeted Advertising

Ian Nathan Anggono[1], Hamed Haddadi[2], Abdelberi Chaabane[3] and Mohamed Ali Kaafar[4]

[1]University of New South Wales, Australia

[2]Queen Mary University of London, UK

[3]Northeastern University, USA

[4]CSIRO Data61, Australia

ABSTRACT

Targeted advertising has inherent privacy risks: ad providers aim to maximize the information inferred about the users in order to increase their click-through ratio. This in turn leads to long-term privacy risks for the users as their information is traded among ad agencies and unknown third parties. In this paper we focus on the privacy associated with information such as location and income, and their relationships with the ads served to users. We first show the possibility for an attacker to use topic modelling and machine learning techniques on ads served to a user as a means to accurately infer their location and income band within a city. We then attempt to reduce this risk of inference using obfuscation, or hiding in plain sight. The idea is to hide the targeted ads, not by relying on encryption, rather by producing just enough noise such that an attacker cannot distinguish between the actual ads served and the ads which are just noise. Our results are promising and demonstrate that efforts such as *TrackMeNot* can help advertising and users achieve a balance between targeted advertising and location privacy exposure.

Keywords

Advertisement, Privacy, Machine Learning

Categories and Subject Descriptors

K.4 [**COMPUTERS AND SOCIETY**]: Privacy

1. INTRODUCTION

The Internet ecosystem is largely fuelled by the advertising industry and the trade of personal data. A major issue with this current Internet norm is the lack of transparency and balance between the users' right to privacy and the effectiveness of targeted online advertising when using social media and web services. The current all-or-nothing approach enforced by tech giants and advertising brokers leads to high privacy risk for individuals. On the 3^{rd} December 2013, Facebook applied for a patent on *"Inferring Household Income for Users of a Social Networking System"* [13]. Using this patented algorithm, Facebook aims to infer a user's income by analysing the data from the posts on their timeline. The need for this algorithm is explained by the economics rationale that ads for cars, home mortgages, or holidays needs to be targeted to users based on their income band. An example of inferences used by such targeted advertising, according to a statement by Facebook, is the assumptions that those with higher income will post more about CNN.com and nytimes.com, instead of tabloids such as TMZ.com and PerezHilton.com. Targeted advertising may be contextual, interest, location-based. The threat of privacy invasion arises when an honest but curious third party can track the users' location based on the delivered advertising. These data can then be traded in a complex web of thousands of third parties [4]. Today, such inference attacks can be done through interception of web traffic since most ads are not encrypted.

Recently, a number of systems and approaches have been proposed to try and improve the advertising ecosystem in terms of user privacy. Privacy advocates and researchers alike have come up with solutions like *Do Not Track*[1] or Privad [6]. However, due to the lack of incentives and regulatory pressure, big companies like Google and Facebook are not interested in giving up data acquisition from users just to increase user privacy. While this is understandable from the business perspective, it has lead to user frustration. As Nissenbaum states, it is important to develop privacy enhancing technology as a reaction to unjust and uncomfortable data collection. Privacy is not complete control of our information nor is it perfect secrecy, instead it is appropriate information flow that is consistent with ideal informational norms [10]. As part of this ideology, obfuscation techniques and tools such as *TrackMeNot*[2] have been proposed which aim to increase user privacy by adding noise to a user's digital footprint.

In this paper we evaluate the effectiveness of obfuscation techniques in-the-wild. We collect thousands of ads from individual volunteers in London, and assess the ability to infer their detailed location (down to the street postcode level) using the type of (text) ads they were served by Google. We first demonstrate the possibility for an attacker to infer a user's accurate location and income by collecting the user's ads and applying topic modelling and machine learning classifiers. In the next part of the paper, we demonstrate the ability to reduce the accuracy of an attacker's prediction by obfuscation in order to improver a user's privacy.

The rest of the paper is organised as follows. In Section 2 we discuss the recent related efforts in the space of advertising privacy.

TargetAd 2016: 2nd International Workshop on Ad Targeting at Scale, WSDM, February 22–25, 2016, San Francisco, CA, USA.

DOI: 10.1145/1235

[1]http://donottrack.us/

[2]http://cs.nyu.edu/trackmenot/

Table 1: London postcode categories in the dataset

Category A	Category B	Category C	Category D	Category E
BH51EE	E1 4NS	CB3 0FD	SW19 8PR	N19
E3 4NZ	E1 4PQ	E15	W14 8BD	-
-	NW1 3HZ	-	-	-
-	E6	-	-	-

In Section 3 we describe our data collection tool and the collected dataset. In Section 4 we demonstrate the results of our inference attack simulations. In Section 5 we demonstrate the effect of obfuscation on privacy and accuracy of inference. Finally we conclude the paper in Section 6 and discuss avenues for future research.

2. RELATED WORKS

Peddinti and Saxena [11] perform a case study of the TrackMeNot tool for preserving search privacy using obfuscation. Mowbray *et al.* [9] have also suggested a client-based method for obfuscating search results and are developing a proof-of-concept tool for evaluation. In [7], Haddadi *et al.*suggest obfuscation and noise addition for aggregate large scale analytics while respecting individuals' privacy.

Angiuli *et al.* [2] demonstrate the data de-identification techniques when k-anonymization is used for anonymizing data in educational platforms. Their techniques demonstrate that perfect anonymity and privacy are hard to achieve while keeping accuracy, even when distortion is applied. We observe similar results in our classification approach when obfuscation is applied.

In this paper we use advertising traces collected from individuals in London as a ground truth for establishing the possibility of inference and obfuscation.

3. DATASET

We developed a Firefox plugin[3] and asked volunteers to install it to collect ads from users around the world. The extension was developed for the purpose of understanding the geographical diversity of online advertising in search engine. Each volunteer user was asked to fill out basic details upon extension installation, e.g., country, city, and postcode. The plugin would then run silently in the background and simulate Google searches using over 500 of the top searched keywords, though we excluded explicit keywords. The keywords include laptop, holiday, hotels and so on. We did not collect any personally identifiable information, nor the user's browsing history. After each search the plugin collects the ads received from the search query. As a result of this experiment, which ran for almost 9 months between 2013 and 2014, a sum of 15,729 unique ads were collected. Associated with these ads were 44 different users from 14 different countries, 19 cities and 33 different postcodes.

Due to the voluntary nature of this experiment, the number of ads collected from different places are great in range, for example about 5k ads are from United Kingdom (UK) while only around 200 are from Iran. Hence we focus on the largest geographic area from which the ads were collected, namely London, UK. A valuable challenge for an advertiser or inference attacker is to try to pinpoint the precise location and income of a user within a big city such as London. In other words, it will be easy for an attacker to figure out whether a user is in London or Stockholm based on ad targeting, while detecting the exact postcode (down to the street level in the UK), is not always feasible.

[3]http://planete.inrialpes.fr/bubbles/

Table 2: Average income band for postcode categories

Category	Income (Avg).	Range of income (weekly)	Number of Ads
A	£260.00	0-520	7
B	£555.00	521-590	3467
C	£630.00	591-670	59
D	£730.00	671-790	1127
E	£790.00	790+	569

Our dataset contains longitudinal data from 11 different postcodes in London. For each of these postcodes we obtain the average weekly household income from the UK government's website.[4] Each postcode is then categorised into 5 different income brackets. Figures 1 and 2 show these categories, their income band, and the number of ads in each category. There are inherent biases in the dataset towards middle class and those in the slightly higher income bands present in our data. Hence when doing our inference, we focus on categories B, D, and E from which more data is available. It is worth noting that the lower number of ads from one category is also a signal of the differences in one community (e.g., those with very low income) which can help with identifying the members of that category, their income band, and eventually their locations in London.

4. INFERRING LOCATION AND INCOME FROM ADS SERVED

In this section we briefly assess the ability of an attacker to infer a user's income by predicting the user's location in our ad dataset from London postcodes. Initially, we used topic modelling, namely LDA (Latent Dirichlet Allocation), to automatically groups the ads in our corpus into 500 topics. We use the conventional 500-topic model to capture the variety present in ads without under-representation of larger issues or over-representing the smaller sub-topics.

After performing topic modelling, we wished to establish the ability to automatically infer user locations in the larger categories (B, D, and E, category A and C are ignored due to the small number of ads) using their ad topics. We applied a variety of machine learning algorithms to the topics, including logistic regressions, decision trees, support vector machine (SVM), k-Nearest Neighbours (k-NN), and naive Bayes classifiers. After a number of trial and errors on the dataset, as a determined attacker would, we found the Nearest Neighbour (IBk algorithm [1]) and SVM (SMO algorithm [12]) on average to have the highest precision and recall. Hence in this paper we present the results of our analysis using these algorithms only. We present the performance of our location classification approach using the F-measure, which takes into account both precision and recall of the classification algorithm when computing the accuracy, hence:

$$F - Measure = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

F-Measure combines precision and recall's harmonic means without bias towards either metric.

The initial results from SMO and IBk classifiers show an almost random probability among the three groups (33%). Our objective here is to show whether it is possible to predict one's location and income by observing their ads and the fact that only a small number of ads are collected. To train our classifiers, the original data is randomly added by labelled ads from any of the three-income bracket category. The result can be seen in Figure 1 for the use of SMO algorithm and Figure 2 for comparison that uses the algorithm IBk.

For looking at the classification power in terms of accuracy, we

[4]http://neighbourhood.statistics.gov.uk/

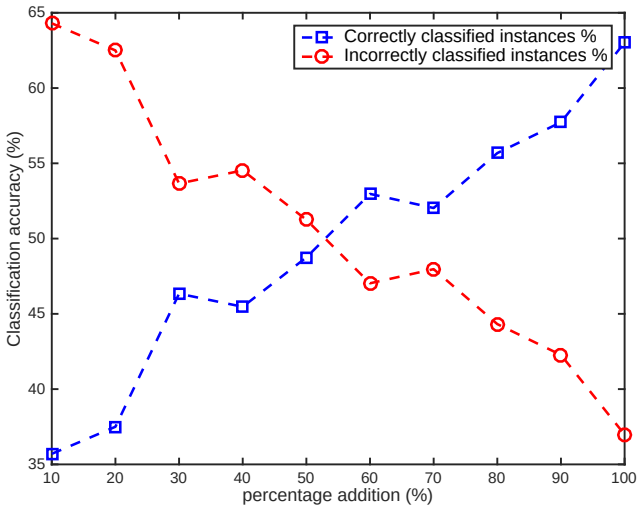

Figure 1: Accuracy of the SVM classifier as more training data is added.

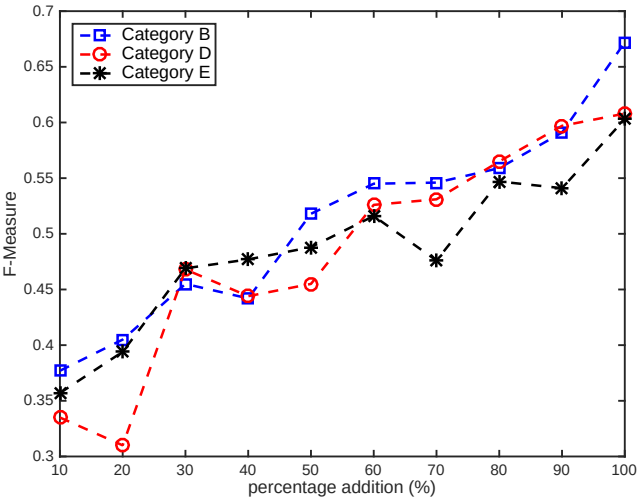

Figure 2: Accuracy of the Nearest Neighbour classifier as more training data is added.

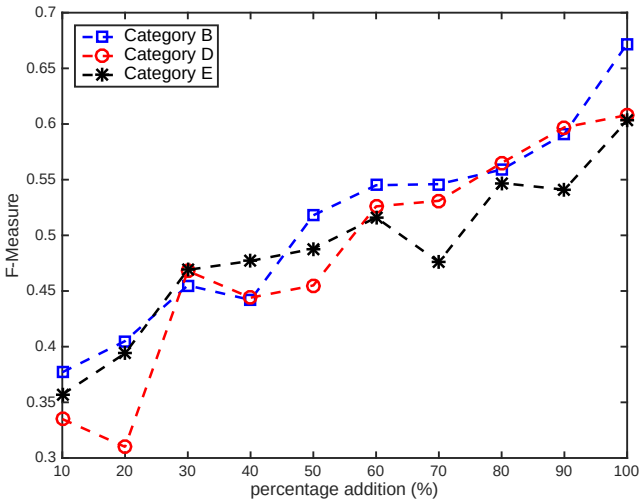

Figure 3: SVM income band classification accuracy, F-Measure vs %added, 80% Split .

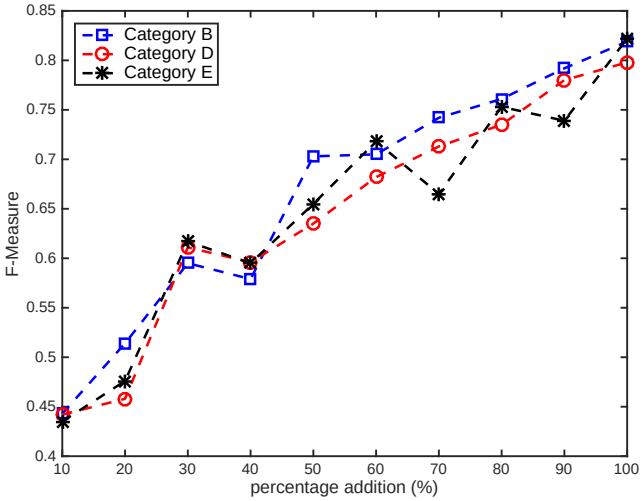

Figure 4: Nearest Neighbour income band classification accuracy, F-Measure vs %added, 80% Split.

use a percentage split of 80%training and 20% test data. Figure 3 and 4 show the F-Measure improvement for all 3 categories by adding ads from category B, D, and E. As expected, the percentage of correctly classified instances increases as more ads are introduced. Hence an attacker can gain great insights about the users' locations by collecting ads from certain areas and training a classifier. In the next section, we show how using obfuscation we can reduce this risk to a certain level. These figures confirm previous research showing that an attacker with access to a small set of Google ads can infer users' interests with an accuracy of more than 79%, and reconstruct as much as 58% of a user's interest profile [3].

5. REDUCING PRIVACY RISKS

In this section, our objective is to decrease the previously obtained F-Measures using obfuscation techniques. Again, we implement SVM and Nearest Neighbour using SMO and IBk algorithms respectively. The question now is what ads will be used to obfuscate the actual ads, i.e. ads that users want to obfuscate. Unlike click-fraud detections techniques such as Bluff ads [5], obfuscation needs to be covert enough such that an attacker cannot distinguish between actual ads and noise ads. For this reason, ads from United Kingdom (UK) besides those from London (77%), ads from Australia (7%) and USA (6%) have been chosen to form the noise ads. There are also ads from Italy, Hungary and other non-English speaking European countries, but an attacker can easily distinguish those ads using English language dictionary lookup. Ads are chosen at random and our evaluation metric of interest here is the F-Measure. The correctly classified instances here also include the machine correctly classifying the noise ads. The objective of an attacker is to predict the right category, not to classify noise ads to the noise category, i.e., precision is more important than recall. We examine three obfuscation strategies: (i) adding noise ads to the actual ads, (ii) removing actual ads while keeping a constant level of noise ads, and (iii) a combination of both techniques.

5.1 Obfuscation by Addition

ADDED ads from around England + Australia + USA (SMO, 80/20 split)						
	Correctly	Incorrectly	F-Measure			
%added	Classified	Classified	Category B	Category D	Category E	Noise
0%	34.0176	65.9824	0.359	0.335	0.326	-
10	31.2	68.8	0.383	0.302	0.3	0.464
20	29.5844	70.4156	0.549	0.549	0.445	0.526
30	25.0564	74.9436	0.328	0.274	0.216	0.174
40	28.3019	71.6981	0.291	0.311	0.206	0.316
50	29.7456	70.2544	0.244	0.279	0.213	0.379
60	32.6606	67.3394	0.233	0.195	0.086	0.485
70	34.715	65.285	0.17	0.165	0.108	0.517
80	42.5775	57.4225	0.144	0.194	0.069	0.603
90	42.813	57.187	0.131	0.092	0.029	0.616
100	47.4302	52.5698	0.108	0.092	0	0.648
Average F-Measure of categories			0.23412121			
Average F-Measure of noise			0.4728			

Figure 5: SVM classifier accuracy and the effect of noise addition.

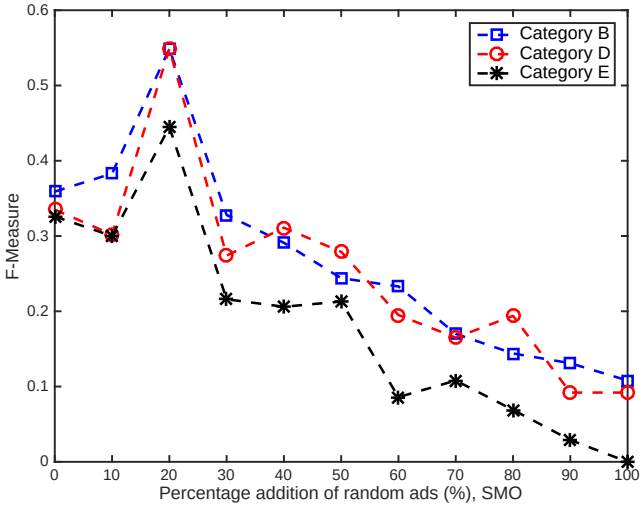

Figure 6: F-Measure response to noise addition using SVM classifier.

ADDED ads from around England + Australia + USA (IBk, 80/20 split)						
	Correctly	Incorrectly	F-Measure			
%added	Classified	Classified	Category B	Category D	Category E	Noise
0%	34.8974	65.1026	0.257	0.417	0.363	-
10	23.7333	76.2667	0.221	0.293	0.243	0.085
20	28.1174	71.8826	0.303	0.351	0.249	0.173
30	23.4763	76.5237	0.228	0.302	0.216	0.178
40	28.0922	71.9078	0.235	0.376	0.191	0.316
50	26.6145	73.3855	0.209	0.269	0.246	0.322
60	26.422	73.578	0.18	0.248	0.181	0.37
70	25.0432	74.9568	0.137	0.226	0.193	0.351
80	26.5905	73.4095	0.127	0.217	0.143	0.409
90	29.5209	70.4791	0.142	0.198	0.158	0.453
100	34.5081	65.4919	0.171	0.24	0.163	0.515
Average F-Measure of categories			0.23312121			
Average F-Measure of noise			0.3172			

Figure 7: NN IBk classifier accuracy and the effect of noise addition.

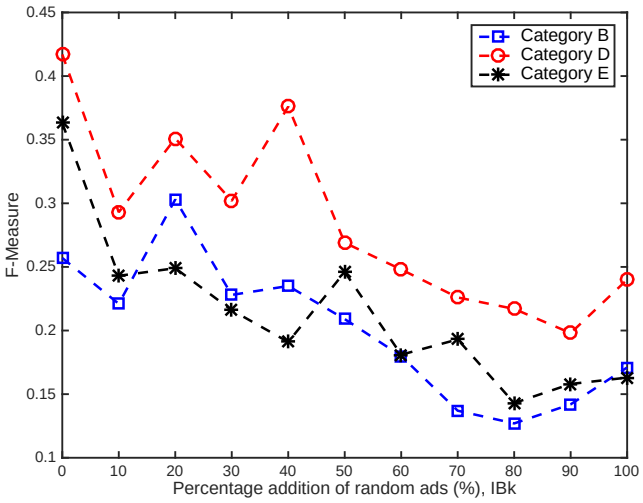

Figure 8: F-Measure response to noise addition using NN IBk classifier.

We first present the results from adding noise. Figures 5 and 6 demonstrate the SVM SMO classifier algorithm and the resulting F-measure, and Figures 7 and 8 show the same for Nearest Neighbour IBk algorithm. A quick comparison from the average F-Measures shows the IBk classifier to be marginally better at identifying categories however considerably better at identifying noises. The x-axis shows the percentage of random ads added to the dataset. For example 3,000 ads will have 10% noise of 300 random ads. As expected the F-Measure for all categories will decrease as the noise increase. The randomness in the added data is the cause of a non-monotonous decrease in F-measure at each step of the noise addition.

5.2 Obfuscation by Deletion

Figures 9 and 10 demonstrate the effect of ad deletion using SMO algorithm and IBk algorithm respectively. Rather than changing the noise threshold, we keep a constant noise for every test and we decrease the number of actual ads. If a user has high number of actual ads, the obfuscation needs to be high as well. The x-axis shows the percentage of the ads removed and the y-axis is the F-Measure. We observe that deletion of ads quickly decreases the F-measure, especially after the 50% cut-off for the SMO classifier. Despite lower initial ad numbers, the IBk algorithm performs pretty well for category E which can potentially be due to the lower diversity and divergence in the ads served to the user in that postcode.

5.3 Obfuscation by Addition and Deletion

In the final obfuscation experiment, we use an ensemble of the two techniques in order to maximise privacy while reducing the amount of noise addition or removal of ads. Figures 11 and 11 demonstrate the precision, recall, and the F-Measure for the SMO algorithm, and Figures 13 and 14 use the IBK algorithm. In this final test, we added extra noise to the original data , and removed actual ads simultaneously. The x-axis is the representative percentage of noise added *AND* the percentage of actual data removed simultaneously. For example at the 10% point on the x-axis, means that the dataset is added extra 10%noise and the dataset's actual ads are removed by 10% as well. The y-axis is the F-Measure scores.

The results of this test further establish that the F-Measure will decrease faster by the two different actions. To prove this, the average of addition only of noise to the dataset has the F-Measure average of 0.29 (mean [additions from 0%-70%]), while the average of noise additions and removal of actual ads has the F-Measure average of 0.2 (mean [add+removals from 0%-70%]). Hence we can deduce that: (i) the increase in the number of actual ads will increase privacy risk; (ii) the increase in the noise will decrease privacy risks; and (iii) Obfuscation can successfully increase the user privacy and decrease privacy risks associated with targeted

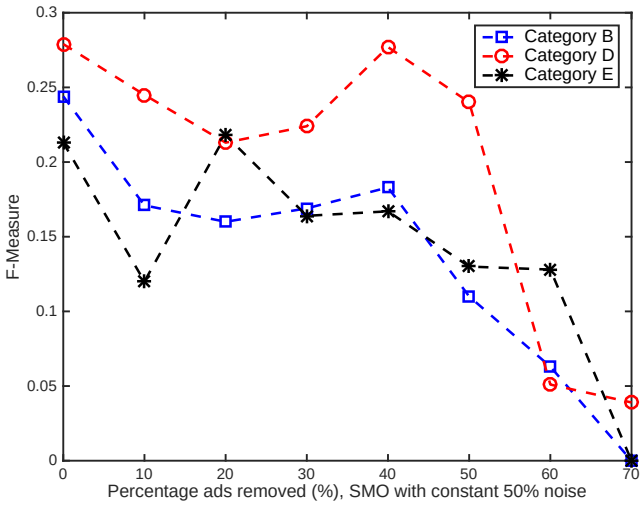

Figure 9: F-Measure response to ad deletion using SVM classifier.

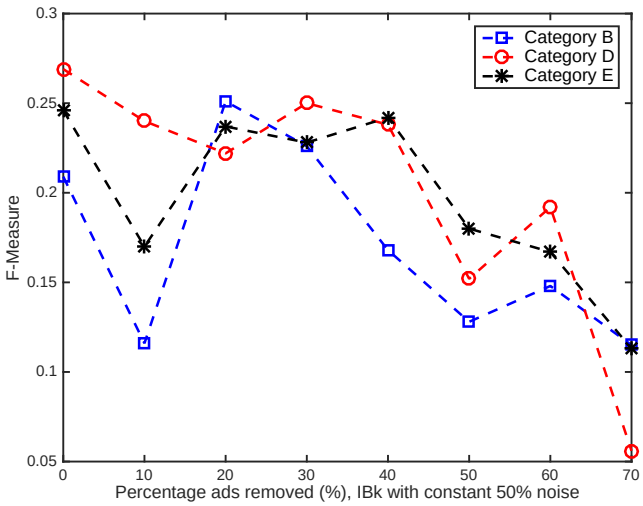

Figure 10: F-Measure response to ad deletion using NN IBk classifier.

advertising and inference of sensitive personal information using large scale data collection and analysis using machine learning algorithms.

6. CONCLUSION

In this paper we showed the possibility for an attacker to infer personal details such as income band and location by collecting ads served to users. This can be done by carefully classifying ads using topic modelling and then using machine learning algorithms such as SVM and Near Neighbour classifier to create a model and test them on the users' ad data. This is an effective technique, mainly due to the ability of advertisers such as Google in providing highly targeted ads associated with each search query. The targeting of Google searches, and hence the localised ads, strongly relies on users' location [8]; hence this sensitive piece of information can be inferred pretty accurately.

We then showed the use of obfuscation techniques as a means to increase the user privacy. We evaluated a number of methods and

ADDED ads from around England + Australia + USA (SMO, 80/20 split)							
Ads being removed are London Ads							
%added noise	%ads removed	Correctly Classified	Incorrectly Classified	F-Measure			
				Category B	Category D	Category E	Noise
0	0	34.0176	65.9824	0.359	0.335	0.326	-
10	10	33.1378	66.8622	0.335	0.327	0.367	0
20	20	24.3402	75.6598	0.246	0.294	0.228	0.177
30	30	27.8592	72.1408	0.231	0.171	0.354	0.296
40	40	30.2053	69.7947	0.165	0.154	0.086	0.482
50	50	41.6422	58.3578	0.182	0.114	0.045	0.598
60	60	56.0117	43.9883	0.161	0.061	0.043	0.72
70	70	68.1287	31.8713	0.103	0.125	0	0.813
Average F-Measure of categories			0.2005				
Average F-Measure of noise			0.44085714				

Figure 11: SVM classifier accuracy and the effect of simultaneous ad deletion and noise addition.

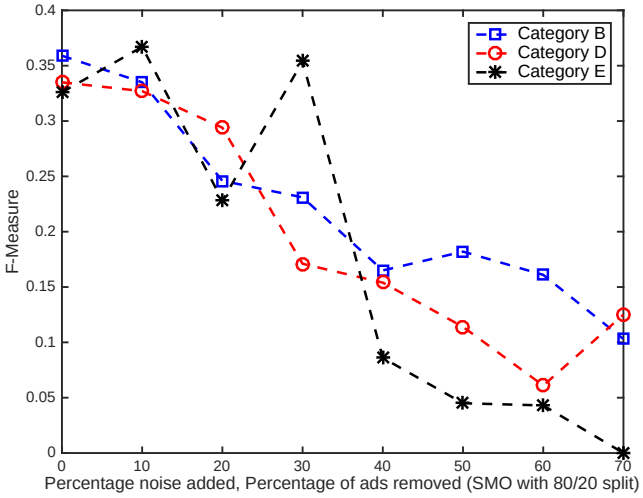

Figure 12: F-Measure response to simultaneous ad deletion and noise addition using SVM classifier.

combinations by addition of noise ads to actual ads, also known as hiding in plain sight. As the results demonstrate, obfuscation is an effective tool in reducing some of the privacy risks of targeting advertising.

Our work is partially motivated by the independent or grassroots movements such as Privacy Badger[5], Do Not Track, and other related user-driven efforts in preserving privacy. Without regulatory and government enforcements, the online advertising industry will most likely continue the trend in aggressive data collection and user tracking. The current ecosystem of the *free* Internet has been balanced in favour of the corporate sector, hence techniques such as obfuscation can act as small yet effective steps for users to protect themselves. In essence, obfuscation can provide a balance between effective targeted advertising and user privacy, without a change to the delivery mechanism of the ads and the economics of the advertising ecosystem.

7. REFERENCES

[1] AHA, D., AND KIBLER, D. Instance-based learning algorithms. *Machine Learning 6* (1991), 37–66.

[2] ANGIULI, O., BLITZSTEIN, J., AND WALDO, J. How to de-identify your data. *Queue 13*, 8 (Sept. 2015).

[3] CASTELLUCCIA, C., KAAFAR, M.-A., AND TRAN, M.-D. Betrayed by your ads!: Reconstructing user profiles from targeted ads. In *Proceedings of the 12th International*

[5]https://www.eff.org/privacybadger

ADDED ads from around England + Australia + USA (IBk, 80/20 split)							
Ads being removed are London Ads							
%added noise	%ads removed	Correctly Classified Instances	Incorrectly Classified Instances	F-Measure			
				Category B	Category D	Category E	Noise
0	0	34.8974	65.1026	0.257	0.417	0.363	-
10	10	27.566	72.434	0.256	0.3	0.311	0.048
20	20	27.2727	72.7273	0.234	0.311	0.484	0.603
30	30	24.3402	75.6598	0.143	0.248	0.291	0.259
40	40	26.9795	73.0205	0.184	0.12	0.146	0.467
50	50	39.0029	60.9971	0.174	0.171	0.159	0.591
60	60	40.4692	59.5308	0.138	0.135	0.103	0.592
70	70	53.2164	46.7836	0.167	0.136	0.08	0.707
Average F-Measure of categories			0.222				
Average F-Measure of noise			0.46671429				

Figure 13: NN IBk classifier accuracy and the effect of simultaneous ad deletion and noise addition.

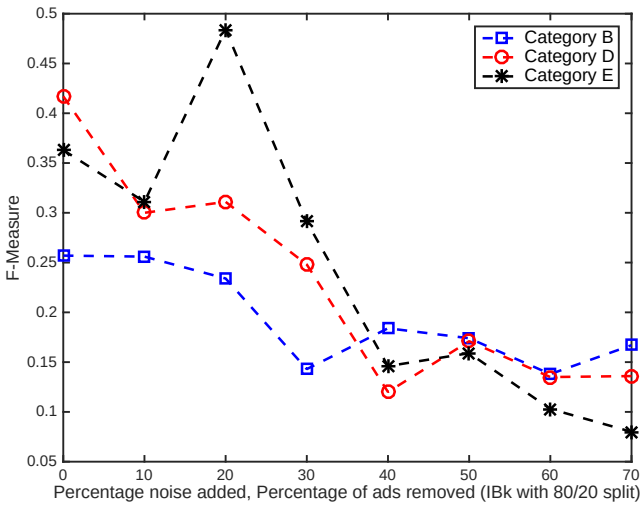

Figure 14: F-Measure response to simultaneous ad deletion and noise addition using NN IBk classifier.

Conference on Privacy Enhancing Technologies (2012), PETS'12, pp. 1–17.

[4] FALAHRASTEGAR, M., HADDADI, H., UHLIG, S., AND MORTIER, R. Anatomy of the third-party web tracking ecosystem. *CoRR abs/1409.1066* (2014).

[5] HADDADI, H. Fighting online click-fraud using bluff ads. *SIGCOMM Comput. Commun. Rev. 40*, 2 (Apr. 2010), 21–25.

[6] HADDADI, H., GUHA, S., AND FRANCIS, P. Not all adware is badware: Towards privacy-aware advertising. In *Software Services for e-Business and e-Society*. 2009.

[7] HADDADI, H., MORTIER, R., AND HAND, S. Privacy analytics. *ACM SIGCOMM Computer Communication Review 42*, 2 (Apr. 2012), 94–98.

[8] KLIMAN-SILVER, C., HANNAK, A., LAZER, D., WILSON, C., AND MISLOVE, A. Location, location, location: The impact of geolocation on web search personalization. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference* (2015), IMC '15.

[9] MOWBRAY, M., AND PEARSON, S. A client-based privacy manager for cloud computing. In *Proceedings of the Fourth International ICST Conference on COMmunication System softWAre and middlewaRE* (2009), COMSWARE '09.

[10] NISSENBAUM, H. F. Privacy as contextual integrity. *Washington Law Review 79*, 1 (Feb. 2004), 119–157.

[11] PEDDINTI, S. T., AND SAXENA, N. On the privacy of web search based on query obfuscation: A case study of trackmenot. In *Proceedings of the 10th International Conference on Privacy Enhancing Technologies* (2010), PETS'10, pp. 19–37.

[12] PLATT, J. C. Advances in kernel methods. MIT Press, Cambridge, MA, USA, 1999, ch. Fast Training of Support Vector Machines Using Sequential Minimal Optimization.

[13] VOSKUHL, J. B., AND VYAGHRAPURI, R. Inferring household income for users of a social networking system, Dec. 3 2013. US Patent 8,600,797.