ISSN 2347 - 3983
Volume 8. No. 4, April 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter76842020.pdf
https://doi.org/10.30534/ijeter/2020/76842020

Privacy Protection in Personalized Web Search using Obfuscation

Krishan Kumar[1], Dr. Bright Keswani[2], Dr. Vivek Jaglan[3]
[1]Suresh Gyan Vihar University, Jaipur, India, krishanchhillar@gmail.com
[21]Suresh Gyan Vihar University, Jaipur, India, bright.keswani@mygyanvihar.com
[3]Graphic Era Hill University, Dehradun, India, jaglanvivek@gmail.com

ABSTRACT

Web search engines are our daily needs now days to find desired information on the internet. The search engines build user profiles from the user search history. Accurate and rich user profile serve better personalized search results but pose high risk to user privacy. Not only this, sensitive information infringement and unsolicited advertisement is also a big issue. So, it is necessary to stop collection of sensitive data of user by obfuscating at client side so that user identity can be safeguarded. So, this work is carried out keeping in mind privacy protection of user and personalization. In this work, actual queries are obfuscated with a number of dummy queries which are semantically related to actual query. Numbers of dummy queries are added with a semantic distance which is controlled by the user.

Key words: Privacy Protection, Personalized web search, Obfuscation.

1. INTRODUCTION

It is very common and regular for us to find information using Web Search as it is very fast, effective and easily available option in our day to day life. Search Engines or Web Search service providers have database of pointers to web pages. These pointers are generally indexed with keywords which relate information in the web pages. To find out information in these pages, user creates a query which comprises one or more words related to user curiosity and then send it to the search engine. Then, search engine compiles the query and returns the result to user based on words in query and web pages.

These search queries are used to create user profiles i.e. search queries can reveal our interest and concerns which creates privacy threats for us. You can say, sensitive information can be inferred from our search queries like income level, health issues, political beliefs, address and so on[1], [2]. So, personalization techniques are double-edged swords in information retrieval systems. On one hand these techniques are providing user desired information but on other hand concealing personal information[3], [4] of the user which poses privacy threat to user.

Most of the users are not aware that traces related to their identity are collected while they browse internet. User activities and behavior [5], [6] is regularly monitored on internet. With a rapid growth in data analysis technique in recent past user profiling and classification is rampant in current internet and data era. Content personalization systems use data collected from users to improve user experience but personalization poses privacy risk[7].

Personalized recommendation systems serve user content and advertisement. Browsing time, search queries, number of clicks, visited web sites, cookies, web forms and web browser configuration are analyzed by adversary which helps to create accurate user profile [8]. Social networks, tagging systems, recommendation platforms and search engines collect user information. This information is used by advertisement industry and political parties as well.

Privacy is multi-dimensional complex perspective which is dependent on individual. The best way to protect privacy is to spread awareness about privacy infringement. It is not possible to measure privacy as there are no tools available, but obfuscation and obstruction of personal information will surely improve privacy. Anonymous web browsing [9],[10]and concealing search profile are very common approaches to connect with web search engine. Anonymizers use search query unlinkability to obstruct search profile creation whereas concealing search profile technique makes it tougher to re-identify anonymous user with their queries. These two techniques are complementary to each other. To conceal search queries, Private Information Retrieval (PIR) technique is also suitable. In this technique, information is retrieved from database without knowledge of database owner. Web search based PIR schemes also proposed. Cryptography based solutions can provide strong privacy but implementation and use of such protocols needs extra cost without benefits at Search Engine end and highly neglected.

In this research paper, we worked on Personalized Web Search (PWS) based on Query Obfuscation (QO)[8], [11], [12],[13],[14], [15][16]–[19][20][21][22]. The basic difference between PIR and PWS based QO is that former technique needs search engine cooperation whereas later one creates dummy search queries to protect user privacy. Dummy queries are generated by software tool will dilute actual interests of user. So concealed user interests obtained from noise added by dummy queries will improve user privacy. It has been observed that QO lower down use of search profiles and reduce the profits to perform large scale profiling besides

protecting user integrity. In this work, query search log based personalization techniques[23],[24] are used for practical systems. In these personalization techniques, past queries and clicks are generally used to create search results after creating user profile. True search intent of user is obfuscated by creating dummy queries and clicks. A noise free user profile is created to re-rank the search result at client side.
We propose framework for PWS based on QO which rudiments of system and adversary. We discussed privacy for profiles, user queries, basics of security analysis, creation of noise. We discussed eight PWS techniques with flaws and gaps. We discussed framework and issues related to it.

2. RELATED WORK

Now days, some tools are available which blocks the personal information of the user. These tools are based on heuristic approach and do not calculate privacy risk and protection level.

2.1 Privacy Extending Tool

Privacy extending tools (PETs)are used to protect user privacy[25]. We can use different approaches like basic anti-tracking technologies, proxy servers, cryptographic methods, group anonymity, and noise injection-based approaches to improvise privacy of the user.

A. Basic Anti Tracking Technologies

Personalized services work on the basis of tracking parameters like IP address and Cookies which help to map identities with their preferences. Hiding or blocking these parameters is basic anti tracking system but it may stop some internet services.

B. Private Information Retrieval

In Private Information Retrieval (PIR) is also termed a local profiling where information is retrieved from the database[26][27], [28] by user and database provider is unaware of the content of information[29]. A local profile is created instead of server side and results are re-ranked locally from the content of information downloaded which helps in improving user privacy. In some cases group profiles are also created instead of creating single user profiles which is again a way of protecting individual's privacy.

C. Proxy Servers

Proxy Server is a good mechanism to anonymize the communication. In this mechanism proxy server receive request from the user and forward it to search engine. Same channel is followed for reply and search engine is kept unaware of actual user. Main drawback of this system is proxy servers create bottlenecks. *Web MIXes*[9] and *Onion Routing*[10], [30] are examples of proxy server mechanism. Both receive messages and forward to destination which prevent tracking of user.

D. Group Anonymity

Undoubtedly group anonymity is a good way to gain privacy. *Crowds*[31], [32] and location based services (LBS) protocols [33] are using group anonymity technique of various entities to improve privacy of users. In [17] users exchange search queries before sending to search engine in such a way that user profiles does not represent interests of single user.

E. Query Obfuscation

Noise injection is a technique in which fake or dummy queries are sent with original ones to the search engine so that adversary cannot prepare precise user profile and individual's privacy can be guaranteed. *Query forgery* is an application to create dummy or fake queries. *TrackMeNot (TMN)*[34], [35] is a web browser plug-in to create fake queries to send search engine. RSS content hosted at information sources is used to create queries. *GooPIR*[36]is another tool for query obfuscation and it use words locally hosted but topics related to health and politics are difficult to obfuscate with this tool. Objective of gaining privacy at recommendation systems based on false ratings[37], forgery and suppression in personalized recommendation systems[38] are also good ideas.

2.2 Privacy Protection Oriented Tools

Some tools which protect user privacy by blocking web browser functions which release personal information are discussed here. Although level of privacy and data protection is not measured for these tools.
Adnostic[39] is browser plug-in for Mozilla Firefox which create local profile of the user on the basis of search queries and click through. The tool organizes personalized advertising without compromising user privacy.
REPRIV[40] is also web browser plug-in proposed by Microsoft. This tool work on the basis of user web browsing history and share it with third party for better search results but third-party trust is risky.
TrackMeNot[25][26] works on mechanism of query obfuscation. It is a Firefox plug-in and issues dummy queries from predefined RSS feeds at random intervals. But false queries can be identified by adversary[41] and lack of privacy measuring tools creates doubts in the mind of user.
Ghostery[https://addons.mozilla.org/en-US/firefox/]is
Firefox plug-in which detect and block trackers or object which track user actions to provide privacy protection to the user. History, images, videos, cookies etc. are not stored in private navigation mode which may block some internet services as well. Other plug-ins like NoScript, Adblock Plus, DoNotTrackMe are also used to block user information. Internet service providers, advertising agencies and social networking sites are considered as main adversaries.

2.3 Obfuscation based personalized web search

User Profiling[42][43][44]In user profiling technique, a user profile is created on the client side based on the privacy settings specified by the user. Sensitive information about the user is not revealed to the search engine and search results are attained with improved privacy. Personalization is performed without exposing the sensitive information specified by the user at the search engine side.
Plausible Deniable Search(PDS)[18][13][45]This is client side privacy protection technique in which dummy or fake queries are created using Latent Semantic Indexing. A set of cover queries is defined in offline fashion although it does not submit a query if it does not have words in predefined

dictionary. Its aim is to provide k-anonymity and emphasis that subsequent queries should be related. When user issues a query, PDS replaces the real query with a similar query from its database which is synonym of the original query.

PRAW-Privacy Model for Web Search[15] prepare a local user profile from the queries submitted and corresponding responses. PRAW issues queries which are very close to actual interests of user.

Optimized Query Forgery for Private Information Retrieval(OQF)[14] is a technique in which profile is obfuscated in such a way that it is equal to average population profile. The difference is calculated with KL divergence technique.

Noise Injection for search Privacy Protection (NIP)[46] issues queries for optimal dummy query distribution among finite number of categories. It creates a metric between observed and real profile.

Knowledge-based scheme(KBS)[23] In knowledge based technique semantically distorted queries are created to safeguard the utility of user profiles. User generated complex queries are analysed with the help of linguistic techniques and then new related queries are generated.

Embellishing Search Queries (ESQ)[47]In this technique search queries are embellished with distracting terms which have same specificity as the original term do have but with plausible different area to protect the user privacy.

Topic based Privacy protection (TPP)[48] In this technique, dummy queries are generated with a pre-trained statistical topic model. For each query TPP find out its topic and then generates queries from the selected topic. This technique does client side re-ranking as well to improve search results.

Anonymizing User Profiles (AUP)[49] This technique anonymize use profiles by grouping them on the basis of semantic relationship between query terms in different user groups and enhance user privacy.

3. USER ACTIVITY AND OBFUSCATION MECHANISM

A User issue queries to WSE to get the desired information and search engine reply with ranked list of web pages. Then the user interacts with the results by clicking, browsing, spending time on web pages or refining the search query further for better results. This activity leaves a sequence of query events which can further be defined as:

e: {u, t, q, r, c}

Where, u is the user identity which can be username, IP address, Cookie identifier or any other information which can link queries with each other if not reveal true identity of the user. t is the time at which query is generated, q is query string, r is the search result page, c is the set of pages clicked by the user. Now series of web search query events from the user can be denoted by S_U.

S_U:{e1, e2....e_n); S_F: {e1, e2..e_m}; S_O: {e1,e2,e3....e_{m+n}}

Obfuscation mechanism at user end will generate fake queries and mix with the real queries of user to protect privacy. Let us suppose S_F is set of fake query events and S_O is set of query events obtained after mixing real and fake queries. The fake queries can be created by getting information from different sources, independently, by observing behaviour of user or some other way. The fake queries can be mixed at regular intervals of time, in a burst or when user issue query.

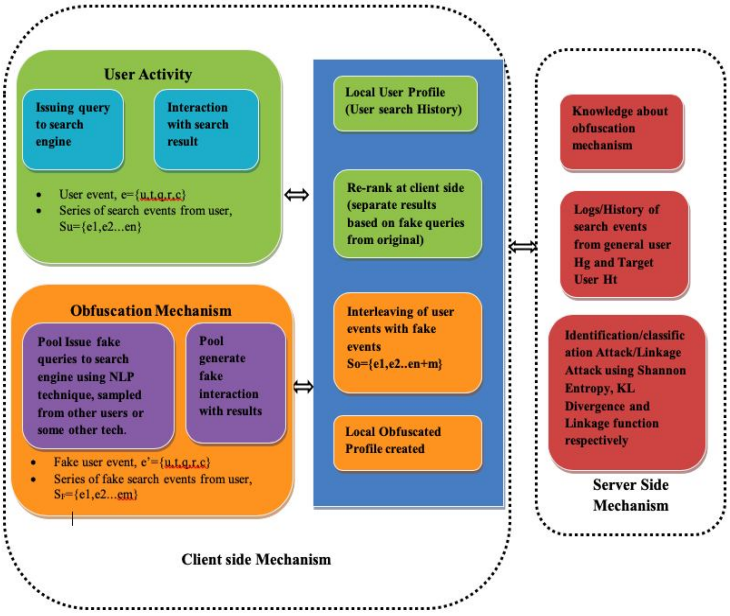

Figure 1: User activity and obfuscation mechanism

4. ADVERSARY MODELS AND PRIVACY METRICS

4.1 Adversary Model

The main target of adversary is to segregate real queries from the fake ones. Probability Mass Function (PMF) is used to create a user profile which is base for privacy metrics as well. Histogram is created with the help of a set of categories of interests. The adversary model defines properties of attacker or resolver as an entity which try to steal information of user and infringe privacy norms. Identification and classification are the main two objectives of attacker. Identification means when attacker can identify a user from a group with the help of deviation of interests from an average profile population. Classification means when attacker can classify the group of users. We assume adversary has knowledge of obfuscation mechanism like how fake queries are generated and mixing with real queries. If it does not have knowledge, then it can detect. Beside this, adversary has log history of the web search activities for a group of users from which it can create generic model and web search behaviour. We also assume, adversary have knowledge of some history of target user which can help to build more specific profile of target user and predict user's queries.

4.2 Privacy Metrics

Shannon's Entropy and Kullback-Leibler (KL) divergence [50] are used as privacy metrics. Bayesian Decision theory[51] is used to measure privacy as estimation error of adversary model. Entropy of discrete random variable X with probability P is defined as equation (1).

$$H(x) = -E \log P(x) = -\sum_x p(x) \log p(x) \quad (1)$$

Where H is Shannon's Entropy, D is KL divergence.

KL divergence also known as relative entropy D (p||q) between two probability distributions p(x) and q(x) for same variable x is defined as equation (2).

$$D(p||q) = E_p \log \frac{p(x)}{q(x)} = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad (2)$$

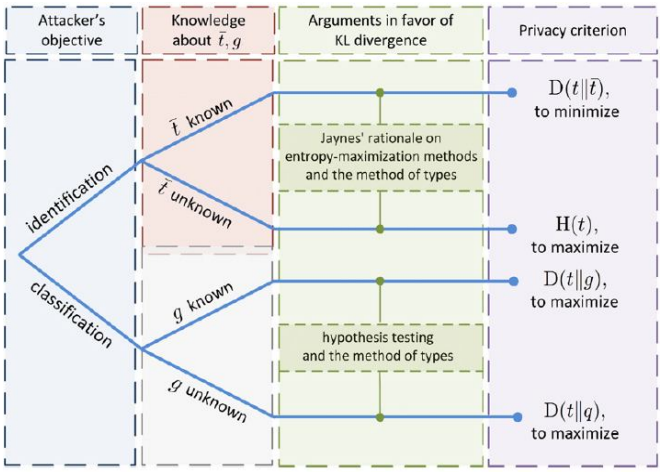

Figure 2: Adversary model and privacy metrics

Here, q is user profile, t is obfuscated profile, t̄ is population profile and g is group profile.

More the divergence between user profile (q) and obfuscated profile (t) more will be the chances of classification. More is the entropy of user profile more be the anonymity of user profile.

5. EVALUATION FRAMEWORK FOR OBFUSCATION

A set of 1000 random queries were picked from real user query log released by AOL in 2006 [2].

5.1 Profile based analysis

Profile Exposure Level (PEL) is privacy breach and can be measured [52].

PEL = $I(x,y)/H(x)$ where x and y are variable from set of real queries and dummy queries respectively. I(x,y) is Mutual Information between x and y.

$I(x,y)$ = H(x)-H(x/y) where H(x) is Entropy of x. H(x/y) is the conditional Entropy of X given Y.

$H(x) = -\sum_x P(x).\log_2 p(x)$

$H(x/y) = -\sum_{xy} p(y).\log_2 \left(p(x/y) / p(y) \right)$

5.2 Query based analysis

The difference between information contentment of actual query to dummy query is considered as semantic difference. Generating dummy queries with random concepts can harshly dilute the user profile which can affect the search results adversely. So, we created cover queries with controlled semantic distance.

As shown in Fig 3, for semantic distance =1, PEL for one dummy query is 94 percent and for four dummy queries PEL is 91 percent. For semantic distance =2, PEL for one dummy query is 81 percent and for four dummy queries PEL is 72.55 percent. For semantic distance =3, PEL for one dummy query is 65.25 percent and for four dummy queries PEL is 46.57 percent. For semantic distance =4, PEL for one dummy query is 29.37 percent and for four dummy queries PEL is -5.59 percent.

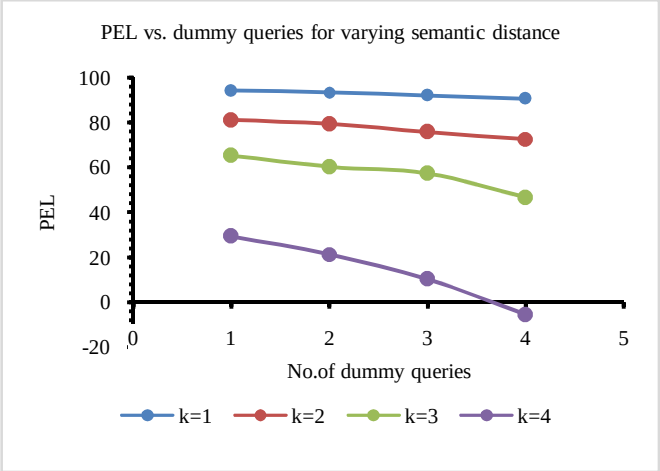

Figure 3: PEL vs. dummy queries for varying semantic distance

As shown in Fig. 4, PEL depends more on semantic distance as compare to number of fake queries. For dummy query k=1, PEL is 94 percent for semantic distance 1 and 29 percent for semantic distance four. For dummy query k=2, PEL is 93 percent for semantic distance 1 and 21.3 percent for semantic distance four.For dummy query k=3, PEL is 92.02 percent for semantic distance 1 and 10.26 percent for semantic distance four. For dummy query k=4, PEL is 90.51 percent for semantic distance 1 and -5.59 percent for semantic distance four.

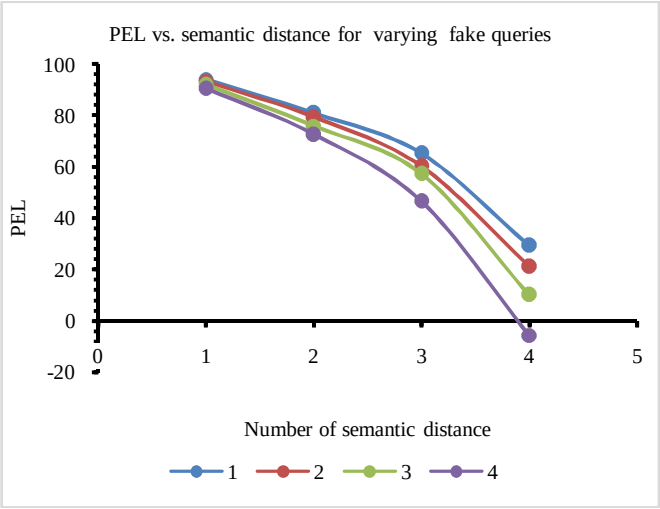

Figure 4: PEL vs. semantic distance for varying fake queries

It can be concluded from the results that PEL depends on both semantic distance and number of fake queries or value of k. A PEL of -5.59 is achieved when semantic distance is 4 and number of fake queries are 4 which indicates that conditional entropy of x given y is greater than the entropy of x or we can say more generalized fake queries are increasing the uncertainty of identifying the original queries. More semantic distance between original and dummy queries topics leads to more disassociation rather than association because of linguistic reasons.

6. KEY CHALLENGES AND ISSUES

Knowledge of adversary is big challenge which keeps track of query pattern, page landed, time spent on page through web history, cookies and page visited.

7. CONCLUSION

This technique brings back the user in control of his privacy against personalization while using search engines. Now user can set obfuscation parameters and decide amount of privacy he/she need. User can decide benefits of user profiling and personalization as well without jeopardizing personal privacy. This technique can support complex queries of user, protect user's profile semantics and obfuscate queries.
Results clearly show that if we neglect knowledge of adversary up to some extent, 100 percent privacy can be achieved with a semantic distance value of 4 and number of dummy queries 4 per query. Profile exposure level is 0%.

REFERENCES

[1] R. Jones, R. Kumar, B. Pang, and A. Tomkins, "'I know what you did last summer' - Query logs and user privacy," in *Proc. of International Conference on Information and Knowledge Management, Proceedings*, 2007, pp. 909–913. https://doi.org/10.1145/1321440.1321573

[2] M. Barbaro and T. Zeller, "A Face Is Exposed for AOL Searcher No. 4417749," *New York Times*, no. 4417749, pp. 1–3, 2006.

[3] T. P. Liang, H. J. Lai, and Y. I. C. Ku, "Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings," *J. Manag. Inf. Syst.*, vol. 23, no. 3, pp. 45–70, 2006. https://doi.org/10.2753/MIS0742-1222230303

[4] J. Teevan, S. T. Dumais, and E. Horvitz, "Personalizing search via automated analysis of interests and activities," *SIGIR 2005 - Proc. 28th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 449–456, 2005.

[5] M. Eirinaki and M. Vazirgiannis, "Web mining for Web personalization," *ACM Trans. Internet Technol.*, vol. 3, no. 1, pp. 1–27, 2003. https://doi.org/10.1145/643477.643478

[6] B. Tan, X. Shen, and C. X. Zhai, "Mining long-term search history to improve search accuracy," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 2006, pp. 718–723, 2006. https://doi.org/10.1145/1150402.1150493

[7] E. Toch, Y. Wang, and L. F. Cranor, "Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems," *User Model. User-adapt. Interact.*, vol. 22, no. 1–2, pp. 203–220, 2012.

[8] P. Eckersley, "How unique is your web browser?," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6205 LNCS, pp. 1–18, 2010.

[9] O. Berthold, H. Federrath, and S. Kopsell, "Web MIXes: A system for anonymous and unobservable internet access," in *Proc. Workshop on Design Issues in Anonymity and UnObservability, LNCS-2009, Springer-Verilag, Heidelberg 2001*, 2009, pp. 115–129.

[10] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-Generation Onion Router," *Sov. At. Energy*, vol. 46, no. 4, pp. 337–337, 1979.

[11] Y. Elovici, C. Glezer, and B. Shapira, "Enhancing customer privacy while searching for products and services on the world wide web," *Internet Res.*, vol. 15, no. 4, pp. 378–399, 2005. https://doi.org/10.1108/10662240510615164

[12] Y. Elovici, B. Shapira, and A. Maschiach, "A new privacy model for hiding group interests while accessing the Web," in *Proc. of the ACM Conference on Computer and Communications Security*, 2002, no. WORKSHOP, pp. 63–70.

[13] M. Murugesan and C. Clifton, "Providing privacy through plausibly deniable search," in *Society for Industrial and Applied Mathematics - 9th SIAM International Conference on Data Mining 2009, Proceedings in Applied Mathematics*, 2009, vol. 2, pp. 764–775.

[14] D. Rebollo-Monedero and J. Forné, "Optimized query forgery for private information retrieval," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4631–4642, 2010.

[15] B. Shapira, Y. Elovici, A. Meshiach, and T. Kuflik, "PRAW - A privacy model for the Web," *J. Am. Soc. Inf. Sci. Technol.*, vol. 56, no. 2, pp. 159–172, 2005. https://doi.org/10.1002/asi.20107

[16] W. U. Ahmad, K. W. Chang, and H. Wang, "Intent-aware query obfuscation for privacy protection in personalized web search," *41st Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR 2018*, pp. 285–294, 2018.

[17] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, "Query profile obfuscation by means of optimal query exchange between users," *IEEE Trans. Dependable Secur. Comput.*, vol. 9, no. 5, pp. 641–654, 2012.

[18] P. Mac Aonghusa and D. J. Leith, "Plausible Deniability in Web Search - From Detection to Assessment," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 4, pp. 874–887, 2018.

[19] S. Punagin and A. Arya, "A Novel Query Obfuscation Scheme with User Controlled Privacy and Personalization," *Int. J. Comput. Appl.*, vol. 158, no. 1, pp. 50–57, 2017.

[20] J. A. Estrada-Jiménez, A. F. Rodríguez, J. Parra, and J. Forné, "Evaluation of a Query-Obfuscation Mechanism for the Privacy Protection of User Profiles," *Netw. Protoc. Algorithms*, vol. 6, no. 2, p. 55, 2014.

[21] E. Balsa, C. Troncoso, and C. Diaz, "OB-PWS: Obfuscation-based private web search," in *Proceedings - IEEE Symposium on Security and Privacy*, 2012.

[22] R. Masood, M. Ikram, and C. Data, "Incognito : A Method for Obfuscating Web Data," *Web Conf.*, no. 2, pp. 267–276, 2018.

[23] D. Sánchez, J. Castellà-Roca, and A. Viejo, "Knowledge-based scheme to create privacy-preserving but semantically-related queries for web search engines," *Inf. Sci. (Ny).*, 2013.

[24] X. Shen, B. Tan, and C. X. Zhai, "Implicit user modeling for personalized search," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 824–831, 2005.

[25] J. Parra-Arnau, D. Rebollo-Monedero, and J. Forné, "Privacy-enhancing technologies and metrics in personalized information systems," *Stud. Comput. Intell.*, vol. 567, pp. 423–442, 2015. https://doi.org/10.1007/978-3-319-09885-2_23

[26] N. P. Kumar, J. K. R. Sastry, and K. R. S. Rao, "Mining negative frequent regular itemsets from data streams," *Int. J. Emerg. Trends Eng. Res.*, vol. 7, no. 8, pp. 85–98, 2019. https://doi.org/10.30534/ijeter/2019/02782019

[27] N. P. Kumar, J. K. R. Sastry, and K. R. S. Rao, "On mining Incremental Databases for Regular and Frequent Patterns," *Int. J. Emerg. Trends Eng. Res.*, vol. 2, no. 11, pp. 52–63, 2014.

[28] A. Sharma, B. Keshwani, and P. Dadheech, "Authentication Issues and Techniques in Cloud Computing Security: A Review," in *SSRN Electronic Journal*, 2019, pp. 2305–2307.

[29] R. Ostrovsky and W. E. Skeith, "A survey of single-database private information retrieval: Techniques and applications," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4450 LNCS, pp. 393–411, 2007.

[30] M. G. Reed, P. F. Syverson, and D. M. Goldschlag, "Anonymous connections and onion routing," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 482–493, 1998.

[31] M. K. Reiter and A. D. Rubin, "Crowds: Anonymity for Web Transactions," *ACM Trans. Inf. Syst. Secur.*, vol. 1, no. 1, pp. 66–92, 1998.

[32] M. K. Reiter and A. D. Rubin, "Anonymous Web Transactions with Crowds," *Commun. ACM*, vol. 42, no. 2, 1999. https://doi.org/10.1145/293411.293778

[33] C. Y. Chow, M. F. Mokbel, and X. Liu, "A peer-to-peer spatial cloaking algorithm for anonymous location-based service," in *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 2006, pp. 171–178.

[34] D. C. Howe and H. Nissenbaum, "Trackmenot: Resisting Surveillance in web search," in *Lessons from the Identity Trail : Anonymity, Privacy and Identity in a Networked Society*, 2018, pp. 418–434.

[35] V. Toubiana, L. Subramanian, and H. Nissenbaum, "TrackMeNot: Enhancing the privacy of Web Search," 2011.

[36] J. Domingo-Ferrer, A. Solanas, and J. Castellà-Roca, "H(κ)-private information retrieval from privacy-uncooperative queryable databases," *Online Inf. Rev.*, 2009.

[37] H. Polat and W. Du, "Privacy-preserving collaborative filtering using randomized perturbation techniques," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 625–628, 2003.

[38] J. Parra-Arnau, D. Rebollo-Monedero, and J. Forné, "A privacy-protecting architecture for collaborative filtering via forgery and suppression of ratings," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7122 LNCS, pp. 42–57, 2012.

[39] V. Toubiana, A. Narayanan, and D. Boneh, "Adnostic: Privacy Preserving Targeted Advertising," in *Proc. of the 17th Annual Network and Distributed System Security Symposium*, 2009.

[40] M. Fredrikson and B. Livshits, "REPRIV: Re-imagining content personalization and in-browser privacy," in *Proceedings - IEEE Symposium on Security and Privacy*, 2011.

[41] S. T. Peddinti and N. Saxena, "On the privacy of web search based on query obfuscation: A case study of trackmenot," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6205 LNCS, pp. 19–37, 2010.

[42] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "UPS: Efficient privacy protection in personalized web search," in *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2011, pp. 615–624.

[43] L. Shou, H. Bai, K. Chen, and G. Chen, "Supporting privacy protection in personalized web search," *IEEE Trans. Knowl. Data Eng.*, 2014.

[44] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-enhancing personalized web search," in *16th International World Wide Web Conference, WWW2007*, 2007.

[45] M. Murugesan and C. Clifton, "Plausibly deniable search," *Work. Secur. Knowl. …*, vol. 1, pp. 3–8, 2008.

[46] S. Ye, F. Wu, R. Pandey, and H. Chen, "Noise injection for search privacy protection," *Proc. - 12th IEEE Int. Conf. Comput. Sci. Eng. CSE 2009*, vol. 3, pp. 1–8, 2009.

[47] H. H. Pang, X. Ding, and kui Xiao, "Embellishing text search queries to protect user privacy," *Proc. VLDB Endow.*, vol. 3, no. 1, pp. 598–607, 2010.

[48] W. U. Ahmad, M. M. Rahman, and H. Wang, "Topic model based privacy protection in personalized web search," in *Proc. of the 39th International ACM SIGIR Conference on Research and Development in*

Information Retrieval, 2016, pp. 1025–1028.

[49] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing user profiles for personalized web search," in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 2010. https://doi.org/10.1145/1772690.1772886

[50] J. Parra-arnau, D. Rebollo-monedero, and J. Forné, "Measuring the privacy of user profiles in personalized information systems," *Futur. Gener. Comput. Syst.*, vol. 33, no. 2014, pp. 53–63, 2016.

[51] D. Rebollo-Monedero, J. Parra-Arnau, C. Diaz, and J. Forné, "On the measurement of privacy as an attacker's estimation error," *Int. J. Inf. Secur.*, vol. 12, no. 2, pp. 129–149, 2013.

[52] G. Navarro-Arribas, V. Torra, A. Erola, and J. Castellà-Roca, "User k-anonymity for privacy preserving data mining of query logs," *Inf. Process. Manag.*, vol. 48, no. 3, pp. 476–487, 2012. https://doi.org/10.1016/j.ipm.2011.01.004