A NOVEL APPROACH TO SECURE WEB SEARCH: PROXY-QUERY BASED QUERY OBFUSCATION

A. Lakshmipathi Rao[1], Palagati Anusha[2]

[1]Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad
[2] Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad

---***---

Abstract - Web search engines are used by people to access content on the internet. To improve the quality of the information received for individualized web searches, search engines keep query logs. User information, both sensitive and non-sensitive, is included in the query log. The query log can expose a great deal of personal information about a person and raise privacy issues if used against them. To achieve privacy-preserving online search, numerous private web search (PWS) systems have been proposed in recent years. While every PWS system claims to have special features that enable web search privacy (WSP), no study clarifies which properties related to private web searches should be taken into account when developing and implementing a PWS scheme. This article aims to accomplish two things. We introduce a new PWS strategy that employs a proxy-query-based query obfuscation method in the first section of the paper. In PWS research, proxy-query-based query obfuscation is a novel research area. Through proxy queries, it offers an IR tool for information retrieval from web search engines. The suggested system has the obvious benefit of preventing users from actually asking search engines questions in order to retrieve information. The second goal of the article is to identify PWS features and examine contemporary PWS systems in light of these attributes. We examined contemporary and suggested PWS systems based on PWS attributes. The investigation showed that only the suggested PWS design satisfies every requirement. Since existing PWS systems do not meet all PWS requirements, it has been found that they are susceptible to WSP attacks.

***Key Words*:** Private web search, Web search privacy, Proxy-query, Query Obfuscation, Information Retrieval, Cryptographic Authentication

1. INTRODUCTION

One key component of contemporary web search engines is personalized information retrieval, or PIR, or personalized web search. It provides the most pertinent data based on several settings. Web search engines employ query logs to obtain users' personal information and use the PIR approach to deduce the queries that users are typing in. Present research indicates that PIR enhances retrieval effectiveness; nevertheless, it also presents significant privacy risks for internet searches. In order to accomplish personalized web search, 75% of users are not happy with search engines that track and log their searches, according to a Pew Internet search poll conducted in the United States. To the best of our knowledge, the majority of existing methods for safeguarding users' online privacy in online contexts focus on the identifiability component of privacy. Providing encrypted data storage and release, as well as secure communication, is the primary goal of these methods. Nonetheless, linkability—an additional crucial component—is frequently disregarded. One key component of web search privacy is linkability. Web search engines employ linkability to link multiple searches to individuals, thus providing detailed information about users' specific interests. Users no longer have much influence over protecting their web search privacy (WSP). Reference has studied linkability and demonstrated through experiments that a minimally supervised classifier could considerably increase the accuracy with which a set of similar queries was linked to a set of users. To have a deeper understanding of WSP concerns, imagine a situation in where a user asks an online search engine multiple times to obtain information about depression, HIV, or pregnancy. The query log may have been hacked, or the search engine may sell this data to other businesses for targeted advertising. Sensitive information regarding the user's health status may be revealed by other organizations or attackers by using the query log's personal information. In order to ensure web search privacy, we developed a PWS technique in this article that uses proxy-query-based query obfuscation[1][2]. In PWS research, proxy-query-based query obfuscation is a new area of study. With the use of proxy queries, it offers an information retrieval (IR) tool for information retrieval from search engines. The suggested method has the obvious benefit of not requiring users to submit actual queries in order to retrieve information. When users submit proxy inquiries, the IR system3 automatically creates true and cover questions based on the proxy queries; it is unable to discern whether the user is attempting to retrieve data for a true or cover query. The client computer uses the proxy-query Posh Men Sunglasses to alter the true query before sending it to the IR system. any cover questions are created by the IR system, which also handles any cover queries related to the proxy query. The client computer receives back-rank lists of every cover query from the IR system. The client computer ignores other queries and only shows the results of the actual inquiry Depression Treatment. Making ensuring the resulting cover inquiries are plausible is the main problem in this research. The suggested plan is deemed plausible not just in relation to the user's current query but also in relation to the order in which their prior queries were submitted.

2. LITERATURE SURVEY

R. Khan et. al., explained that the increasing use of web search engines (WSEs) for searching healthcare information has resulted in a growing number of users posting personal health information online. A recent survey demonstrates that over 80% of patients use WSE to seek health information. However, WSE stores these user's queries to analyze user behavior, result ranking, personalization, targeted advertisements, and other activities. Since health-related queries contain privacy-sensitive information that may infringe user's privacy. Therefore, privacy-preserving web search techniques such as anonymizing networks, profile obfuscation, private information retrieval (PIR) protocols etc. are used to ensure the user's privacy. In this paper, we propose Privacy Exposure Measure (PEM), a technique that facilitates user to control his/her privacy exposure while using the PIR protocols. PEM assesses the similarity between the user's profile and query before posting to WSE and assists the user in avoiding privacy exposure. The experiments demonstrate 37.2% difference between users' profile created through PEM-powered-PIR protocol and other usual users' profile. Moreover, PEM offers more privacy to the user even in case of machine-learning attack[19].

S. Bashir et. al., discussed in their paper that search engines store users' queries in a query log for performing personalized information retrieval. However, query logs cause privacy concerns and reveal a lot of information about individuals if used against them. Private web search (PWS) provides a privacy-preserving information retrieval (IR) facility which allows users to retrieve information from an IR system without revealing true search queries. Current PWS techniques that are explored in the domain of web search are query obfuscation-based private web search (OB-PWS). These techniques achieve web privacy by injecting cover queries into the user profiles. However, existing OB-PWS techniques submit true queries along with cover queries and achieve query obfuscation in an isolated manner without considering the similarity between consecutive queries. In this article, we propose proxy-terms based query obfuscation technique that allows users to retrieve information from an IR system through proxy queries without submitting true queries. IR system automatically generates cover queries and true queries from the proxy queries and cannot differentiate whether the user is trying to retrieve information for the cover queries or true query. We analyzed the effectiveness of the proposed technique on test queries, and develop a similarity metric for testing the accuracy of the proposed technique. Promising results of experiments confirm the effectiveness of the proposed technique.

M. Ullah et. al., explained that Web search querying is an inevitable activity of any Internet user. The web search engine (WSE) is the easiest way to search and retrieve data from the Internet. The WSE stores the user's search queries to retrieve the personalized search result in a form of query log. A user often leaves digital traces and sensitive information in the query log. WSE is known to sell the query log to a third party to generate revenue. However, the release of the query log can compromise the security and privacy of a user. In this work, we propose a Profile Aware Obscure Logging (PaOSLo) Web search privacy-preserving protocol that mitigates the digital traces a user leaves in Web searching. PaOSLo systematically groups users based on profile similarity. The primary objective of this work is to evaluate the impact of the systematic group compared to random grouping. We first computed the similarity between the users' profiles and then clustered them using the K-mean algorithm to group the users systematically. Unlikability and indistinguishability are the two dimensions in which we have measured the privacy of a user. To compute the impact of systematic grouping on a user's privacy, we have experimented with and compared the performance of PaOSLo with modern distributed protocols like OSLo and UUP(e). Results show that, at the top degree of the ODP hierarchy, PaOSLo preserved 10% and 3% better profile privacy than the modern distributed protocols mentioned above. In addition, the PaOSLo has less profile exposure for any group size and at each degree of the ODP hierarchy[19].

3. METHODOLOGY

In this study we examined contemporary and suggested PWS systems based on PWS attributes. The investigation showed that only the suggested PWS design satisfies every requirement. Since existing PWS systems do not meet all PWS features, it has been found that they are susceptible to WSP attacks.

Existing System:

The aforementioned goal is accomplished by the prior system (ProxyTermPWS) on this research, which maps the terms of topics containing sensitive information with the terms of proxy topics. The primary flaw with ProxyTermPWS is the way it creates a proxy-term mapping between specific terms in the cover topic and the proxy. If there are multiple terms in the queries, this is not very successful. This is because finding the best proxy-term mapping for each possible combination of legal query phrases is computationally challenging[3].

Existing System Disadvantages:

Moreover, a disadvantage of the current method is that it is less effective at obscuring queries when a user asks a sequence of related questions one after the other.

Proposed System:

The query log can expose a great deal of personal information about a person and raise privacy issues if used against them. To achieve privacy-preserving online search, numerous private web search (PWS) systems have been proposed in recent years. While every PWS system claims to have special features that enable web search privacy (WSP), no study clarifies which properties related to private web searches should be taken into account when developing and implementing a PWS scheme. This article aims to accomplish two things. We introduce a new PWS strategy that employs a proxy-query-based query obfuscation method in the first section of the paper. The issues we wish to tackle in the paper include figuring out how much it costs to search for the best proxy-query mapping and suggesting a strategy for doing so. The best mapping is made available to all users by the suggested scheme as a proxyquery dictionary once it has been determined[4][5].

Proposed System Advantages:

- Increased security, efficiency, and authentication service
- Furthermore, the suggested method does not produce a comprehensive collection of cover questions for processing and recovering the real query from an ideal proxy-query mapping, nor does it provide a heuristic for finding an optimal mapping because each proxy group has a legitimate set of proxy and cover queries.

System Architecture

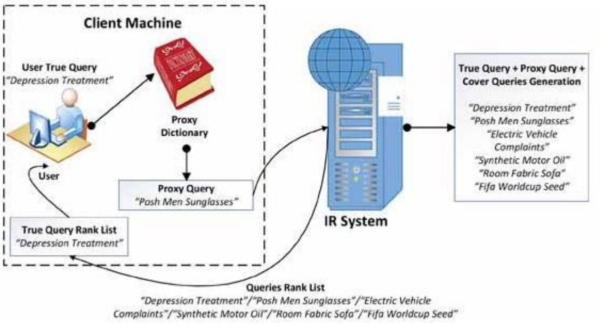


Fig -1: System Architecture

In order to protect user privacy, the client sends obfuscated queries through a proxy server as part of the client-server system architecture for the private online search utilising proxy-query based query obfuscation scheme. The search engine receives the queries via the proxy server, which then retrieves, anonymizes, and returns the results to the client.

Modules:

User Interface Design: A user's login and password are required in order to establish a connection with the server. The user can log in straight to the server if they have previously left; otherwise, they must register their information, including their email address, password, and username. In order to maintain the upload and download rates, the server will create an account for each user.

Query Processing: In this module, user requests for data are sent to the server. When a user queries the client, the proxy creates a runtime user profile based on the query parameters. A generalised user profile that satisfies the privacy standards is the step's output.

Two measures that appear to be at odds with each other—the privacy risk and the personalisation utility—are used to steer the generalisation process. These metrics are created for user profiles in which the administrator is in charge of managing all file maintenance and cloud storage.

Combining Proxy Query and User Profile (Similarity Computation): In this architecture, the generalised user profile and the user-provided query are transmitted to the server for search simultaneously. query that uses associated user preferences that are kept in a user profile to improve search results.

Online Generalisation or IR System: This approach is unable to detect a genuine inquiry. The user's computer ignores all other requests and only shows the answers to legitimate questions. The proposed method works well with customised image retrieval settings, which are widely used in commercial search engines and have been the subject of extensive research. Search engines save user queries in query logs in personalised IR to increase the efficacy of subsequent inquiries.

QOS Ranking Prediction: This model uses QOS & PREDICTION ACCURACY methods to determine whether or not to personalise search results depending on user-provided queries and privacy restrictions.

Search Personalisation: In this model, search results are sent back to the query proxy after being customised based on the user's profile. once the user is provided the results.

Admin: Using their name and password, the admin logs in to this model. Once logged in, the user has access to many features such as user profiles, file uploading with comprehensive support, data viewing for uploaded files, search log file details, and detailed analysis of user search queries. Admin keeps track of every user's information as well as certain other processes.

4. IMPLEMENTATION

Proxy-Term Based Query Obfuscation:

To search information from IR systems in secret, refer to the suggested proxy-term based query obfuscation approach (ProxyTermPWS). The suggested method uses proxy queries to retrieve information in order to achieve web search privacy. The user uses the proxy terms kept in the proxy dictionary to change the query terms for each search query. The IR system receives the modified (proxy) query. The true query is hidden by the proxy query. The IR system cannot discover the genuine query through proxy-query since it maps numerous cover queries that are stored in the proxy dictionary. Cover queries are produced by the IR system via proxy queries. All cover queries are processed, and the user receives the results. Other inquiries are ignored by the client computer, which only shows the outcome of the genuine inquiry[7-14].

A. Proxy-Query Based Query Obfuscation:

The typical method we covered above creates a proxy-term mapping between each term in the cover topic and the proxy. If there are multiple terms in the queries, this is not very successful. This is because finding the best proxy-term mapping for each possible combination of legal query phrases is computationally challenging. The suggested method creates proxy mapping using topic queries rather than term queries in order to get around this constraint. This is helpful for a series of connected searches as well as for individual queries. The suggested method establishes a proxy mapping between queries given a set of subjects and an exhaustive set of queries to obtain documents of topics. Offering the best query obfuscation possible is the aim of mapping, allowing users to get data through proxy requests. Assume that the system divides the set of exhaustive queries Q into P proxy-query sets. There are M questions in every set. Every set is guaranteed to have a minimum of one query from the topic that contains sensitive data by the system. Finding the best proxy mapping is challenging since the system tries to hide not just the user's current query but also a set of all related questions that they can submit sequentially. Given Q questions and a proxy-query collection of size M, the challenge is computationally difficult. The system must find an ideal mapping from O(QM) combinations[15-18].

5. RESULTS AND DISCUSSION:

HOME PAGE:

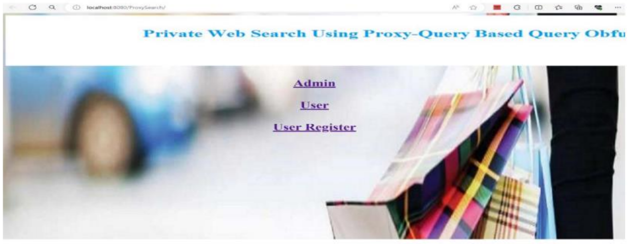
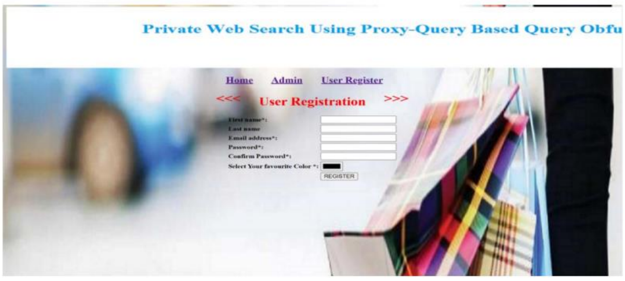

localhost:8080/ProxySearch/
Private Web Search Using Proxy-Query Based Query Obfu
Admin
User
User Register

Fig -2: Home Page

A private web search tool with an interface that is easy to use and maintains a clean look will probably have a proxy-query based query obfuscation method. It might have a search bar placed in a noticeable spot for convenience. Furthermore, tools like filters or advanced search parameters may be used for further refining searches. To reassure users, privacy guarantees and details of the obfuscation strategy could be emphasised. Finally, for further details and help, links to the terms of service, privacy policy, and support resources could be offered.

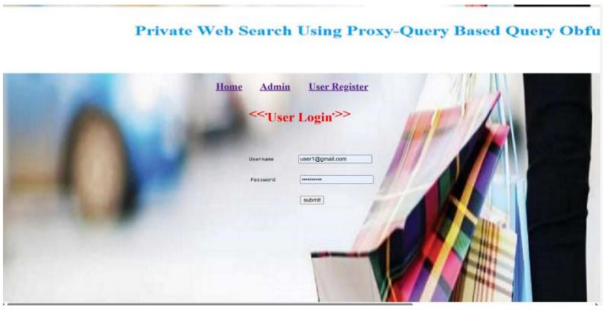
USER REGISTER:


Private Web Search Using Proxy-Query Based Query Obfu
Home
Admin
User Register
<<< User Registration >>>

First name*:
Last name
Email address*:
Password*:
Confirm Password*:
Select Your favourite Color *: ■
REGISTER

Fig -3: User Register

LOGIN PAGE:

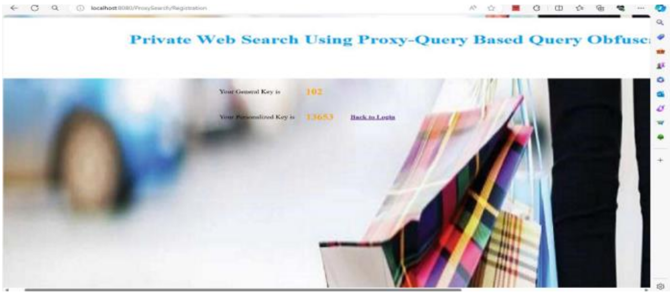

Private Web Search Using Proxy-Query Based Query Obfu

<<User Login>>

Username user1@gmail.com
Password ••••••••••

Fig -4: Login Page

CRYPTOGRAPHY: The login page may include a secret key and private key authentication as an extra degree of protection in a private online search that uses a proxy-query based query obfuscation strategy. Along with the fields for the secret key and private key, users would enter their username and password as usual. While the private key offers an additional layer of encryption to verify the user's identity, the secret key functions as an additional password. To improve privacy, these keys are kept in a secure location and are used to decrypt user data.


localhost:8080/ProxySearch/Registration

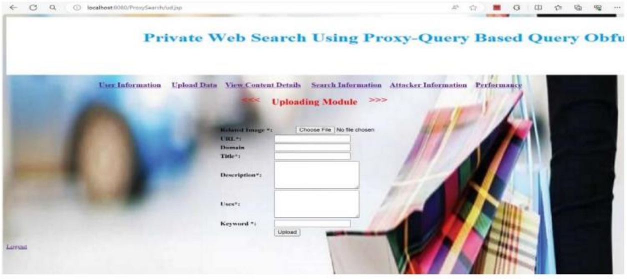
Private Web Search Using Proxy-Query Based Query Obfusc

Your General Key is 102

Your Personalized Key is 13653

Fig -5: Cryptography

UPLOADING MODULE:


localhost:8080/ProxySearch/ud.jsp

Private Web Search Using Proxy-Query Based Query Obfu

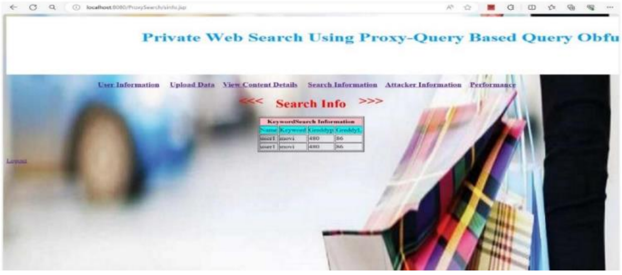
<<< Uploading Module >>>

Related Image *: Choose File No file chosen
URL*:
Domain
Title*:
Description*:
Uses*:
Keyword *:

Fig -6: Uploading Module

An uploading module might be a key element in improving user security and privacy. While not mentioned specifically, this scheme's integration of an uploading module would allow users to safely submit files or documents without immediately disclosing private information to search engines.

SEARCH INFO:


localhost:8080/ProxySearch/sinfo.jsp

Private Web Search Using Proxy-Query Based Query Obfu

<<< Search Info >>>

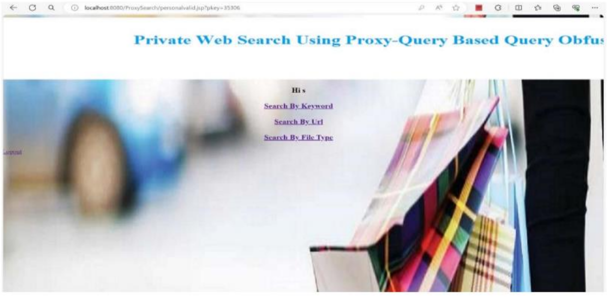
KeywordSearch Information

Name	Keyword	[illegible]	[illegible]
user1	movi	480	86
user1	movi	480	86

Fig -7: Search Info

Maintaining user privacy requires protecting the search information description. PWS methods use techniques such query obfuscation, encryption, and access limitations to try and alleviate privacy issues related to search information description. By using these strategies, the risk of privacy breaches is reduced and unauthorised access to sensitive search data is prevented.

SEARCH METHODS:


localhost:8080/ProxySearch/personalvalid.jsp?pkey=35306

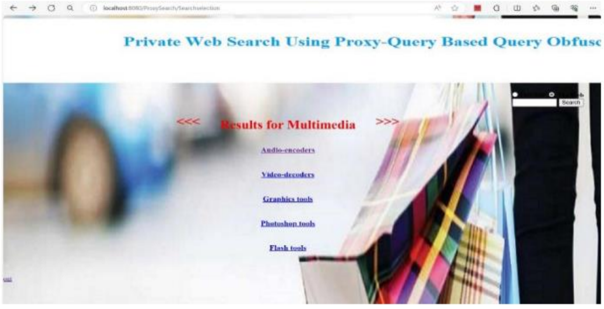
Private Web Search Using Proxy-Query Based Query Obfus

Hi s

Fig -8: Search Methods

These techniques are essential to guaranteeing that users can query data without jeopardising their private information. The classic keyword-based search technique is one popular strategy, in which user-entered terms are matched with indexed content to retrieve pertinent results. However, more advanced techniques are frequently used in the context of PWS.

RESULTS FOR MULTIMEDIA:


localhost:8080/ProxySearch/Searchselection

Private Web Search Using Proxy-Query Based Query Obfusc

<<< Results for Multimedia >>>

Fig -9: Results for Multimedia

By concealing users' search queries through a proxy service, it prevents search engines and intermediaries from tracking and profiling individuals based on their browsing habits.

6. CONCLUSION

Web search privacy is a crucial topic to take into account. When obtaining information from a search engine, web users need to protect their privacy regarding their searches. To accomplish web search privacy, we present a WSP technique in the first section of the article that uses proxy-query-based query obfuscation. In PWS, proxy-query-based query obfuscation is a new area of study. Through proxy queries, it offers an IR tool for information retrieval from search engines. The suggested method has the obvious benefit of avoiding the need for users to submit actual search engine queries in order to retrieve information. An existing proxy-term mapping between individual terms of the proxy and cover subjects is produced using the proxy-term-based query obfuscation approach. If there are multiple terms in the queries, this is not very successful. This is because finding the best proxy-term mapping for each feasible combination of query terms is computationally challenging. The suggested method creates mapping using topic queries rather than term queries in order to get around this restriction. This offers excellent efficacy for both individual inquiries and a series of connected queries.

7. FUTURE ENHANCEMENT

In order to have a deeper understanding of WSP concerns, future research should examine a situation in where a user queries a web search engine multiple times in an attempt to obtain depression-related content.

REFERENCES

[1] J. Liu, C. Liu and N. J. Belkin, "Personalization in text information retrieval: A survey", J. Assoc. Inf. Sci. Technol., vol. 71, no. 3, pp. 349-369, Mar. 2020.

[2] T.-P. Liang, H.-J. Lai and Y.-C. Ku, "Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings", J. Manage. Inf. Syst., vol. 23, no. 3, pp. 45-70, Dec. 2006.

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

[4] B. Tan, X. Shen and C. Zhai, "Mining long-term search history to improve search accuracy", Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), pp. 718-723, Aug. 2006.

[5] K. Sugiyama, K. Hatano and M. Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users", Proc. 13th Conf. World Wide Web (WWW), pp. 675- 684, May 2004.

[6] M. Eirinaki and M. Vazirgiannis, "Web mining for web personalization", ACM Trans. Internet Techn., vol. 3, no. 1, pp. 1-27, 2003.

[7] B. Mobasher, J. Srivastava and R. Cooley, "Automatic personalization based on web usage mining", Commun. ACM, vol. 43, no. 8, pp. 142-151, Aug. 2000.

[8] A. El-Ansari, A. Beni-Hssane, M. Saadi and M. El Fissaoui, "PAPIR: Privacy-aware personalized information retrieval", J. Ambient Intell. Hum. Comput., vol. 12, no. 10, pp. 9891- 9907, Oct. 2021. [9] 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-Adapted Interact, vol. 22, no. 1, pp. 203-220, Apr. 2012.

[10] R. K. Chellappa and R. G. Sin, "Personalization versus privacy: An empirical examination of the online Consumer’s dilemma", Inf. Technol. Manage., vol. 6, no. 2, pp. 181-202, Apr. 2005.

[11] R. Dingledine, "Tor and the censorship arms race: Lessons learned", Progress in Cryptology, vol. 7107, pp. 1, Dec. 2011.

[12] 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. 482-494, May 1998.

[13] N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222-233, Nov. 2014.

[14] L. Fan, L. Bonomi, L. Xiong and V. Sunderam, "Monitoring web browsing behavior with differential privacy", Proc. 23rd Int. Conf. World Wide Web (WWW), pp. 177-188, Apr. 2014.

[15] A. Inan, M. Kantarcioglu, G. Ghinita and E. Bertino, "Private record matching using differential privacy", Proc. 13th Int. Conf. Extending Database Technol. (EDBT), pp. 123-134, 2010.

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

[17] R. Jones, R. Kumar, B. Pang and A. Tomkins, "‘I know what you did last summer’: Query logs and user privacy", Proc. 16th ACM Conf. Inf. Knowl. Manag. (CIKM), pp. 909-914, 2007.

[18] D. Asonov, "Private information retrieval an overview and current trends", GI Jahrestagung, pp. 889-894.

[19] R. Khan, M. A. Islam, M. Ullah, M. Aleem and M. A. Iqbal, "Privacy exposure measure: A privacy-preserving technique for health-related web search", J. Med. Imag. Health Informat., vol. 9, no. 6, pp. 1196-1204, Aug. 2019.

BIOGRAPHIES

Mr. A. LAKSHMIPATHI RAO is working as Asst. Professor at Guru Nanak Institute of Technology, Hyderabad. He completed M.Tech CSE with Distinction from JNTUH, Kukatpally in 2010. He has 13+ years of teaching experience. He has published 02 international journals and attended 6 FDPs and 2 National level workshops. He has Member of IAENG, Membership Number: 162690. He has Qualified GATE with 88.68 percentile

Mrs. Palagati Anusha is working as Assistant Professor in the Department of Computer Science and Engineering at Guru Nanak Institute of Technology, Hyderabad with a diverse educational background. She completed her B.Tech IT from JB Women’s Engineering College, Tirupati in 2012, M.Tech CSE with Distinction from Geethanjali Institute of Technology, Nellore in 2015, and is currently pursuing a Ph.D. at Anna University, Chennai. She has 8 years of experience in teaching field. She has made significant contributions to her field with 10 articles published in reputed journals, 7 patents and she has attended 5 conferences at national and international levels.