IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

EEDS: An Efficient Multi Keyword Search Scheme over Encrypted Data on Mobile Cloud

M. Akhila[1*], K. Madhavi[2]

[1]Department of CSE, JNTUA University College of Engineering, Anantapuramu
[2]Department of CSE, JNTUA University College of Engineering, Anantapuramu

[*]*Corresponding Author: akhilayadav888@gmail.com, Tel.: +91-97057-92011*

Available online at: www.ijcseonline.org

Accepted: 08/Jun/2018, Published: 30/Jun/2018

***Abstract*—** With increasing number of websites the Web users are increased with the massive amount of data available on the internet which is provided by the Web Search Engine (WSE). The aim of the WSE is to provide the relevant search result to the user with the behavior of the user click were they performed. WSE provides the relevant result on behalf of the user frequent click based method. From this method, no assurance to the user privacy and also no securities were providing to their data. Hence users were afraid for their private information during the search has become a major barrier. There were many techniques proposed by researchers of those most of them are based on the server side where it encompasses many security concerns. For minimizing the privacy risk this paper addresses a client-side based technique with the combination of a Greedy method to prevent the user data that we applied in Knowledge mining area.

***Keywords*—** Web Search Engine, personalized search, user query logs, content search and privacy preserving.

I. INTRODUCTION

Web search engines are very important in web life. Web search engines are built for all users and not specified for any individual user. Generic internet search engines like google and yahoo cannot become aware of the distinct needs of various users if the consumer enters a fallacious key-word or ambiguous keywords and absence of customers able to express what they need are a few demanding situations faced via ordinary net search engines. To deal with this trouble we should customize those effects. As it is becoming a vital element, to provide such environments, distinct strategies and approaches have evolved. But at the same time protection of personalized internet searches has additionally won significance, in which the consumer's private or private information cannot be disclosed thru net searches.

User's hesitation to reveal their personal facts for the duration of the search has turned out to be a major difficulty on personalization technology. For instance, a system that is personalized in some advertisements according to the physical area of the person or their seek records introduces new privacy challenges which can discourage the huge adoption of personalization technologies. Personalized internet seek is proving its effectiveness however also elevating the problem of privateness and securing personal statistics. Many personalization methods had been exposed and been in practice. But it isn't always positive that those strategies will make sure their performance in distinct queries for distinctive users. The answers to PWS can usually be categorized into types, specifically click on-log-based totally methods and profile-primarily based ones. The click on-log primarily based techniques are straightforward; they honestly impose bias to clicked pages in the consumer's query history. Although this strategy has been established to perform continuously and notably nicely, It can best paintings on repeated queries from the equal consumer, that's a robust hindrance confining its applicability. In contrast, profile-based total strategies improve the search experience with complex person-interest models generated from user profiling strategies. Profile-primarily based strategies may be probably effective for nearly all styles of queries but are mentioned to be risky beneath some occasions. The two contradicting effects [4] throughout the quest manner to be taken into consideration. Improve the quest quality with the personalization application of the consumer profile and the need to cover the privateness contents existing within the personal profile to the region the privateness hazard under manipulate. This survey investigates the numerous privacy-maintaining strategies and presents an idea about the brand new efficient approach within the destiny. The major goal of this work is to assure the privateness guarantee to the consumer who is worried in the customized net search.

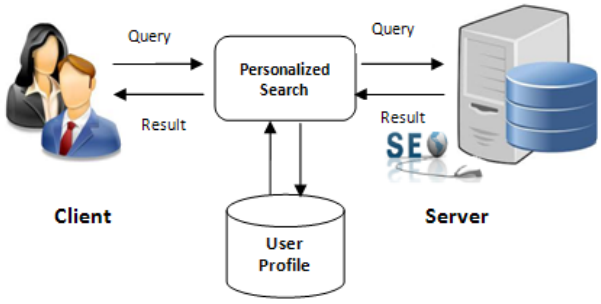


Fig 1: Personalized Search Engine Architecture

By these methods, private data without problems revealed. While many search engines like google take benefit of facts about human beings is not unusual, or concerning precise groups of people, personalized search primarily based on a consumer profile this is particular to the individual person. Research systems that customize seek effects model their users in exclusive methods. The Personalized Web Search affords a completely unique opportunity to consolidate and scrutinize the work from business labs on personalizing net seek using person logged search conduct context. It offers a totally anonymized dataset, which has anonymized consumer identification, queries primarily based at the keywords, their phrases of the query, supplying URLs, the domain of URL and the consumer clicks. This dispute and the shared dataset will enable a whole new set of researchers to observe the problem of personalizing net search experience. It decreases the chance of finding new statistics by using biasing search consequences toward what the person has already observed. By using these methods privateness of the user might be a loss because of clicking the relevant seek, often visited websites and supplying their non-public data like their call, address, etc. In this case, their privateness is probably a leak. For this privateness issue, many present works proposed capability privateness problems in which a consumer might not be aware that their seek consequences are customized for them [6, 7].

II. RELATED WORK

There are mainly two types of personalized web search they are Click-log-based and Profile-based personalized web Seek. Before 2000, there was hardly any painting aimed to offer an answer for searching on encrypted records. In 2000, D. Song, D. Wagner, and A. Perrig proposed the distinct strategies for searching operation over encrypted records [4]. These strategies for faraway looking at encrypted data have been furnished with safety proofs and have a number of critical benefits. All those techniques have been based totally on Boolean keyword seek. Boolean key-word seek isn't always suitable for cloud storage since it sends all matching files to the customers, and therefore incur a larger amount of network site visitors and a heavier publish-processing overhead for the cell gadgets. TF-IDF is a statistic which displays how essential a phrase is to a report in a collection [5]. Y. Chang and M. Mitzenmacher supplied key-word seek scheme, but it does not send returned the maximum applicable files [6]. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu proposed a one-to-one mapping OPE in order to result in Statistics Information Leak Control [3]. A. Swaminathan, Y. Mao, G. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard proposed a confidentiality-keeping rank-ordered seek [7]. This scheme displays low performances as the relevance ratings are computed on the consumer aspect, increasing its workload. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou presented a comfortable ranked key-word search over encrypted cloud facts [8]. However, of their paintings, the terms are closely related to the documents which can cause potential statistics leak. In 2015, Jian Li, Ruhui Ma, Haibing Guan proposed TEES: An Efficient Search Scheme over Encrypted Data on Mobile Cloud [9]. TEES structure was delivered to create a traffic and energy efficient encrypted keyword search tool over cell cloud storages. It offloaded the relevance rating calculation to the cloud server decreasing the burden on the cell clients. It additionally shortened the retrieval manner so that the information user can receive the most relevant documents within simplest one verbal exchange. However, TEES structure makes use of Single keyword seek consequently yielding a long way too coarse consequences. To enhance the quest end result accuracy in addition to enhance the consumer looking to revel in, it's far necessary to guide a couple of keyword searches to slim down the results. Multi-key-word is potentially the destiny mainstream encrypted seek scheme with higher looking accuracy. Table 1 compares all the preceding strategies and mentions their shortcomings

A. Click-Log-Based Method

Here, personalization is performed on the basis of clicks made by means of the user. The statistics recorded through clicks in query logs simulate user experience. The web pages regularly clicked by way of the person in beyond for a specific question is recorded in the history and rating is computed for the specific web page and based on that internet search effects are provided. This approach will carry out constantly and notably nicely while it is operating on common queries. When a by no means requested query is entered by means of the consumer; it'll no longer provide any particular seek outcomes, that's the main drawback of this approach.

B. Profile Based Personalization

The primary concept of those works is to tailor the search consequences by referring to a personal profile, implicitly or explicitly which famous a person statistics aim. Many profile

representations are to be had inside the literature to facilitate one-of-a-kind personalization techniques.

• Lists/vectors or bag of words: Earlier techniques make use of term lists/vectors or bag of words to symbolize their profile. It is the simple representation in statistics retrieval system. Here a text is represented as the bag of its phrases, brushing off grammar and even word order [3]. But it maintains the multiplicity of those phrases. In each vector the second access
could be the be counted of that word.

• Hierarchical representation: latest works construct person profiles in hierarchical structures. The cause is their more potent descriptive ability, better scalability, and higher get admission to the performance. Majority of the hierarchical representations are constructed with existing weighted topic hierarchy/graph, which includes ODP, Wikipedia, and DMOZ and so on. Using the term-frequency analysis on the personal information, the hierarchical profile may be built robotically additionally.

III. PROPOSED WORK

There are lessons of privateness safety problems for PWS in general. One class includes those works, treat privacy as the identity of a person. The other includes the ones don't forget the sensitivity of the information, in particular, the user profiles, uncovered to the PWS server.

A. Identification Of An Individual

Typical works within the literature of protective consumer identifications (class one) try and clear up the privateness problem on different tiers, which include the pseudo-identification, the institution identification, no identification, and no private information [13]. Solution to the primary degree is proved fragile. The 0.33 and fourth tiers are impractical due to the excessive fee of conversation and cryptography. So the present efforts cognizance on the second one degree.

• Online anonymity: It works based totally on user profiles via generating a group profile of k users. Using this technique, the linkage between the question and an unmarried consumer is damaged.

• Useless user profile (UUP): This protocol is proposed to shuffle queries among a set of users who trouble them. As an end result, any entity cannot profile a pure character. These works assume the life of a truthful third-birthday celebration anonymizer, which isn't always conveniently available over the Internet all of the time in big number.

• Legacy social networks: Instead of the 0.33 birthday party to provide a distorted user profile to the net search engine, here every user acts as a search business enterprise of his/her associates. They can decide to submit the question on behalf of who issued it or ahead of it to different associates.

B. Sensitivity Of Data

The answers in class do now not require 0.33-party help or collaborations among social community entries. In those answers, users handiest believe themselves and can not tolerate the exposure of their entire profiles to an anonymity server.

(i) Statistical Techniques: To examine a probabilistic model, after which use this model to generate the near-foremost partial profile. One primary difficulty of this work is that it builds the person profile as a finite set of attributes, and the probabilistic model is skilled through predefined common queries. These assumptions are impractical within the context of PWS.

(ii) Generalized Profiles: Proposed a privateness protection answer for PWS based on hierarchical profiles. Using a consumer-distinctive threshold, a generalized profile is acquired in effect as a rooted subtree of the complete profile.

C. Issues

The shortcomings of current solutions in class one are the excessive price added due to the collaboration and communication. The statistical methods construct the user profile as a finite set of attributes, and the probabilistic model is trained thru predefined common queries in class. These assumptions are impractical inside the context of PWS and the generalized profile does no longer address the question application, that's vital for the provider high-quality of PWS.

IV. METHODOLOGY

Indeed, the privacy concern is one of the major barriers in deploying serious personalized search applications, and how to attain personalized search though preserving users' privacy. Here we propose a client-side personalization which deals with the preserving privacy and envisions possible future strategies to fully protect user privacy. For privacy, we introduce our approach to digitalized multimedia content based on user profile information. For this, two main methods were developed:

Automatic creation of user profiles based on our profile generator mechanism and on the other hand recommendation

system based on the content to estimates the user interest based on our client side metadata.

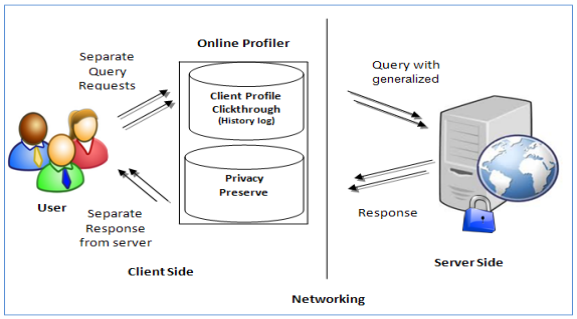


Fig 2: Proposed Architecture

Above figure shows our proposed architecture which is built in the client side mechanism and here we protect the data from the server, so only we provide a privacy to the client user.

Every query from the client user was provided by the separate requests to the server, this hides the frequent click-through logs or content based mechanism, from this user can protect the data from the server. In the equal case, our mechanism continues the net profiler approximately the person, therefore, it hides the press logs and provides a shield to the person statistics. After that, online profiler question becomes processed in the manner of generalization system, it's miles used to fulfill the unique prerequisites to deal with the user profile and its miles based totally on the preprocessing the user profiles. Our structure, not handiest the consumer's seek performance but also their history activities (e.G., viewed before) and private facts (e.G., emails, browser bookmarks) might be blanketed into the personal profile, allowing for the structure of a much richer user model for customization.

The sensitive contextual statistics is normally now not the principle thing considering that it's miles strictly stored and used on the client side. A person's personal information together with person queries and click on logs history is living at the person's private computer, and is exploited to higher assume the user' records require and provide applicable search outcomes.

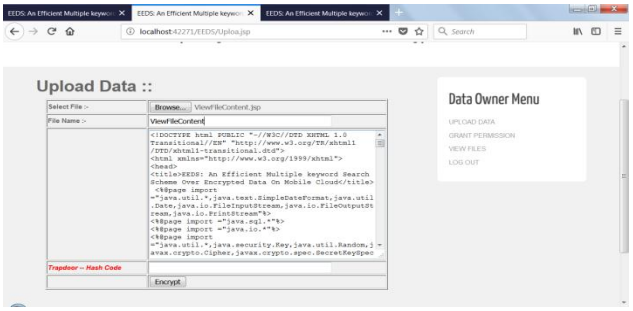
Our proposed algorithm makes use of the greedy method primarily based on the discriminating power and facts loss protection to inherit the relations. Here it makes use of the inherited technique to generalize the question.

It permits appearing the customization system to shield the information and use the User customizable Privacy-keeping Search framework addressed the privateness problems. This objective at protecting the privacy in man or woman person profiles.

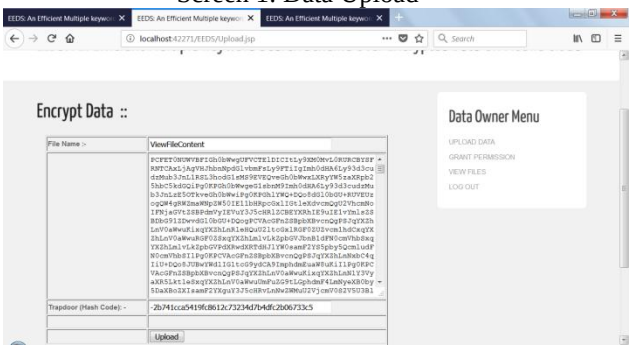
Web customers were will increase due to the availability of facts from the web browser primarily based at the seek engine. With the increasingly wide variety of user carrier engine should provide the applicable search end result based on their behavior or based on the user overall performance. Providing applicable end result to the person is based on their click on logs, query histories, bookmarks, by means of this privateness of the consumer is probably a loss. For offering relevant seek by means of using those tactics the privacy of the consumer may also lose. The maximum existing machine affords the first-rate barrier to the personal statistics all through person seek. That technique does not protect privateness issues and rising statistics loss for the person facts. For this problem, this paper proposes a customer primarily based architecture primarily based at the greedy set of rules to save you the user data and offer the relevant seek result to the user in destiny it is able to include this work inside the cell utility.

V. RESULT AND DISCUSSIONS

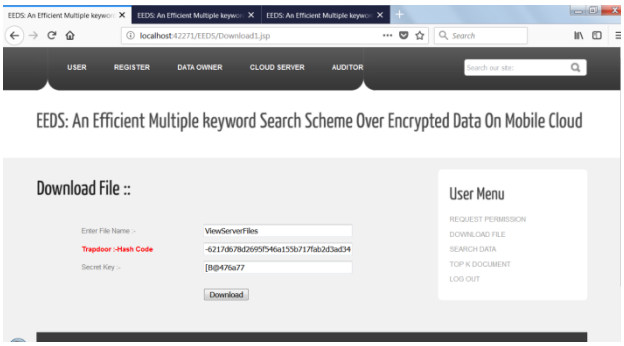
Document Summarization This graph shows the efficiency of the proposed system is better than the existing system. The result shows that retrieval rations in a millisecond. Here compare three different algorithms from which Some all text algorithm required minimum retrieval ratio.



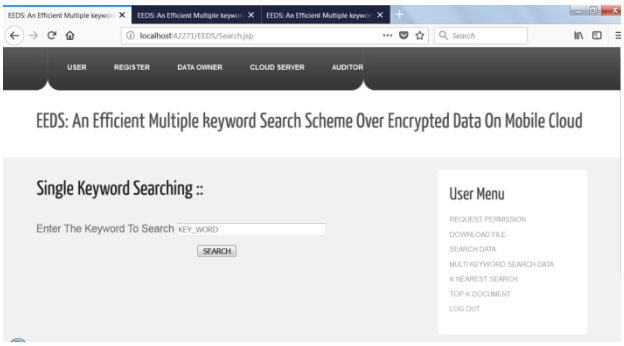
Screen 1: Data Upload



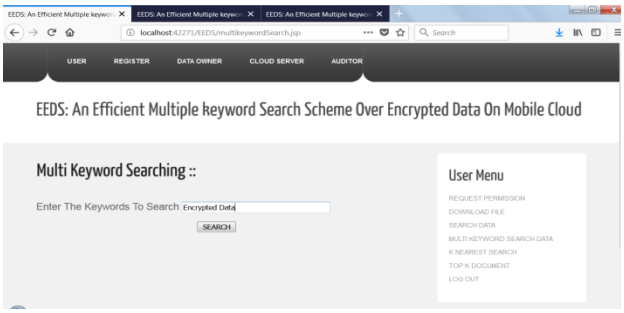
Screen 2: Encrypted Data



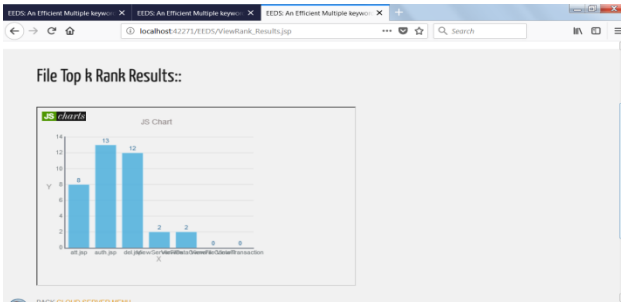
Screen 3: File Downloading



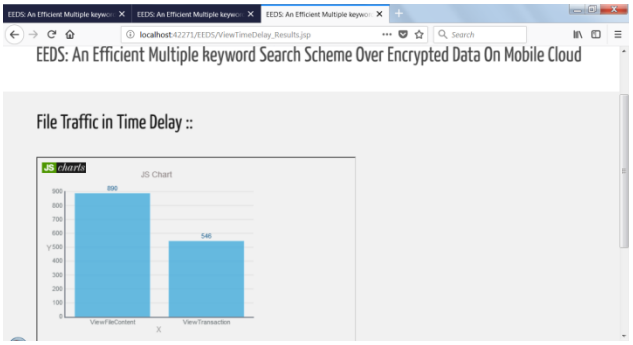
Screen 4: Single Keyword Search



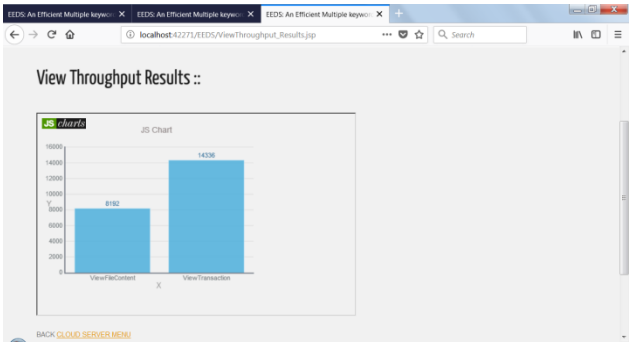
Screen 5: Multi Keyword Search



Screen 6: Top K Search Result



Screen 7: File Traffic in Time Delay



Screen 8: View Through Results

V. CONCLUSION AND FUTURE SCOPE

This paper provides a review of personalized web search and the related security concepts. The PWS techniques are developed remarkably in the last decades. A variety of techniques have emerged to increase search effectiveness and to protect privacy using multiple algorithms. Different techniques conclude that privacy protection is not dealt with well. UPS framework that's proposed to offer privacy for every consumer makes use of the net profiler to take an online selection on whether to personalize a query or not. This framework can substantially reduce the danger of assault and performs better in comparison to others. The predominant aim of this work is to guarantee the privacy guarantee to the consumer who is involved in the personalized internet search.

REFERENCES

[1] D. Huang, "Mobile cloud computing," IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter, 2011.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," the Journal of Machine Learning Research, vol. 3, 2003, pp. 993–1022.

[3] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, " Order-preserving encryption for numeric data," in Proceedings of the 2004 ACM SIGMOD international conference on Management of data. ACM, 2004, pp. 563-574.

[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44–55.

[5] A. Aizawa, " An Information-theoretic perspective of tf-IDF measures," Information Processing and Management, 2003, vol. 39, pp. 45-65.

[6] Y. Chang and M. Mitzenmacher, " Privacy-preserving keyword searches on remotely encrypted data," in Applied Cryptography and Network Security. Springer, 2005, pp. 391-421.

[7] A. Swaminathan, Y. Mao, G. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, " Confidentiality-preserving rank-ordered search," in Proceedings of the 2007 ACM workshop on Storage security and survivability. ACM 2007, pp. 7-12.

Authors Profile

M. Akhila received B.Tech in Computer Science and Engineering from Srinivasa Ramanujan Institute of Technology, Anantapur in 2016. Currently, she is pursuing M.Tech in Computer Science from JNTUA College of Engineering, Ananthapuramu, Andhra Pradesh, India. Her areas of interests include Database Systems, Cloud Computing.

K. Madhavi is Associate Professor in Computer Science and Engineering, JNTUA, Ananthapuramu, Andhra Pradesh, India. She received Ph.D. from JNTUA University. Her areas of interests include Wireless Networks, Image Processing, and Cloud Computing.

