IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Privacy Preservation in Content Based Retrieval of Large-Scale Information

Prajakta Pramod Mane[1], Suhas B. Bhagate[2]

PG Student, Department of CSE, D.K.T.E.'s Textile and Engineering Institute, Ichalkaranji, India[1]

Assistant Professor, Department of CSE, D.K.T.E.'s Textile and Engineering Institute, Ichalkaranji, India [2]

Abstract: A privacy protection framework for large-scale content-based information retrieval is proposed, which offers 2 layers of protection. To start with, strong hash values square measure used as queries to avoid uncovering distinctive content or options. Second, the client will favour to exclude bound bits in an exceedingly hash values to additional expand the anomaly for the server. Because of the reduced info, it's computationally tough for the server to grasp the customer's interest. The server has to return the hash values of each single client. The client performs a research at intervals the candidate list to find the simplest match. Since simply hash values square measure changed between client and the server, the privacy of each side is ensured. The thought is to highlight vector into items and list each piece with a sub hash value. Every sub hash price is connected with associate inverted index list. The outcomes demonstrate that the privacy upgrade somewhat enhances the retrieval performance.

Keywords: Multimedia database, indexing, content-based retrieval, data privacy.

I. INTRODUCTION

In the net era, multimedia system content is massively made and distributed. So as to expeditiously find content during a large-scale info, content-based search techniques are developed. They're utilized by content primarily based information retrieval (CBIR)[1] systems to enhance typical keyword-based techniques in applications like near-duplicate detection, automatic annotation, recommendation, etc. In such a typical scenario, a user might offer a retrieval system with a group of criteria or examples as a query; the system returns relevant info from the info as a solution. Recently, with the emergence of latest applications, a difficulty with content-based search has arisen typically the question or the info contains privacy-sensitive info. During a networked atmosphere. A privacy issue arises once an untrusted party needs to access the personal info of another party. There in case, measures ought to be taken to safeguard the corresponding info.

The main challenge is that the search must be performed while not revealing the initial question or the info. This motivates the requirement for privacy-preserving CBIR (PCBIR) systems. Privacy raised early attention in biometric systems, wherever the question and therefore the info contain biometric identifiers. Biometric systems seldom keep information within the clear, fearing thefts of such extremely valuable information[2]. Recently, the one way privacy model for CBIR was investigated. The one way privacy setting assumes that solely the user needs to stay his info secret as a result of the info is public. A number of them already integrate similarity search mechanisms, like Google pictures or Google spectacles. PCBIR is one among several aspects on privacy protection within the huge information era wherever identification becomes omnipresent. Survey on private information retrieval is done from many years[3]. Latest analysis discovers that websites are literally procedure users on the net by their system (e.g. browser) configurations. There's already some initiatives in net search privacy. The trend shows that privacy protection can become an essential a part of future content-based search systems.

II. LITERATURE REVIEW

"On privacy preserving search in large scale distributed systems: A signal processing view on searchable encryption" by S. Voloshynovskiy, F. Beekhof, O. Koval, and T. Holotyak [4] Which allows us to find similar matches in encrypted domain. The user can encrypt his query and then send it to the server. The server searches to finds a match, if match is found then the corresponding encrypted database part is sent to the user. It obtains a one single match or a list of the most similar matches in the database.

"Data-Oriented Locality Sensitive Hashing "by Wei Zhang, Ke Gao, Yong-dong Zhang, and Jin-tao Li [5] Locality Sensitive Hashing (LSH) has been projected as a scalable and high-dimensional index for approximate similarity search. Geometer LSH may be a variation of LSH and has been with success utilized in several transmission applications. However, hash functions of the essential geometer LSH project information points over arbitrarily directions that reduce accuracy, once information non-uniformly distributed. Thus a lot of hash tables are required to ensure the accuracy, and therefore a lot of memory is consumed. Since significant memory value may be an important in geometer LSH, we tend to

Copyright to IJARCCE

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

propose Data-Oriented LSH to cut back memory consumption once information area unit non-uniformly distributed. We tend to centred on the hash table construction, and therefore the query-directed ways may be applied to our index. The experiment shows that to realize a similar accuracy, our methodology uses less time and less memory compared with original geometer LSH.

"One-Way Private Media Search on Public Databases" by Giulia Fanti, Matthieu Finiasz [7] Automated media classification is changing into more and more common in areas starting from mobile location recognition to police investigation systems to automatic annotation. Whereas these tools will add nice value to the general public circle, media searches typically method non-public information; in such things, it's necessary to safeguard the interests of 1 or each parties. A lot of attention has been given to the scenario wherever each the server and therefore the client want to stay their information secret, however relatively very little work has been done on searches within which solely the purchasers information is sensitive. This case study is a scenery for a discussion on the role of signal process techniques within the style of privacy-preserving media search systems.

"A Privacy-Preserving Framework for Large-Scale Content-Based Information Retrieval" by Li Weng, Laurent Amsaleg, April Morton, and Stéphane Marchand-Maillet [8]. We propose a privacy protection framework for large-scale content-based data retrieval. It offers two layers of protection.

First, strong hash values are only used as queries to avoid revealing original content or options. Second, the client will value more highly to omit sure bits during a hash worth to more increase the anomaly for the server. Because of the reduced data, it's computationally tough for the server to grasp the client's interest. The server should come the hash values of all potential candidates to the consumer. The consumer performs an enquiry among the candidate list to search out the simplest match. Since solely hash values measure changed between the client and also the server, the privacy of each party is protected. We have a tendency to introduce the construct of tuneable privacy, wherever the privacy protection levels are often adjusted consistent with a policy. It's complete through hash-based piecewise inverted categorization. The thought is to divide a feature vector into items and index every bit with a subhash worth. Every subhash worth is related to an inverted index list. The framework has been extensively tested employing on huge information. We've got evaluated each retrieval performance and privacy-preserving performance for a selected content identification application

III.LITERATURE REVIEW TABLE

Summary of literature view is shown in Table I

TABLE I

Paper Name	Technique	Advantages	Disadvantages	Result
On privacy preserving search in large scale distributed systems: A signal processing view on searchable encryption	bit reliability techniques	It provides new best to find similar matches in encrypted domain	an unsecure server by multiple requests	find similar matches in encrypted domain.
Data-Oriented Locality Sensitive Hashing	Locality Sensitive Hashing (LSH)	LSH project information points over arbitrarily hand-picked directions, that reduces accuracy once information area unit non-uniformly distributed	We use to centred on the hash table construction	to realize a similar accuracy, our methodology uses less time and fewer memory compared with original geometer LSH.
One-Way Private Media Search on Public Databases	signal process techniques	Used for mobile location recognition to police investigation systems to automatic annotation	we have a tendency to create the case that unidirectional non-public media search is a crucial	A lot of attention has been given to the state of affairs

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

A Privacy-Preserving Framework for Large-Scale Content-Based Information Retrieval	strong hash algorithms	Since solely hash values square measure changed between the consumer and also the server, the privacy of each parties is protected	Existing privacy protection framework for large-scale content-based data retrieval	Proposed privacy protection framework for large-scale content-based data retrieval

IV. FUTURE SCOPE

A privacy issue arises once associate untrusted party needs to access the personal data of another party[9]. In this case, measures ought to be taken to safeguard the corresponding data. The most challenge is that the search has got to be performed while not revealing the initial question or the info[10]. This motivates the necessity for privacy-preserving CBIR (PCBIR) systems. In order to safeguard privacy, original content can't be used as queries. Generally even options aren't safe, because they still reveal data concerning the initial content . Rather than encoding, we tend to generate queries from original content by strong hashing.

A. System Architecture

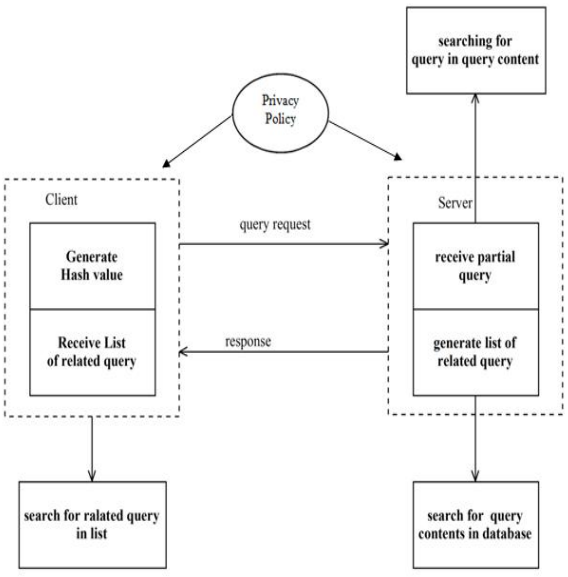

Fig. 1 System Architecture

The architecture of privacy preserving content based information retrieval works as follows:

1. A client generates a query and then sends it to server.
2. A server then generates a query list dependent on the query generated by client.
3. Then server performs a search in the list and returns the matching items.
4. The client performs a search in the list of matching items sent by the server by using his original query.

V. CONCLUSION

The framework has been enforced and extensively evaluated in numerous situations. We tend to show that the privacy level, e.g., the amount and also the diversity of candidates is tuned by the privacy policy. While accessing single database to maintain privacy of user, whole database must be downloaded [11].

Some tips area unit given on a way to select the omitted bits. We've got unquestionable retrieval performance and privacy-preserving performance for a specific content identification application. Experiment results show that question things with near-duplicates area unit probably to be at risk of majority selection. The possibility of success is appreciate the possibility that a question item has a lot of near-duplicates than different digressive things within the candidate list[12]. The results additionally show that the success rate decreases with variety the amount the quantity of omitted bits and also the number of distinct thing.

REFERENCES

[1] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," ACM Trans. Multimedia Comput., Commun., Appl., vol. 2, no. 1, pp. 1–19, Feb. 2006.
[2] J. Bringer, H. Chabanne, and A. Patey, "Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends," IEEE Signal Process. Mag., vol. 30, no. 2, pp. 42–52,Mar. 2013.
[3] W. Gasarch, "A survey on private information retrieval," in Bulletin of the EATCS, vol. 82. Rio, Greece: EATCS, 2004, pp. 72–107
[4] S. Voloshynovskiy, F. Beekhof, O. Koval, and T. Holotyak, "On privacy preserving search in large scale distributed systems: A signal processing view on searchable encryption," in Proc. Int. Workshop Signal Process. Encrypted Domain, Lausanne, Switzerland, 2009.
[5] W. Zhang, K. Gao, Y.-D. Zhang, and J.-T. Li, "Data-oriented locality sensitive hashing," in Proc. ACM Int. Conf. Multimedia, 2010,pp. 1131–1134.
[6] Privacy-preserving nearest neighbor methods: Comparing signals without revealing them March 2013
[7] G. Fanti, M. Finiasz, and K. Ramchandran, "One-way private media search on public databases: The role of signal processing," IEEE Signal Giulia Fanti, Matthieu Finiasz, and Kannan Ramchandran 2013
[8] A Privacy-Preserving Framework for Large-Scale Content-Based Information Retrieval Li Weng, Laurent Amsaleg, April Morton, and Stéphane Marchand-Maillet 2016
[9] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," J. ACM, vol. 45, no. 6, pp. 965–981, 1998.
[10] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 439–450

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

[11] R. Ostrovsky and W. E. Skeith, III, "A survey of single-database private information retrieval: Techniques and applications," in Proc. 10th Int. Conf. Pract. Theory Public-Key Cryptogr., 2007, pp. 393–411.
[12] W. Lu, A. L. Varna, and M. Wu, "Security analysis for privacy preserving search of multimedia," in Proc. 17th IEEE Int. Conf. ImageProcess. (ICIP), Sep. 2010, pp. 2093–2096.

BIOGRAPHIES

Prajakta Pramod Mane has completed B.E. In Computer Science and Engineering from Shivaji University, Kolhapur. She is currently pursuing M.E. in Computer Science and Engineering at D. K. T. E.'s Textile and Engineering Institute, Ichalkaranji, India. Her areas of interest in research includes Information Security and Data Mining.

Suhas B. Bhagate, received M.Tech (C.S.E.) degree from Walchand College of Engineering, Sangli in year 2011. He is working as an Assistant Professor in Computer Science and Engineering department at D.K.T.E. Society's Textile and Engineering Institute, Ichalkaranji (India), since July 2004. His areas of interest in research includes Visual Cryptography, Data structures, Algorithms, Big Data.