ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. IV, Issue IV, April 2018

A DISTRIBUTED SYSTEM FOR PRACTICAL PIR IN PRIVATE CLOUD

Dr.R.Sumathi[1],V.Elavarasi[2], G.Monishree[3],B.Praveena[4]

[1]Professor, Saranathan College of Engineering
[2,3,4]Student, Saranathan College of Engineering

Abstract—Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally.

Computational Private Information Retrieval (cPIR) protocols allow a client to retrieve one bit from a database, without the server inferring any information about the queried bit. These protocols are too costly in practice and the bandwidth of the database is high so it will take more time to retrieve the database

A pCloud is presented, a distributed system that constitutes the first attempt towards practical cPIR. The approach assumes a disk-based architecture that retrieves one page with a single query. Using a striping technique, we distribute the database to a number of cooperative peers, and leverage their computational resources to process cPIR queries in parallel. PCloud reduces considerably the query response time compared to the traditional client/server model, and has a very low communication overhead as well as increase a no of peers to speed up the process. The real estate business is used as an application in order to explain our pcloud concepts.

Key terms: Computational Private Information Retrieval (cPIR), pCloud, striping technique.

1. Introduction

Cloud Computing is an internet based computing whereby shared servers provide resources software and data to computers and other devices on demand. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. Example – Gmail, Yahoo, Email

A Private cloud is setup within an organization's internal enterprise datacenter. It is easier to align with security, compliance, regulatory requirements and provides more enterprise control over deployment and use. In the private cloud, scalable resources and virtual applications provided by the cloud vendor are pooled together and available for cloud users to share and use. It differs from the public cloud in that all the cloud resources and applications are managed by the organization itself, similar to intranet functionality. Utilization on the private cloud can be much more secure than that of the public cloud because of its specified internal exposure. Only the organization and designated stake holders may have access to operate on a specific private cloud.

Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third party. Computational Private Information Retrieval (cPIR) protocols allow a client to retrieve one bit from a database, without the server inferring any information about the queried bit. These protocols are too costly in practice and bandwidth of the database is high so it will take more time to retrieve from the database.

Pcloud, a distributed system that constitutes the first attempt towards a practical cPIR. The approach assumes a disk-based architecture that retrieves one page with a single query. Using a striping technique, is distributed the database to a number of cooperative peers and leverage their computational resources to process cPIR queries in parallel. Pcloud reduces time compared to the traditional client/server model, and has a very low communication cost. This system is implemented in a real estate application which n stores and retrieves the information from the private cloud.

2. Literature survey

Benny Chor, Oded Goldreich, Eyal Kushilevitz & Madhu Sudan[1] examined that publicly accessible databases are an indispensable resource for retrieving up to date information. But they also pose a significant risk to the privacy of the user, since a curious database operator can follow the user's queries and infer what the user is after. Indeed, in cases where the users' intentions are to be kept secret, users are often cautious about accessing the database. It can be shown that when accessing a single database, to completely guarantee the privacy of the user, the whole database should be down-loaded; namely n bits should be communicated (where n is the number of bits in the database).

While replicating the database, more e□cient solutions to the private retrieval problem can be obtained. The schemes that enable a user to access k replicated copies of a database ($k \geq 2$) and privately retrieve information stored in the database. This means that each individual server (holding a replicated copy of the database) gets no information on the identity of the item retrieved by the user. This schemes use the replication to gain substantial saving. In particular, two-server scheme with communication complexity O (n1/3) is presented.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. IV, Issue IV, April 2018

The question of proving lower bounds on private information retrieval schemes remains one of the most intriguing open problems. The only obvious lower bound is a log n bit which holds for any number of servers (this follows from communication complexity considerations without using any privacy argument). Thus a simple lower bound on the communication complexity in single-server PIR schemes is proved.

Craig Gentry & Zulfikar Ramzan[2] proposed a single-database private information retrieval (PIR) scheme with communication complexity O (k+d) is presented, where k ≥ log n is a security parameter that depends on the database size n and d is the bit-length of the retrieved database block. This communication complexity is better asymptotically than previous single-database PIR schemes. The scheme also gives improved performance for practical parameter settings whether the user is retrieving a single bit or very large blocks. For large blocks, this scheme achieves a constant "rate" (e.g., 0.2), even when the user-side communication is very low (e.g., two 1024-bit numbers). This scheme and security analysis is presented using general groups with hidden smooth subgroups; the scheme can be instantiated using composite moduli, in which case the security of this scheme is based on a simple variant of the "Φ-hiding" assumption by Cachin, Micali and Stadler.

A single-database computational block retrieval schemes based on the decision subgroup problem with communication complexity O (k+d), where d is the size of the block to be retrieved and k is the security parameter. There is only an additive communication overhead of the security parameter k. Indeed, this scheme has better asymptotic performance compared to previous schemes. It is an open problem to construct an instantiation of this scheme that achieves rate arbitrarily close to 1, while circumventing Coppersmith's attack. Clearly, based on analysis of the Decision Subgroup Problem in the generic group model, any attack that prevents the scheme from achieving rate close to 1 must exploit the encoding of the elements.

Gabriel Ghinita1, Panos Kalnis, Ali Khoshgozaran & Cyrus Shahabi, Kian-Lee Tan[3] modelled mobile devices equipped with positioning capabilities (e.g., GPS) that ask location-dependent queries to Location Based Services (LBS). To protect privacy, the user location must not be disclosed. Existing solutions utilize a trusted anonymizer between the users and the LBS. This approach has several drawbacks: (i) All users must trust the third party anonymizer, which is a single point of attack. (ii) A large number of cooperating, trustworthy users is needed. (iii) Privacy is guaranteed only for a single snapshot of user locations; users are not protected against correlation attacks (e.g., history of user movement).

A novel framework to support private location dependent queries, based on the theoretical work on Private Information Retrieval (PIR) is proposed. This framework does not require a trusted third party, since privacy is achieved via cryptographic techniques. Compared to existing work, this approach achieves stronger privacy for snapshots of user locations; moreover, it is the first to provide provable privacy guarantees against correlation attacks. This framework is used to implement approximate and exact algorithms for nearest-neighbor search. Query execution is optimized by employing data mining techniques, which identify redundant computations. Contrary to common belief, the experimental results suggest that PIR approaches incur reasonable overhead and are applicable in practice.

In this, the Private Information Retrieval theory is employed to guarantee privacy in location-dependent queries. This is the first work to provide a practical PIR implementation with optimizations that achieve reasonable communication and CPU cost. Compared to previous work, our architecture is simpler, more secure (i.e., does not require an anonymizer or collaborating trustworthy users), and is the first one to protect against correlation attacks.

Currently, there is a working on sophisticated heuristics to generate better optimized execution plans, in order to reduce further the CPU cost. In the future, there is a plan to investigate the extension of this framework to di□erent types of queries, such as spatial joins.

Ali Khoshgozaran & Cyrus Shahabi[4] proposed The ubiquity of smartphones and other location-aware handheld devices has resulted in a dramatic increase in popularity of location based services (LBS) tailored to user locations. The comfort of LBS comes with a privacy cost. Various distressing privacy violations caused by sharing sensitive location information with potentially malicious services have highlighted the importance of location privacy research aiming to protect user privacy while interacting with LBS.

The anonymity and cloaking-based approaches proposed to address this problem cannot provide stringent privacy guarantees without incurring costly computation and communication overhead. Furthermore, they mostly require a trusted intermediate anonymizer to protect a user's location information during query processing. A set of fundamental approaches is reviewed based on private information retrieval to process range and k-nearest neighbour queries, the elemental queries used in many Location Based Services, with significantly stronger privacy guarantees as opposed to cloaking or anonymity approaches.

The PIR-based approaches to location privacy presented an open door to a novel way of protecting user privacy. Full privacy guarantees of these approaches come at the cost of computationally intensive query processing. Therefore, reducing the costs associated with PIR operations can greatly increase the popularity of these approaches. For the approximate nearest neighbour queries utilizing the excessive object information returned to a user to guarantee exact results is one promising research direction. As for the hardware-based approaches, employing more e□cient shu□ing techniques and moving as much non-secure processing as possible away from the SC can result in significant improvements.

The privacy/e□ciency trades of various location privacy approaches discussed. While cloaking and transformation-based approaches enable e□cient spatial query processing they su□er from various privacy implications.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. IV, Issue IV, April 2018

Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek & Hari Balakrishnan[5] solved A fundamental problem that confronts peer-to-peer applications is to efficiently locates the node that stores a particular data item. Chord, a distributed lookup protocol that addresses this problem. Chord provides support for just one operation: given a key, it maps the key onto a node. Data location can be easily implemented on top of Chord by associating a key with each data item, and storing the key/data item pair at the node to which the key maps. Chord adapts efficiently as nodes join and leave the system, and can answer queries even if the system is continuously changing. Results from theoretical analysis, simulations, and experiments show that Chord is scalable, with communication cost and the state maintained by each node scaling logarithmically with the number of Chord nodes.

Many distributed peer-to-peer applications need to determine the node that stores a data item. The Chord protocol solves this challenging problem in decentralized manner. It offers a powerful primitive: given a key, it determines the node responsible for storing the key's value, and does so efficiently. In the steady state, in an N-node network, each node maintains routing information for only about O (log N) other nodes, and resolves all lookups via O (log N) messages to other nodes. Updates to the routing information for nodes leaving and joining require O ($\log^2$ N) only messages.

Attractive features of Chord include its simplicity, provable correctness, and provable performance even in the face of concurrent node arrivals and departures. It continues to function correctly, albeit at degraded performance, when a node's information is only partially correct. The theoretical analysis, simulations, and experimental results confirm that Chord scales well with the number of nodes, recovers from large numbers of simultaneous node failures and joins, and answers most lookups correctly even during recovery.

It is believed that Chord will be a valuable component for peer to-peer, large-scale distributed applications such as cooperative file sharing, time-shared available storage systems and distributed indices for document and service discovery, and large-scale distributed computing platforms.

3. System Architecture

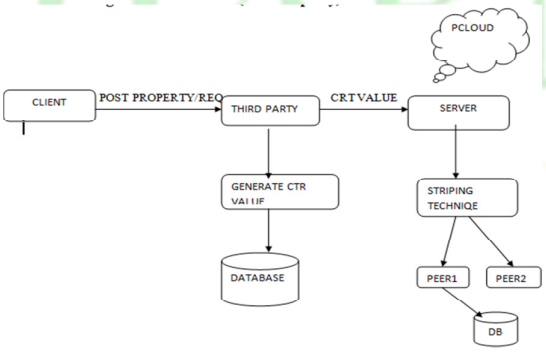

Fig1: Storing the data in PCloud

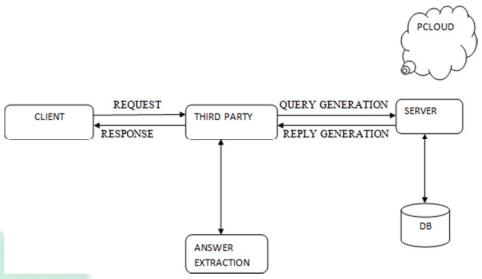

Fig2 : Retrieving the data from PCloud

4. Proposed System

In order to avoid all the problems in the existing system, a private cloud i.e. pcloud which will reduce the query response time when compare with traditional client/server is introduced. In the proposed system we use computational Private Information Retrieval (cPIR) protocols allow a client to retrieve one bit from a database, without the server interfering any information about the queried bit.

In this system the server does not know the original client rather every process is done by the intermediate (third party). When everything is accessed through the third party, the server would infer that the intermediate is a original client so the actual client will have more security and privacy. When the client request the data, the server will give the exact response for that query and not the whole database and so the response time is get reduced.

In this proposed system Chinese Remainder Theorem is used to produce a random number for the particular file. This CRT is used to rename the original file name. It will provide security to the client files. In order to store the data in the different peers the striping technique is used. The stripping technique is useful when a processing device request access to data more quickly than a single storage device can provide.
For each and every data, it randomly selects one peer and stores the data in that peer.

The advantages of proposed system: (1)It will hide the user location using PIR protocol (2)It will reduce the query response time (3)It reduces the cost by retrieving the exact information not a database (4)It provides security to the user data in the cloud.
+

5. Experimental Outputs

The outputs taken in various stages during the implementation are presented below.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE
International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. IV, Issue IV, April 2018

Fig3 : Posting the property from client login

Fig4 : Post property stored in the file name 4240.xml

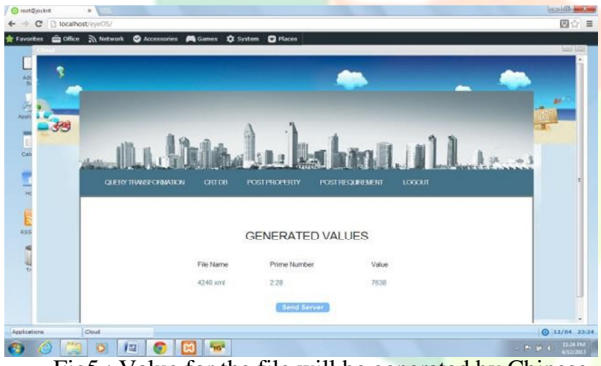

Fig5 : Value for the file will be generated by Chinese Remainder Theorem and stored in server

Fig6 : Searching the property using original file name

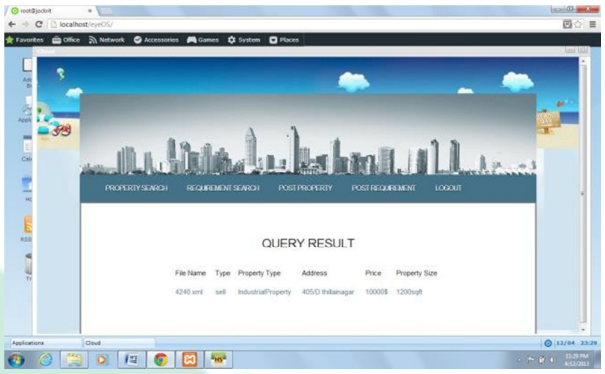

Fig7 : Property details fetched

Fig8 : 4240.xml file transformed into 7638.xml

6. Conclusion

Private information retrieval (PIR) is an important field with several practical applications. The proposed pCloud solution embeds a state-of-the-art PIR protocol in a distributed environment, by utilizing a novel striping technique.

pCloud can retrieve arbitrarily large blocks of information with a single query. A comprehensive solution that includes a data placement policy, result retrieval and authentication mechanisms are presented. This system enables collaboration and communication among users in the cloud environment. Specifically, compared to the traditional client/server architecture, pCloud drops the query response time by orders of magnitude, and its performance improves linearly with the number of peers.

7. References

[1] Benny Chor, Oded Goldreich, Eyal Kushilevitz & Madhu Sudan. Private Information Retrieval, April 21, 1998

[2] Craig Gentry & Zulfikar Ramzan. Single-Database Private Information Retrieval with Constant Communication Rate, DoCoMo Communications Laboratories USA, Inc.@docomolabs-usa.com

[3] Gabriel Ghinita1, Panos Kalnis, Ali Khoshgozaran & Cyrus Shahabi, Kian-Lee Tan. Private Queries in Location

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. IV, Issue IV, April 2018

Based Services: Anonymizers are not necessary, Dept. of Computer Science National University of Singapore@comp.nus.edu.sg.

[4] Ali Khoshgozaran & Cyrus Shahabi. Private Information Retrieval Techniques for Enabling Location Privacy in Location-Based Services, University of Southern California Department of Computer Science Information Laboratory (InfoLab) Los Angeles, CA 90089-0781.

[5] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek and Hari Balakrishnan. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications, MIT Laboratory for Computer Science chord@lcs.mit.edu

[6] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, & K.-L.Tan. Private queries in location based services: Anonymizers are not necessary, In SIGMOD, 2008.

[7] A. Khoshgozaran, H. Shirani-Mehr, and C. Shahabi. SPIRAL: A scalable private information retrieval approach to location privacy. In PALMS, 2008

[8] C. A. Melchor and P. Gaborit. A fast private information retrieval protocol. In ISIT, 2008.