Enhanced Performance of PIR in Outsourcing Database Services -A Real-life E-Commerce Application

[1]Mrs. Durga Puja, [2]Dr. Raghav Mehra
[1]Research Scholar, [2]Professor
[1,2]Computer Science & Engineering Department
[1,2]Bhagwant University, Ajmer, Rajasthan, India

Abstract: In this paper, E-commerce is normally measured to be the sales phase of electronic business. The main purpose of this solution is to present a scheme to employ database outsourcing protocol like Private Information Retrieval to develop an e-commerce application. When a customer buys some products online, then he/she has to fill the necessary details. These details may include his/her personal information along with credit card details. For a trustworthy system, it is important to keep the customer's information secret from thethird party service provider. The use of PIR protocol within e-commerce applications would make it possible for a customer to keep his/her preferences private from others including the service provider. In this model, we propose a new design for e-commerce applications using PIR to address the issue of keeping a customer's private information secret from the service provider. This paper work addresses the various security issues that arise in the Outsourced Database. These range from providing data privacy, user privacy, data confidentiality, and key management, to enabling anun-trusted server to run queries over encrypted data. We also focus on the performance aspects of our solutions.

Keywords-E- commerce application , Private Information Retrieval, User Privacy, Data Privacy.

Introduction: E-commerce is mostly used via the Internet, but before the Internet was available, a form of electronic transactions occurred over Electronic Data Interchange (EDI). Businesses and customers used EDI by setting up a data link specifically reserved for commerce between them. For example, businesses in Europe can ship items to customers in America, and vice versa. Just like any other technology, a user needs to know how e-commerce works and how to navigate through it. One of the basic concepts is knowing what a "shopping cart" is and how to use it. A "shopping cart" is a software that allows buyers to have a virtual chart, just like you would at a physical store, to collect multiple items and be able to buy all the items at once when you "check out".

E-commerce if we consider E-Commerce as the networking of business communities and digitization of business information. EDI, which stretched from financial transactions to other

business deal processing and enlarged the participating companies from economic institutions to manufacturers, retailers, services, and so on. One reason for the rapid growth of the technology was the improvement of networks, protocols, software, and specifications.

Objective

This research work focuses on security issues in databases that are outsourced and/or are shared among many users. In such databases, catering to security requirements listed above are difficult, error prone and performance hindering. The research work proposes solutions to address security requirements for various scenarios encountered in outsourced databases. The broad activities carried out for this areto study various approaches used for Outsourcing Database as Service (ODBS) Model and too research various possible scenarios in such ODBS Models . The experiments conducted to demonstrate the applicability of the proposed PIR technique to a practical real-time application

Literature Review : Carlos Melchor and Philippe Gaborit [11]: presented a lattice-based PIR scheme, in which the computational cost is a few thousand bit-operations per bit in the database. They claimed that the protocol improved computational performance by two orders of magnitude when compared to existing approaches. Also, they have shown that practical usability of PIR scheme is not as dependent on communication performance as the literature suggests.

Protection of Indexes in Outsourced Databases: The indexing scheme proposed in [60] suggests encrypting the whole database row and assigning a set identifier to each value in this row. When searching a specific value, its set identifier is calculated and then passed to the server, who, in turn, returns to the client a collection of all rows with values assigned to the same set. Finally, the client searches the specific value in the returned collection and retrieves the desired rows. In this scheme, equal values are always assigned to the same set, so some information is revealed when statistical attacks are applied, as stated in[27]. The indexing scheme in [27] suggests building a B-Tree index over the table plaintext values and then encrypting the table at the row level and the B-Tree at the node level. The main advantage of this approach is that the B-Tree content is not visible to the untrusted database server. However, only the client can now perform the B-Tree traversal, by executing a sequence of queries. Each query retrieves a node located at a deeper level of the B-Tree. The indexing scheme provided in [67] is based on constructing the index on the plaintext values and encrypting each page of the index separately. Whenever a specific page of the index is needed for processing a query, it is loaded into the memory and decrypted. Since the uniform encryption of all pages is likely to provide many cipher breaking clues.

The indexing scheme provided in [9] proposes encrypting each index page using a different key depending on the page number. However, these schemes, which are implemented at the level of the operating system, are not satisfactory, since in most cases it is not possible to modify the operating system implementation. Moreover, in these schemes, it is not possible to encrypt different portions of the database using different keys.

Use of PIR in E-commerce Applications: We extend Goldberg's multi-server information-theoretic private information retrieval (PIR) with a suite of protocols for privacy-preserving e-commerce. Our first protocol adds bear for single-pay tiered pricing, wherein users acquire database records without illuminating the indices or prices of those records. Tiered pricing lets the seller set prices based on each user's status within the system; Next, we show how to do some basic secretarial to execute a novel top-K replication approach that enables the servers to assemble bestsellers lists, which facilitate earlier retrieval for these most admired records. No other priced PIR or oblivious transfer protocol supports tiered pricing, access control lists, multiple payees, or top-K replication, whereas ours supports all of these features while preserving PIR's sub linear communication complexity.

Overview of PIR: Private Information Retrieval (PIR) is privately retrieving the ith bit out of an N-bit string stored at the service provider. Private means that the server does not know about i, that is, the server does not learn which bit the client is interested.

Need for PIR: E-commerce mainly focuses on buying and selling of products on the Internet and deals with different sensitive activities. In this case, information such as credit card details, personal information, bank account details, etc. of the customer is very important and confidential. If this information is leaked, there are chances that it can be misused for someone's personal use. With the use of PIR protocol, the customer can retrieve a particular record or product from the database in such a way that the service provider or any attacker will not be able to identify which item the user wants. Hence, the protocol helps to achieve security issues like user privacy, data privacy, confidentiality, etc. in e-commerce applications.

Entire Database Download: It is the traditional approach of achieving the PIR protocol. In this approach, the whole database from the service provider is downloaded to the client side. After that the client chooses the desired records from this downloaded dataset.

Single-database PIR: In this section we consider single-database PIR, which deals with the schemes that allow a user to repossess privately an constituent of a non-replicated database. In these schemes [22].In a PIR protocol, a user wants to retrieve an element of index *i* from a database, uses a PIR query generation algorithm with input *i* and sends the

resulting query to the database (see Figure 1). The database combines its elements to the query using a reply generation algorithm and obtains a result which is sent back to the user. Finally, the user decodes the answer through a reply decoding algorithm. The protocol is said to be correct if the decoding results in the i^{th} element of the database, and private if the database is unable to learn anything about *i* from the query.

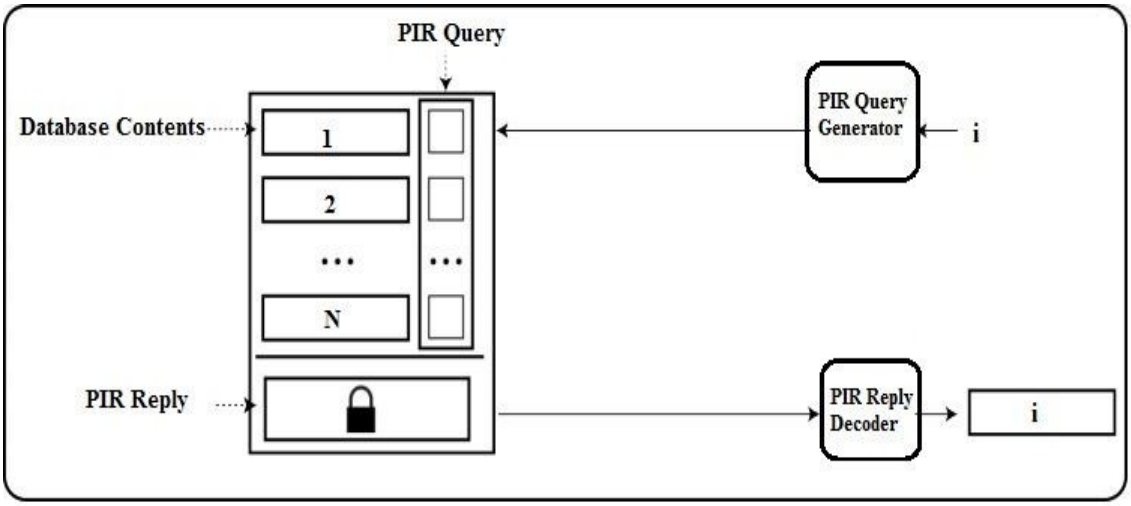

Figure 1: PIR scheme

The entire process involves following steps:

i)Query Generation-A query generation algorithm is used by the PIR client to generate a PIR query with input *i* given by the user as described above.

ii)Reply Generation-The PIR server performs complex concatenation of the query with contents of the data base to generate there ply.

iii)Reply Encoding-The reply generated is encoded and sent back to the client.

iv)Reply Decoding-The encoded reply is decoded and noise is removed as described above to get the intended element. Further sections deal with the overall proposed system design and how this PIR scheme can be integrated with an e-commerce application to ensure user privacy.

Security Issues addressed in the Proposed System: The merchant can still predict the item from the amount that the customer pays for the item purchased. The solution for this is to employ a prepaid system where the customer recharges his account with some fixed denomination and balance details are preserved in the user's own private database. Thus, each transaction results in deduction from the amount already paid to the merchant. To further strengthen the PIR scheme, the data sent to the customer can be symmetrically encrypted with customer's private key using some strong encryption scheme like AES or

3DES. Thus, the integrity and confidentiality of data is preserved and ensures that the data is free from any kind of tampering Customer private database contains information regarding his/her balance and transaction which can only be accessed by the customer. It acts as his/her personalized e-wallet. Payments are made through the payment server using the prepaid scheme as mentioned before. The retrieval will happen through the PIR client and server. Connections are secured by SSL and digital signatures are implemented to safe guard authenticity.

Experimentation and Results: This section deals with the results of the experimentation carried out on implementation of the proposed scheme. We have used the single database C-PIR scheme proposed by Aguilar-Melchor and Gaborit [7]. The front end is a basic setup of an e-commerce website and databases have been setup in MySql. The model was run on an Intel Core2 Duo T5550, with 1.83 GHz clock speed, 2 GB of RAM, and 2 MB of L2 cache. The results of experimentation are shown in the following three tables. Table 1.1 summarizes the time taken by the system at various stages of retrieval, which includes operation of PIR protocol, site-to-site communication, and data transfer via LAN. The values have been measured for different file sizes and are taken as an average of results obtained after multiple trials.

DB Size (MB)	**Returned bits**	**PIR reply generation**	**PIR reply decoding**	**Inter-site communication**	**LAN**
0.5	2^{21}	54 sec	18 sec	2 sec	0.5 sec
2	2^{23}	56 sec	33 sec	2 sec	1.2 sec
5	2^{24}	60 sec	58 sec	2 sec	3.8 sec
10	2^{25}	63 sec	93 sec	2 sec	9.5 sec

Table 1.1: Time taken by different phases of retrieval for different DB sizes

Table 1.1 gives a list of different PIR schemes available and a performance comparison of their retrieval times. The metric selected for comparison is time required to retrieve a 10 MB file from a database containing 1000 files. From the table it is clearly observed that the C-PIR scheme by Aguilar-Melchor and Gaborit is the fastest amongst the given set and therefore has been used by us. Further, parallelizing the PIR scheme will result in faster

operations and hence will reduce the retrieval time considerably.

C-PIR Protocol schemes	**Result retrieval Time (10 MB file from 1000 files)**
Limpaa	31 hrs
Gentry and Ramzan	16 hrs
Anonymity based	73 min
Aguilar-Melchor (We have used this scheme)	11.32 min

Table 1.1(A): Comparison of retrieval time of different C-PIR schemes

Security Features Addressed	**Implementation detail in our system**
User Authentication	Authentication module
User Privacy	Single-database PIR scheme
Data privacy	Private databases
Data Integrity	Private key encryption
Encryption used	3-DES encryption scheme
SSL (Secure Socket Layer)	Custom made certificate

Table 1.1(B): Security features addressed by our system

Conclusions: Private Information Retrieval protocol gives an idea of data privacy as well as data confidentiality. This paper also propose the use of PIR E-Commerce appication for implementation to increase speed of execution of protocol, which will leads to practical usability of PIR schemes in real world. By looking systematically at the body of work done on outsourcing databases, it is clear that several areas have received deep treatment and some clear recommendations for outsourcing databases as a service have emerged. For each approach, several methods have been studied in isolation and the currently proposed techniques fall short of the desirable protection issues. One should identify several performance metrics like cryptographic overhead, message overhead and transaction

overhead related to different database transactions. One should identify protocols that will reduce this overhead especially for outsourced databases research and define protocols that will reduce the cryptographic overhead for maintaining the integrity of data in such solutions. Many existing protection techniques including encryption, data fragmentation or slicing etc. compromise data integrity. Thus future developments in this issue needs to be considered and propose solutions for protecting the constitution of database from inference attacks. Towards this, one should propose protocols that will randomize queries and results so that no information is leaked to the attackers. One idea to reduce the inference is to make the transaction activity uniform across all queries.

References:

[1]. Carlos Aguilar Melchor, "High-Speed Single-Database PIR Implementation" in2008.

[2]. L.Ballard, S. Kamara, and F. Monrose. Achieving efficient conjunctive keywordsearchesoverencrypteddata.InS.Qing,W.Mao,J.López,andG.Wang,editors, ICICS 05, volume 3783 of LNCS, pages 414–426. Springer, Dec.2005.

[3]. M. Bellare, A. Boldyreva, and A. O'Neill. Deterministic and efficiently searchable encryption. In A. Menezes, editor, CRYPTO 2007, volume 4622 of LNCS, pages 535–552. Springer, Aug. 2007.

[4]. D. J. Bernstein. Faster square roots in annoyingfinitefields.http://cr.yp.to/papers/sqroot.pdf,2001.

[5]. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. In C. Cachin and J. Camenisch, editors, EUROCRYPT 2004, volume 3027 of LNCS, pages 506–522. Springer, May 2004.

[6]. D. Boneh and B. Waters. Conjunctive, subset, and range queries on encrypted data. In S. P. Vadhan, editor, TCC 2007, volume 4392 of LNCS, pages 535–554. Springer, Feb.2007.

[7]. J. W. Byun, D. H. Lee, and J. Lim. Efficient conjunctive keyword search on encrypted data storage system. In EuroPKI, pages 184–196, 2016.

[8]. D.Cash,J.Jaeger,S.Jarecki,C.Jutla,H.Krawczyk,M.-C.Roşu,andM.Steiner. Dynamic Searchable Encryption in Very Large Databases: Data Structures and Imple- mentation. 21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, 2014.,2014.

[9]. D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner. Highly-scalable searchable symmetric encryption with support for boolean queries.

Crypto'2013.CryptologyePrintArchive,Report2013/169,Mar.2013.http://eprint.iacr.org/2013/169.

[10]. Y.-C. Chang and M. Mitzenmacher. Privacy preserving keyword searches on remote encrypted data. In J. Ioannidis, A. Keromytis, and M. Yung, editors, ACNS 05, volume 3531 of LNCS, pages 442–455. Springer, June 2005.

[11]. M. Chase and S. Kamara. Structured encryption and controlled disclosure. In ASI- ACRYPT 2010, LNCS, pages 577–594. Springer, Dec. 2010.

[12]. E. D. Cristofaro, Y. Lu, and G. Tsudik. Efficient techniques for privacy-preserving sharing of sensitive information. In J. M. McCune, B. Balacheff, A. Perrig, A.-R. Sadeghi, A. Sasse, and Y. Beres, editors, TRUST, volume 6740 of Lecture Notes in Computer Science, pages 239–253. Springer, 2011.

[13]. R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric en- cryption: improved definitions and efficient constructions. In A. Juels, R. N. Wright, and S. Vimercati, editors, ACM CCS 06, pages 79–88. ACM Press, Oct. / Nov. 2006.

[14]. M. J. Freedman, Y. Ishai, B. Pinkas, and O. Reingold. Keyword search and oblivious pseudorandom functions. In J. Kilian, editor, TCC 2005, volume 3378 of LNCS, pages 303–324. Springer, Feb. 2005.

[15]. E.-J. Goh. Secure indexes. Cryptology ePrint Archive, Report2003/216,2003.http://eprint.iacr.org/.

[16]. P. Golle, J. Staddon, and B. R. Waters. Secure conjunctive keyword search over en- crypted data. In M. Jakobsson, M. Yung, and J. Zhou, editors, ACNS 04, volume 3089 of LNCS, pages 31–45. Springer, June2004.

[17]. Y. Huang and I. Goldberg. Outsourced private information retrieval with pricing and accesscontrol.TechnicalReport2013-11,CentreforAppliedCryptographicResearch (CACR), University of Waterloo, Feb.2013.

[18]. IARPA. Security and Privacy Assurance Research (SPAR) Program - BAA, 2011.

[19]. S. Kamara and K. Lauter. Cryptographic cloud storage. In Financial Cryptography Workshops, pages 136–149, 2018.

[20]. S. Kamara, C. Papamanthou, and T. Roeder. Dynamic searchable symmetric encryp- tion. In Proc. of CCS'2012, 2012.

[21]. K. Kurosawa and Y. Ohtaki. UC-secure searchable symmetric encryption. In

Financial Cryptography, page 285, 2012.

[22]. M. Naor and O. Reingold. Number-theoretic constructions of efficient pseudo-random functions. In 38th FOCS, pages 458–467. IEEE Computer Society Press, Oct. 1919.

[23]. V. Pappas, B. Vo, F. Krell, S. G. Choi, V. Kolesnikov, A. Keromytis, and T. Malkin.Blind Seer: A Scalable Private DBMS. Manuscript,2013.

[24]. E. Shi, J. Bethencourt, H. T.-H. Chan, D. X. Song, and A. Perrig. Multi-dimensional range query over encrypted data. In 2007 IEEE Symposium on Security and Privacy, pages350–364.IEEEComputerSocietyPress,May2007.

[25]. B. R. Waters, D. Balfanz, G. Durfee, and D. K. Smetters. Building an encrypted and searchable audit log. In NDSS 2004. The Internet Society, Feb. 2004.

[26]. WSJ. U.S. Terrorism Agency to Tap a Vast Database of Citizens. Wall Street Journal 12/13/12. http://alturl.com/ot72x

[27]. Carlos Aguilar Melchor, Philippe Gaborit "A Fast Private Information Retrieval Protocol" in ISIT 2008, Toronto, Canada, July 6 - 11,2008.

[28]. Andrew Clarke, Eric Pardede "Outsourced XMLDatabase:Query Assurance Optimization," 24th IEEE International Conference on Advanced Information Networking and Applications, ICEGOV2019,

[29]. Carlos Aguilar Melchor, Philippe Gaborit, "A Fast Private Information Retrieval Protocol," ISIT 2, pp.1848-1852, July 6 - 11, 2020.

[30]. D. Abril, G. Navarro-Arribas, V. Torra, "Towards Privacy Preserving Information Retrieval Through Semantic Microaggregation,"IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010.

[31]. Wang J, Chen X, Huang X, You I, Xiang Y (2015) Verifiable auditing for outsourced database in cloud computing. IEEE Trans Compute 64(11):3293–3303.