

Private Database on the Cloud: Models, Issues

Madhuri P.Gole¹, Prof. Sachin Vyawhare²

M. E Scholar, Department of CSIT, Sanmati Engineering College, Washim, Maharashtra, India¹

Assistant Professor, Department of CSIT, Sanmati Engineering College, Washim, Maharashtra, India²

ABSTRACT: Privacy and security of big data is emerging as one among the most relevant research challenges of recent years, also stirred-up by a wide family of critical applications ranging from scientific computing to social network analysis and mining, from data stream management to smart cities, and so forth. Traditionally, the issue of making (even very-large) databases private and secure has a long history in the context of encrypted databases but, when specifically considered in the Cloud setting, it poses new requirements and challenges to deal with, with particular regard to the scalability of solutions. In line with this emerging research trend, this paper focuses the attention on state-of-the-art proposals in the area of private databases over Clouds, and proposes critical comments about pros and cons of actual research efforts along with future research directions to be considered in future years.

KEYWORDS: Big Data, Privacy and Security of Big Data, Private Databases on the Cloud, Encrypted Cloud Databases

I. INTRODUCTION

Nowadays, Clouds are among the most challenging computational infrastructures of interest for actual research communities. Clouds not only pose architectural and infrastructural issues but also paradigm-oriented research challenges. Among these, privacy and security of big data on the Cloud play, without doubts, a dominant role (e.g., [13], [27]). This happens, for instance, in the so-called hybrid Clouds (e.g., [37]), which are emerging as one of the most popular Cloud environments for modern applications, such as energy-efficient computing (e.g., [5]), due to the fact that, in such Clouds, outsourced databases (e.g., [28]) are very often introduced and managed. Indeed, it is easy to note that a plethora of Cloud-based applications, such as scientific computing, social network analysis and mining, data stream management, smart cities and so forth, expose to the requirement of accessing, managing and processing big data over Clouds. From this, it is natural to derive the need for models, techniques and algorithms for defining privacy-preserving database management solutions over Clouds, with scalable extensions towards more specific big data processing. It is a matter of fact, indeed, the strong correlation between Cloud architectures and big data management (e.g., [3]).

Private databases on the Cloud introduce a wide family of research issues. One is represented by the issue of securely accessing a Cloud database, which has attracted a relevant amount of attention from the research community (e.g., [29]). Basically, here the problem consists in providing effective and efficient secure access schemes to Cloud databases, by also ensuring a well-recognized features like: grant management, grant revocation, dynamic access, and so forth. Another relevant issue is represented by the requirement of making Cloud databases privacy-preserving, meaning that they must ensure the privacy of data during common (data) management tasks like indexing, query processing, information retrieval, and so forth.

II. HYBRID CLOUDS: MODELS, DEFINITIONS AND REFERENCE ARCHITECTURE

Hybrid Clouds are a special case of inter-Clouds (e.g.,[10]) where private and public Clouds are interconnected and integrated. The International Data Corporation (IDC) study predicted that, in 2015 [2], chief information managers will move to hybrid Cloud and, as part of this migration, existing deficiencies in service management will become evident, forcing investment in automation and consumption of externally managed services as alternatives to on-premises deployment of Cloud. Key trends that will emerge in the Cloud market include demand for integrated software development methods that stress communication, collaboration, integration, automation and measurement of

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 3, March 2018

cooperation between software developers and other IT professionals. There is also demand for higher adoption of virtualization solutions and increased emphasis on operational governance and security management. Among these challenges, privacy of databases “disseminated” in the Cloud is also comprised.

As a reference architecture of hybrid-Cloud systems, in the following we provide a description of Eco Multi Cloud [18], a hierarchical framework for the efficient distribution of the workload on a multi-site scenario among geographically distributed interconnected DCs. It allows for an integrated and homogeneous management of heterogeneous platforms but at the same time it preserves the autonomy of single sites. Through the self-organizing and adaptive nature of the approach, the Virtual Machines (VM) migrations are performed asynchronously, both location-wise and time-wise, and with a tunable rate managed by DC administrators. The Eco Multi Cloud hierarchical architecture is composed of two layers:

- The lower layer is used to allocate the workload within single DCs: each site adopts its own strategy to assign VMs internally, with local consolidation algorithms (possibly different from site to site). The lower layer collects information about the state of the local DC, and passes it to the upper layer.
- The upper layer is able to exchange information among geographically dispersed and interconnected sites and drive the distribution of VMs among the DCs.

III.STATE-OF-THE-ART TECHNIQUES FOR SUPPORTING PRIVATE DATABASES OVER CLOUDS

The research community has devoted quite a relevant amount of attention to the issue of supporting private databases over Clouds. In the following, we provide an overview on some recent relevant proposals. [19] Focuses the attention on the problem of supporting private Nearest-Neighbor (NN) queries over databases in the context of distributed mobile environments, which well-marry with the case of (mobile) Clouds. In particular, authors are interested in supporting the privacy of users via hiding their exact user coordinates to un-trusted entities that are delivering mobile services associated to Points of Interest (POI). To this end, a two-step approach for supporting private location-based query processing is defined. At the first step, the locations of users are hidden inside so-called Cloaking Regions (CR); at the second step, location data (which are embedded in mobile databases) are encrypted by using a suitable Private Information Retrieval (PIR) protocol (e.g., [41]). In addition to this, ad-hoc algorithms for supporting approximate and exact NN queried are provided and experimentally assessed.[17] Considers again distributed settings as reference application scenario and studies the problem of supporting efficient and private approximations of distributed databases calculations. Authors depict a target environment where data are collected in different and non-cooperative databases. For such specific settings, the issue of supporting privacy-preserving distributed calculations over these kind of databases has been largely investigated. While some existent approaches exist, computational costs are still a terrible bottleneck to deal with. Inspired by this main evidence, authors provide an innovative solution that trades-off between performance and accuracy of distributed calculations via data sampling (e.g., [22]). They focus on the specific case of supporting sampling for separate, non-collaborating, vertically-partitioned data sets, yet supporting privacy-preservation requirements. In order to provide a proof-of-concept, authors apply the proposed method to approximation of intersection set both without and with privacy preserving mechanism, and derive an interesting theoretical analysis on the error found that, along with the experimental assessment and analysis of the proposed method, completes the contributions of the paper.

IV.PRIVATE DATABASES OVER CLOUDS: FUTURE RESEARCH PERSPECTIVES

Several research issues will play a prominent role in next generation techniques for supporting private databases over Clouds. In the following, we report on some of most noticeable of them.

4.1 Indexing Data Structures for Private Cloud Databases

When dealing with private Cloud databases, the issue of defining and building effective and efficient indexing data structures enabled with data management features (e.g., searching a certain item) is very relevant. Indeed, the encryption task must not prevent the support of such operations. This need, combined with the obvious scalability requirement, is a clear research challenge of future efforts.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 3, March 2018

4.2 Advanced Query Operators over Private Cloud Databases

Querying private Cloud databases not only implies the support of basic query operators (e.g. SELECT, UPDATE, and so forth), but also advanced query operators such as: aggregate operators, rank-based operators, IR-style operators, OLAP style operators, and so forth. The issue of supporting this rather-wide family of query operators on top of encrypted databases is significant as well, especially when these operators are used as “basic core layer” for more complex big data analytics tools

4.3 Dealing with Mixed-Encrypted Cloud Databases Database

Encryption usually works at a certain granularity that, for instance, can be attribute-level or tuple-level. Both have advantages and disadvantages. For instance, attribute-level encryption may be useful during JOIN query processing, while tuple-level encryption ensures a more-compact and better-for access-procedures solution. Starting from these considerations, an important line of research for private Cloud databases will be represented by the issue of introducing and devising mixed encryption schemes, i.e. encryption schemes that incorporate different granularities, which may turn to be extremely useful in complex analytical settings.

4.4 Preference-Aware Query Processing over Encrypted Cloud

Databases Preference-aware query processing (e.g., [33]) is a fortunate research niche within the database community. It foresees paradigms according to which users can embed so called “preferences” in their query interaction, and the query processor filters-out the results by pruning those tuples that do not satisfy those preferences. Skyline queries are a kind of preference-aware query. While skyline query processing over encrypted data has been studied recently (e.g., [8]), to the best of our knowledge there are not relevant research efforts in the context of preference-aware query processing over encrypted data. We recognize the latter as one of the most prominent research challenge for private Cloud databases.

4.5 Distributed Query Processing over Private Cloud Databases

With the goal of supporting big data analytics tasks, query processing over private Cloud databases usually performs in a distributed manner, i.e. private databases stored in different nodes of the reference Cloud architecture must be accessed and queried. This calls for solutions capable of devising distributed query optimization plans in the presence of encrypted data. A possible alternative is represented by the idea of trading off accuracy of the query answers with the privacy of target Cloud databases. Adaptive query optimization plans are also interesting to this purpose.

4.6 Scalability Issues One of the requirements of private Cloud

Databases are, without doubts, scalability. Indeed, Cloud databases are used to run in complex environments where performance plays a leading role. To this end, not only the specific query algorithms, which are likely to run online, but also the encryption algorithms, which are likely to run offline, must marry elastic metaphors and scale-up on massive big data. This is still an open issue to be widely investigated, as it heavily impacts on the reliability of real-life big data applications and systems.

4.7 Integration with MapReduce Engines

MapReduce [16] is a novel computational paradigm according to which complex procedures are decomposed in simpler sub-procedures (the map phase) and later the final computation result is composed by combining the local results of such sub-procedures (the reduce phase). This paradigm is likely to be integrated with private Cloud databases in a smoothly way, even with the goal of computing the encrypted versions of massive Cloud databases in a parallel/distributed-manner. The integration among private Cloud databases and MapReduce engines is really likely to become one of the most interesting and exciting research avenues in the field of big data processing.

4.8 Integration with IoT Architectures

When integrated with IoT architectures, private Cloud databases expose a higher risk of privacy breaches. This is due to several reasons. First, devices are characterized by very heterogeneous drivers and communication protocols.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 3, March 2018

Second, several devices do not have an explicit. Internet connection protocol. Third, for some devices (e.g., those used in domotics applications) the mobility is really high, hence this complicates the security- and privacy preserving process. All these concepts are likely to be further studied in next research efforts.

4.9 Encryption Mechanisms over Uncertain Cloud Databases

Nowadays, dealing with privacy-preserving issues over uncertain data sets (e.g., [15]) is a gold research topic for the database community. Uncertainty can be found in a plethora of real-life application scenarios, hence from this evidence several research efforts have been devoted recently. With the same spirit, this problem will soon touch the field of encryption mechanisms for Cloud databases. Hence, devising models, algorithms and techniques for dealing with encryption mechanisms over uncertain Cloud databases is regarded as a rich line of research for the future years.

V. CONCLUSION

Recently, a great deal of attention has been devoted to private databases over Clouds, as stirred-up by both (i) interesting theoretical problems and challenges and (ii) novel applications where secure and privacy-preserving access and management of embedded databases pose strict requirements, like, for instance, the case of emerging IoT architectures. In response to this “call for arms”, a lot of research efforts has been devoted to design and implement a wide spectrum of solutions and achievements. Inspired by this trend, this paper has focused the attention on state-of-the-art proposals in the area of private databases over Clouds. A critical evaluation of current research efforts along with perspectives on future research directions have been proposed as well.

REFERENCES

- [1] Cisco and Partners to Build World's Largest Global Intercloud. <http://newsroom.cisco.com/press-release-content?articleId=1373639>, 2015. Accessed: 2014-03-24.
- [2] IDC Reveals Cloud Predictions for 2015. <https://www.idc.com/getdoc.jsp?containerId=prUS25350114>, 2015. Accessed: 2014-12-15.
- [3] D. Agrawal, S. Das, and A. El Abbadi. Big data and cloud computing: current state and future opportunities. In EDBT 2011, 14th International Conference on Extending Database Technology, Uppsala, Sweden, March 21-24, 2011, Proceedings, pages 530–533, 2011.
- [4] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Order-preserving encryption for numeric data. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Paris, France, June 13-18, 2004, pages 563–574, 2004.
- [5] R. Bianchini, S. U. Khan, and C. Mastroianni. Guest editors introduction: Special issue on green and energy-efficient cloud computing: Part i. *IEEE Transactions on Cloud Computing*, 4(2):119–121, April 2016.
- [6] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee. Linked data on the web. In Proceedings of the 17th International Conference on World Wide Web, WWW 2008, Beijing, China, April 21-25, 2008, pages 1265–1266, 2008.
- [7] S. B'orz's'onyi, D. Kossmann, and K. Stocker. The skyline operator. In Proceedings of the 17th International Conference on Data Engineering, April 2-6, 2001, Heidelberg, Germany, pages 421–430, 2001.
- [8] S. Bothe, A. Cuzzocrea, P. Karras, and A. Vlachou. Skyline query processing over encrypted data: An attribute-order-preserving-free approach. In Proceedings of the First International Workshop on Privacy and Security of Big Data, PSBD@CIKM 2014, Shanghai, China, November 7, 2014, pages 37–43, 2014.
- [9] C. Branciard, N. Gisin, B. Kraus, and V. Scarani. Security of two quantum cryptography protocols using the same four qubit states. *Physical Review A*, 72(3):032301, 2005.
- [10] R. Buyya, R. Ranjan, and R. N. Calheiros. Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services. In Algorithms and Architectures for Parallel Processing, 10th International Conference, ICA3PP 2010, Busan, Korea, May 21-23, 2010. Proceedings. Part I, pages 13–31, 2010.
- [11] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: Improved definitions and efficient constructions. *Journal of Computer Security*, 19(5):895–934, 2011.
- [12] A. Cuzzocrea. Analytics over big data: Exploring the convergence of datawarehousing, OLAP and data-intensive cloud infrastructures. In 37th Annual IEEE Computer Software and Applications Conference, COMPSAC 2013, Kyoto, Japan, July 22-26, 2013, pages 481–483, 2013.
- [13] A. Cuzzocrea. Privacy and security of big data: Current challenges and future research perspectives. In Proceedings of the First International Workshop on Privacy and Security of Big Data, PSBD@CIKM 2014, Shanghai, China, November 7, 2014, pages 45–47, 2014.
- [14] A. Cuzzocrea. A reference architecture for supporting secure big data analytics over cloud-enabled relational databases. In 40th IEEE Annual Computer Software and Applications Conference, COMPSAC Workshops 2016, Atlanta, GA, USA, June 10-14, 2016, pages 356–358, 2016.
- [15] A. Cuzzocrea, C. K. Leung, and R. K. MacKinnon. Mining constrained frequent itemsets from distributed uncertain data. *Future Generation Comp. Syst.*, 37:117–126, 2014.
- [16] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, 2008.
- [17] P. Derbeko, S. Dolev, E. Gudes, and J. D. Ullman. Efficient and private approximations of distributed databases calculations. *CoRR*, abs/1605.06143, 2016.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 3, March 2018

- [18] A. Forestiero, C. Mastroianni, M. Meo, G. Papuzzo, and M. Sheikhalishahi. Hierarchical approach for efficient workload management in geo-distributed data centers. *IEEE Transactions on Green Communications and Networking*, August 2016. Early Access.
- [19] G. Ghinita, P. Kalnis, M. Kantarcioglu, and E. Bertino. Approximate and exact hybrid algorithms for private nearest-neighbor queries with database protection. *GeoInformatica*, 15(4):699–726, 2011.
- [20] C. A. Goble, A. J. G. Gray, L. Harland, K. Karapetyan, A. Loizou, I. Mikhailov, Y. Rankka, S. Senger, V. Tkachenko, A. J. Williams, and E. L. Willighagen. Incorporating commercial and private data into an open linked data platform for drug discovery. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II*, pages 65–80, 2013.
- [21] S. Guastello and M. Rieke. A review and critique of honesty test research. *Behavioral Sciences & the Law*, 9(4):501–523, 1991.
- [22] A. Gupta. Sampling techniques for statistical databases. In *Encyclopedia of Database Systems*, page 2467. 2009.
- [23] B. Hore, S. Mehrotra, and H. Hacig`um`us. Managing and querying encrypted data. In *Handbook of Database Security - Applications and Trends*, pages 163–190. 2008.
- [24] IBM, P. Zikopoulos, and C. Eaton. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media, 1st edition, 2011.
- [25] Y. Ishai, E. Kushilevitz, S. Lu, and R. Ostrovsky. Private large-scale databases with distributed searchable symmetric encryption. In *Topics in Cryptology - CT-RSA 2016 - The Cryptographers' Track at the RSA Conference 2016, San Francisco, CA, USA, February 29 - March 4, 2016, Proceedings*, pages 90–107, 2016.
- [26] V. Karwa, S. Raskhodnikova, A. D. Smith, and G. Yaroslavtsev. Private analysis of graph structure. *ACM Trans. Database Syst.*, 39(3):22:1–22:33, 2014.
- [27] P. Li, S. Guo, T. Miyazaki, M. Xie, J. Hu, and W. Zhuang. Privacy-preserving access to big data in the cloud. *IEEE Cloud Computing*, 3(5):34–42, 2016.
- [28] Z. Liu, X. Chen, J. Yang, C. Jia, and I. You. New order preserving encryption model for outsourced databases in cloud environments. *J. Network and Computer Applications*, 59:198–207, 2016.
- [29] S. Namasudra and P. Roy. Secure and efficient data access control in cloud computing environment: A survey. *Multiagent and Grid Systems*, 12(2):69–90, 2016.
- [30] M. Naor and M. Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, May 13-17, 1990, Baltimore, Maryland, USA, pages 427–437, 1990.
- [31] P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptology - EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2-6, 1999, Proceeding*, pages 223–238, 1999.
- [32] P. Peng, L. Zou, L. Chen, X. Lin, and D. Zhao. Answering subgraph queries over massive disk resident graphs. *World Wide Web*, 19(3):417–448, 2016.
- [33] H. Qu and A. Labrinidis. Preference-aware query and update scheduling in web-databases. In *Proceedings of the 23rd International Conference on Data Engineering, ICDE 2007, The Marmara Hotel, Istanbul, Turkey, April 15-20, 2007*, pages 356–365, 2007.
- [34] S. Sakr, L. Zhao, and A. Liu. Clouddb autoadmin: A consumer-centric framework for SLA management of virtualized database servers. In *Large Scale and Big Data - Processing and Management.*, pages 357–388. 2014.
- [35] Y. Sang and H. Shen. Efficient and secure protocols for privacy-preserving set operations. *ACM Trans. Inf. Syst. Secur.*, 13(1), 2009.
- [36] E. Shmueli, R. Vaisenberg, Y. Elovici, and C. Glezer. Database encryption: an overview of contemporary challenges and design considerations. *SIGMOD Record*, 38(3):29–34, 2009.
- [37] B. Sotomayor, R. S. Montero, I. M. Llorente, and I. T. Foster. Virtual infrastructure management in private and hybrid clouds. *IEEE Internet Computing*, 13(5):14–22, 2009.
- [38] Q. To, B. Nguyen, and P. Pucheral. Private and scalable execution of SQL aggregates on a secure decentralized architecture. *ACM Trans. Database Syst.*, 41(3):16, 2016.
- [39] S. Wang, D. Agrawal, and A. El Abbadi. Towards practical private processing of database queries over public data. *Distributed and Parallel Databases*, 32(1):65–89.
- [40] F. Xia, L. T. Yang, L. Wang, and A. V. Vinel. Internet of things. *Int. J. Communication Systems*, 25(9):1101–1102, 2012.
- [41] S. Yekhanin. Private information retrieval. *Commun. ACM*, 53(4):68–73, 2010.
- [42] M. L. Yiu, C. S. Jensen, J. Møller, and H. Lu. Design and analysis of a ranking approach to private location-based services. *ACM Trans. Database Syst.*, 36(2):10, 2011.
- [43] F. Yu, D. Qiu, H. Situ, X. Wang, and S. Long. Enhancing user privacy in sarg04-based private database query protocols. *Quantum Information Processing*, 14(11):4201–4210, 2015.
- [44] Q. Zhang, L. Cheng, and R. Boutaba. Cloud computing: state-of-the-art and research challenges. *J. Internet Services and Applications*, 1(1):7–18, 2010.