Data Security Issues in Cloud Scenarios

Sabrina De Capitani di Vimercati, Sara Foresti, and Pierangela Samarati

Computer Science Department
Università degli Studi di Milano – 26013 Crema, Italy
firstname.lastname@unimi.it

Abstract. The amount of data created, stored, and processed has enormously increased in the last years. Today, millions of devices are connected to the Internet and generate a huge amount of (personal) data that need to be stored and processed using scalable, efficient, and reliable computing infrastructures. Cloud computing technology can be used to respond to these needs. Although cloud computing brings many benefits to users and companies, security concerns about the cloud still represent the major impediment for its wide adoption.
We briefly survey the main challenges related to the storage and processing of data in the cloud. In particular, we focus on the problem of protecting data in storage, supporting fine-grained access, selectively sharing data, protecting query privacy, and verifying the integrity of computations.

1 Introduction

The wide use and advancements of Information and Communication Technologies (ICTs) have profoundly changed our lives. The proliferation of any kind of smart devices that can easily connect to the Internet together with the availability of (almost free) wireless connections anywhere have led to a more distributed computing environment, which is expected to grow in the near future. The huge amount of data coming from these devices, which must be rapidly stored and processed, introduces the need of developing scalable, efficient, and reliable computing infrastructures. Cloud computing is a collection of technologies and services that provides an answer to these needs, making virtually unlimited storage space and computing power available at affordable prices. Users and companies can therefore store their data at external cloud providers, access reliable and efficient services provided by third parties, and use computing power available in multiple locations across the network. Although the use of cloud computing has clear economic advantages, the collection, storage, processing, and sharing of (often personal) data in the cloud pose several security concerns (e.g., [25,32,33,34,35,43,44]). In particular, when data are moved to the cloud, the data owner loses control over them and often even knows neither the location where the data are stored nor the organizations responsible for their management.

It is important to observe that the protection of data is a key aspect not only for the success of today's cloud infrastructures but also for the proper development of applications in emerging areas such as Internet of Things and Big Data analytics, which are characterized by huge amounts of data that need to be shared and processed by different parties. In addition to data privacy, another important concern is the security and privacy of the data processing that may involve different parties with the need of sharing information and performing distributed computations. Ensuring that the data processing is carried out securely is a significant challenge. The goal of this paper is to provide an overview of the data protection challenges that need to be addressed when using the cloud to store and process data, and to illustrate existing proposals addressing them. In particular, we focus on the problem of protecting data in storage while supporting fine-grained access, selectively sharing data among different users/data owners, supporting query privacy, and verifying the integrity of data computations.

2 Protection of data in storage and fine-grained access

A well-known problem that characterizes the use of a cloud infrastructure is the *loss of control* over data. A data owner storing her data in the cloud often knows neither where her data are stored nor the organizations involved in their management. Encryption services are therefore at the basis of current solutions for protecting the confidentiality and integrity of the data from malicious users and from the providers themselves (e.g., [7,25,43]). Encryption can be applied both at the server side or at the owner side. In the first case, data owners have full trust in the cloud provider managing their data (e.g., Google Cloud Storage, iCloud), which has full access to their data, and can enjoy full functionality. In the second case, the data are encrypted before outsourcing them to a cloud provider (e.g., Boxcryptor, SpiderOak), and data owners can enjoy protection but limited functionality. In fact, the cloud provider cannot access the data it stores in plaintext, and hence cannot directly evaluate queries or execute computations, because they would require operating on encrypted data. Current approaches enabling the execution of queries on encrypted data are based on: the use of specific cryptographic techniques supporting keyword-based searches (e.g., [8]); homomorphic encryption supporting any operation but with high performance overhead (e.g., [29]); different layers of encryption each supporting specific operations [40]; or indexes, that is, metadata attached to the data and used for fine-grained information retrieval and execution of specific queries, depending on the kind of index (e.g., [9,31,46]).

Although all these approaches provide the capability of evaluating queries on encrypted data, they make query evaluation more expensive or not always possible. To avoid this problem, alternative solutions to encryption use *data fragmentation* to protect data confidentiality. Data fragmentation is based on the observation that often data are not sensitive per se but the association among data is sensitive and needs to be protected. For instance, while a list of names

and a list of illnesses are not sensitive, the association of each name in the first list with the illness from which she suffers in the second list needs to be kept confidential. In this case, encrypting both names and illnesses may not be necessary. Data fragmentation comes at this point: the use of encryption is limited or avoided by splitting data in different *fragments* (e.g., [1,10,11,12,15]). Fragments are computed in such a way that sensitive associations, called *confidentiality constraints*, are broken. To guarantee that sensitive associations cannot be reconstructed, fragments are designed in such a way to guarantee their unlinkability or are stored at different cloud providers.

Besides protecting data confidentiality and integrity, attention has been also dedicated to solutions aimed at proving to remote parties guarantees that the management of data by a cloud provider complies with the service level agreement, guaranteeing their availability (e.g., [6,36]).

3 Selective information sharing

The proposals aimed at protecting the confidentiality of the data in the cloud are typically based on the implicit assumption that any authorized user can access the whole data content. This assumption, however, is in contrast with the nature of today's open and dynamic scenarios, where different users might need to have different views on the outsourced data (e.g., [2,18]). Since neither the cloud provider storing the data nor the data owner can enforce the access control policy for confidentiality and efficiency reasons, respectively, existing solutions are based on the use of *attribute-based encryption* (ABE) and *selective encryption* techniques. ABE is a public-key encryption schema that regulates access to resources on the basis of policies defined on *descriptive attributes* associated with users [30,48]. ABE-based approaches have been recently widely investigated, and several solutions have been proposed for improving the support of policy updates (e.g., [28,42]). Selective encryption is based on the idea that different resources are encrypted with different keys and that a key derivation strategy (e.g., [5]), which relies on the definition of a *key derivation hierarchy*, is adopted to translate read access privileges into the ability of users to derive the encryption keys used to protect the data they are authorized to access [19,20]. To easily support changes to the access control policy on a resource (i.e., grant/revoke operations) without downloading the resource, decrypting it, changing the key derivation hierarchy, and re-encrypting the resource with the new key, two layers of encryption (each characterized by its own encryption policy) are used. One layer is managed by the data owner, while the other is managed directly by the cloud provider and access to data is granted to users who know the encryption keys of both the layers. Therefore, the management of grant and revoke operations can be partially delegated to the cloud provider storing the data.

Few works have also extended the selective encryption techniques to enforce write privileges (e.g., [14,41,42]), to support the presence of multiple data owners selectively sharing their data among each other (e.g., [17]), and to support the release of data according to a subscription-based policy [13].

4 Query privacy

Several efforts have been dedicated to the development of efficient techniques for protecting *access confidentiality* and *pattern confidentiality*. Access confidentiality means that an observer (including the cloud provider storing the data) should not be able to infer the target of an access singularly taken. Pattern confidentiality means that an observer should not be able to infer whether the target of two different accesses is the same. Private Information Retrieval (PIR) techniques have been traditionally used to provide both these protection guarantees but they are computationally expensive and operate on plaintext data, therefore not protecting data confidentiality. Recent alternative solutions enhance existing index structures (e.g., B+-trees and Oblivious RAM [9,26,49]) to protect confidentiality of data, accesses, and patterns thereof. However, these solutions, while more efficient than PIR approaches, cause an overhead in access times (either for each access or when triggering expensive reordering of the underlying data structure), which make them not always applicable in real-life scenarios. Dynamically allocated data structures (e.g., [22,23,24,45,51]) represent a different approach to provide data, access, and pattern confidentiality, while guaranteeing a limited overhead in query evaluation and supporting concurrent accesses to the data. The basic idea of these solutions consists in moving the physical location where data are stored after each access (without leaving traces of such reallocations) so that an observer cannot make any inference on the data accessed.

5 Integrity of computations

When data are elaborated by cloud providers that are not fully trustworthy, there is the problem of verifying the integrity of such a computation, that is, verifying whether the result is *correct*, *complete*, and *fresh*. A result is: correct if the computation involves only genuine data; complete if the computation has been performed on the whole data collection and includes all resources satisfying the computation; fresh if the computation has been performed on the most recent version of the data. At a high level, existing solutions addressing this problem can be divided into two main classes: *deterministic* and *probabilistic*. Deterministic approaches are based on the definition of *authenticated data structures*, which are structures built over specific attributes (e.g., Merkle hash trees or signature chaining schemas [38,39]). A user submits a query to a cloud provider that executes it and returns the query result along with the information necessary for the user to verify the correctness and completeness of the query result. Such an information, called *verification object*, is computed with the help of an authenticated data structure. These techniques provide deterministic integrity guarantees but only for queries with conditions on the attribute(s) on which the data structure has been built. Probabilistic approaches complement the data with fictitious information or checks whose absence in a query result signals an integrity violation (e.g., [21,47,50]). Probabilistic approaches can detect an integrity violation for any query but with only probabilistic guarantees.

This means that while the absence of the expected fictitious information implies an integrity violation, their presence does not provide full guarantees of the integrity of the query result (the cloud provider might have just not missed the fictitious information inserted by the data owner). The possible presence of multiple providers in the computation complicates the scenario and requires the use of additional controls (e.g., [16]).

6 Conclusions

The adoption of cloud technologies to store and process huge amount of data, while bringing many benefits, also introduces novel security risks on the data. In this paper, we described challenges related to the management of data in the cloud, and described current solutions.

Acknowledgement. This work was supported in part by: the EC within the 7FP under grant agreement 312797 (ABC4EU) and within the H2020 under grant agreement 644579 (ESCUDO-CLOUD); the Italian Ministry of Research within PRIN project "GenData 2020" (2010RTFWBH).

References

1. Aggarwal, G., Bawa, M., Ganesan, P., Garcia-Molina, H., Kenthapadi, K., Motwani, R., Srivastava, U., Thomas, D., Xu, Y.: Two can keep a secret: A distributed architecture for secure database services. In: Proc. of the 2nd Biennial Conference on Innovative Data Systems Research (CIDR 2005). Asilomar, CA, USA (January 2005)
2. Ardagna, C., De Capitani di Vimercati, S., Paraboschi, S., Pedrini, E., Samarati, P., Verdicchio, M.: Expressive and deployable access control in open web service applications. IEEE Transactions on Service Computing (TSC) 4(2), 96–109 (April-June 2011)
3. Ardagna, C.A., Jhawar, R., Piuri, V.: Dependability certification of services: a model-based approach. Computing 97(1), 51–78 (October 2013)
4. Ardagna, C., Jajodia, S., Samarati, P., Stavrou, A.: Providing users' anonymity in mobile hybrid networks. ACM Transactions on Internet Technology (TOIT) 12(3), 1–33 (May 2013), article 7
5. Atallah, M., Frikken, K., Blanton, M.: Dynamic and efficient key management for access hierarchies. In: Proc. of the 12th ACM Conference on Computer and Communications Security (CCS 2005). Alexandria, VA, USA (November 2005)
6. Ateniese, G., Burns, R., Curtmola, R., Herring, J., Khan, O., Kissner, L., Peterson, Z., Song, D.: Remote data checking using provable data possession. ACM Transactions on Information and System Security (TISSEC) 14(1), 12:1–12:34 (May 2011)
7. Bowers, K., Juels, A., Oprea, A.: Hail: a high-availability and integrity layer for cloud storage. In: Proc. of the 16th ACM Conference on Computer and Communications Security (CCS 2009). Chicago, IL, USA (November 2009)
8. Cao, N., Wang, C., Li, M., Ren, K., Lou, W.: Privacy-preserving multikeyword ranked search over encrypted cloud data. In: Proc. of the 30th IEEE International Conference on Computer Communications (INFOCOM 2011). Shanghai, China (April 2011)

9. Ceselli, A., Damiani, E., De Capitani di Vimercati, S., Jajodia, S., Paraboschi, S., Samarati, P.: Modeling and assessing inference exposure in encrypted databases. ACM Transactions on Information and System Security (TISSEC) 8(1), 119–152 (February 2005)
10. Ciriani, V., De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Fragmentation and encryption to enforce privacy in data storage. In: Proc. of the 12th European Symposium On Research In Computer Security (ESORICS 2007). Dresden, Germany (September 2007)
11. Ciriani, V., De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Keep a few: Outsourcing data while maintaining confidentiality. In: Proc. of the 14th European Symposium On Research In Computer Security (ESORICS 2009). Saint Malo, France (September 2009)
12. Ciriani, V., De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Combining fragmentation and encryption to protect privacy in data storage. ACM Transactions on Information and System Security (TISSEC) 13(3), 22:1–22:33 (July 2010)
13. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Livraga, G.: Enforcing subscription-based authorization policies in cloud scenarios. In: Proc. of the 26th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy (DBSec 2012). Paris, France (July 2012)
14. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Livraga, G., Paraboschi, S., Samarati, P.: Enforcing dynamic write privileges in data outsourcing. Computers & Security 39, 47–63 (November 2013)
15. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Livraga, G., Paraboschi, S., Samarati, P.: Fragmentation in presence of data dependencies. IEEE Transactions on Dependable and Secure Computing (TDSC) 11(6), 510–523 (November/December 2014)
16. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Livraga, G., Paraboschi, S., Samarati, P.: Integrity for distributed queries. In: Proc. of the 2nd IEEE Conference on Communications and Network Security (CNS 2014). San Francisco, CA, USA (October 2014)
17. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Pelosi, G., Samarati, P.: Encryption-based policy enforcement for cloud storage. In: Proc. of the 1st ICDCS Workshop on Security and Privacy in Cloud Computing (SPCC 2010). Genova, Italy (June 2010)
18. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Psaila, G., Samarati, P.: Integrating trust management and access control in data-intensive web applications. ACM Transactions on the Web (TWEB) 6(2) (May 2012)
19. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Over-encryption: Management of access control evolution on outsourced data. In: Proc. of the 33rd International Conference on Very Large Data Bases (VLDB 2007). Vienna, Austria (September 2007)
20. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Encryption policies for regulating access to outsourced data. ACM Transactions on Database Systems (TODS) 35(2), 12:1–12:46 (April 2010)
21. De Capitani di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Integrity for join queries in the cloud. IEEE Transactions on Cloud Computing (TCC) 1(2), 187–200 (July-December 2013)
22. De Capitani di Vimercati, S., Foresti, S., Paraboschi, S., Pelosi, G., Samarati, P.: Efficient and private access to outsourced data. In: Proc. of the 31st International

Conference on Distributed Computing Systems (ICDCS 2011). Minneapolis, MN, USA (June 2011)
23. De Capitani di Vimercati, S., Foresti, S., Paraboschi, S., Pelosi, G., Samarati, P.: Supporting concurrency in private data outsourcing. In: Proc. of the 16th European Symposium On Research In Computer Security (ESORICS 2011). Leuven, Belgium (September 2011)
24. De Capitani di Vimercati, S., Foresti, S., Paraboschi, S., Pelosi, G., Samarati, P.: Distributed shuffling for preserving access confidentiality. In: Proc. of the 18th European Symposium On Research In Computer Security (ESORICS 2013). Egham, U.K. (September 2013)
25. De Capitani di Vimercati, S., Foresti, S., Samarati, P.: Selective and fine-grained access to data in the cloud. In: Jajodia, S., et al. (eds.) Secure Cloud Computing. Springer (2014)
26. Ding, X., Yang, Y., Deng, R.: Database access pattern protection without full-shuffles. IEEE Transactions on Information Forensics and Security (TIFS) 6(1), 189–201 (March 2011)
27. Donida Labati, R., Genovese, A., Piuri, V., Scotti, F.: Touchless fingerprint biometrics: a survey on 2d and 3d technologies. Journal of Internet Technology 15(3), 325–332 (May 2014)
28. Fangming, Z., Takashi, N., Kouichi, S.: Realizing fine-grained and flexible access control to outsourced data with attribute-based cryptosystems. In: Proc. of the 7th International Conference on Information Security Practice and Experience (ISPEC 2011). Guangzhou, China (May-June 2011)
29. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: Proc. of the 41st ACM Symposium on Theory of Computing (STOC 2009). Bethesda, MD, USA (May-June 2009)
30. Goyal, V., Pandey, O., Sahai, A., Waters, B.: Attribute-based encryption for fine-grained access control of encrypted data. In: Proc. of the 13th ACM Conference on Computer and Communications Security (CCS 2006). Alexandria, VA, USA (October-November 2006)
31. Hacigümüş, H., Iyer, B., Li, C., Mehrotra, S.: Executing SQL over encrypted data in the database-service-provider model. In: Proc. of the 21th ACM SIGMOD International Conference on Management of Data (SIGMOD 2002). Madison, WI, USA (June 2002)
32. Jhawar, R., Piuri, V.: Fault tolerance management in IaaS clouds. In: Proc. of the IEEE Conference in Europe about Space and Satellite Telecommunications (ESTEL 2012). Rome, Italy (October 2012)
33. Jhawar, R., Piuri, V., Samarati, P.: Supporting security requirements for resource management in cloud computing. In: Proc. of the 15th IEEE International Conference on Computational Science and Engineering (CSE 2012). Paphos, Cyprus (December 2012)
34. Jhawar, R., Piuri, V., Santambrogio, M.: A comprehensive conceptual system-level approach to fault tolerance in cloud computing. In: Proc. of the 2012 IEEE International Systems Conference (SysCon 2012). Vancouver, Canada (March 2012)
35. Jhawar, R., Piuri, V., Santambrogio, M.: Fault tolerance management in cloud computing: A system-level perspective. IEEE Systems Journal 7(2), 288–297 (June 2013)
36. Juels, A., Kaliski, B.: PORs: Proofs of retrievability for large files. In: Proc. of the 14th ACM Conference on Computer and Communications Security (CCS 2007). Alexandria, VA, USA (October-November 2007)

37. Labati, R.D., Piuri, V., Scotti, F.: Touchless Fingerprint Biometrics. Series in Security, Privacy and Trust, CRC Press (August 2015)
38. Li, F., Hadjieleftheriou, M., Kollios, G., Reyzin, L.: Authenticated index structures for aggregation queries. ACM Transactions on Information and System Security (TISSEC) 13(4), 32:1–32:35 (December 2010)
39. Pang, H., Jain, A., Ramamritham, K., Tan, K.: Verifying completeness of relational query results in data publishing. In: Proc. of the 24th ACM SIGMOD International Conference on Management of Data (SIGMOD 2005). Baltimore, MD, USA (June 2005)
40. Popa, R., Redfield, C., Zeldovich, N., Balakrishnan, H.: CryptDB: Protecting confidentiality with encrypted query processin. In: Proc. of the 23rd ACM Symposium on Operating Systems Principles (SOSP 2011). Cascais, Portugal (October 2011)
41. Raykova, M., Zhao, H., Bellovin, S.: Privacy enhanced access control for outsourced data sharing. In: Proc. of the 16th International Conference on Financial Cryptography and Data Security (FC 2012). Kralendijk, Bonaire (February-March 2012)
42. Ruj, S., Stojmenovic, M., Nayak, A.: Privacy preserving access control with authentication for securing data in clouds. In: Proc. of the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2012). Ottawa, Canada (May 2012)
43. Samarati, P.: Data security and privacy in the cloud. In: Proc. of the 10th International Conference on Information Security Practice and Experience (ISPEC 2014). Fuzhou, China (May 2014)
44. Samarati, P., De Capitani di Vimercati, S.: Cloud security: Issues and concerns. In: Murugesan, S., Bojanova, I. (eds.) Encyclopedia on Cloud Computing. Wiley (2016)
45. Stefanov, E., van Dijk, M., Shi, E., Fletcher, C., Ren, L., Yu, X., Devadas, S.: Path ORAM: An extremely simple Oblivious RAM protocol. In: Proc. of the 20th ACM Conference on Computer and Communications Security (CCS 2013). Berlin, Germany (November 2013)
46. Wang, H., Lakshmanan, L.: Efficient secure query evaluation over encrypted XML databases. In: Proc. of the 32nd International Conference on Very Large Data Bases (VLDB 2006). Seoul, Korea (September 2006)
47. Wang, H., Yin, J., Perng, C., Yu, P.: Dual encryption for query integrity assurance. In: Proc. of the 17th Conference on Information and Knowledge Management (CIKM 2008). Napa Valley, CA, USA (October 2008)
48. Waters, B.: Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In: Proc. of the 14th IACR International Conference on Practice and Theory of Public Key Cryptography (PKI 2011). Taormina, Italy (March 2011)
49. Williams, P., Sion, R.: Single round access privacy on outsourced storage. In: Proc. of the 19th ACM Conference on Computer and Communications Security (CCS 2012). Raleigh, NC, USA (October 2012)
50. Xie, M., Wang, H., Yin, J., Meng, X.: Integrity auditing of outsourced data. In: Proc. of the 33rd International Conference on Very Large Data Bases (VLDB 2007). Vienna, Austria (September 2007)
51. Yang, K., Zhang, J., Zhang, W., Qiao, D.: A light-weight solution to preservation of access pattern privacy in un-trusted clouds. In: Proc. of the 16th European Symposium On Research In Computer Security (ESORICS 2011). Leuven, Belgium (September 2011)