Asian Journal of Information Technology 15 (16): 3103-3107, 2016
ISSN: 1682-3915

Securing Provenance for Efficient Maintenance and Lineage Tracking of Data in Cloud Environment: A Literature Survey

Divya Vadlamudi and K. Thirupathi Rao
Department of Computer Science and Engineering, KL University, Vaddeswarm, A.P, India

Abstract: Cloud computing is an elegant popular approach in dealing with large data sets and to perform huge computations. Therefore, now a days organizations are becoming highly reliant on cloud computing for servicing their data storage and computational requirements. This increased the importance of governing access control at finer levels. This is where it is to be realized that cloud systems have their own significance in maintaining provenance. Provenance is a meta-data of object history which help in maintaining process history and the ownership of the data and therefore ensures verifiability and lineage (pedigree) tracking. Including provenance in cloud systems aid in rebuilding data and providing provenance based access control which in turn ensures secure data storage in cloud. This study focus on identifying various challenges in securing provenance which helps in providing confidentiality for collecting provenance along with unforgebility and thus can provide trusted evidence for scenarios like forensics and hence pushes cloud computing for wide acceptance.

Key words: Cloud computing, provenance, secure provenance, trusted evidence, confidentiality

INTRODUCTION

Cloud computing is a newly extended computing terminology based on utility and resource consumption. It is an elegant popular approach in dealing with large data sets and performing huge computations which includes maintaining groups of remote servers and software defined networks that allow online access to resources and services of computer. NIST defines cloud computing as "a model for enabling ubiquitous, convenient, on demand network access to a shared pool of computing resources that can be provisioned and released numerously with minimal management effort or service provider interaction" (Mell and Grance, 2011).

The cloud model is collection of service models, set of characteristics and deployment models three, five and four, respectively. The five essential characteristics of cloud computing are (Mell and Grance, 2011). On-demand self-service ability to stipulate computing capabilities automatically as required with customer's wish and also involves minimum human interaction with the cloud service provider. Broad network access, capabilities are available over the network and accessed through electronic devices like mobile phones, laptops etc. Resource pooling computing resources of the cloud service provider are shared to produce multiple customers employing a multi-tenant model with totally different physical and virtual resources dynamically allocated and re-allocated in keeping with customer's demand. There is a way of location independence during which the client typically has no data over the precise location of the provided resources however could also be able to specify location at the next level of abstraction. Rapid elasticity capability of being elastically provisioned and revealed with demand. To customer, the capabilities obtainable for provisioning typically seem to be unlimited and may be seized at any time. Measured service cloud systems consequently manage and optimize resource use by investment a metering capability at some level of abstraction relevent to the kind of service. Resources may be in the form of storage, processing, bandwidth and active user accounts

Cloud computing make available to use infinite "virtualized" resources to users as services across the web where it abstracts particulars from the users. With the emergence of business cloud computing platforms there exists 3 service models like IaaS (Infrastructure as a Service), PaaS (Platform as a sService), SaaS (Software as a Service). IaaS is to provision the virtualized hardware on which the client runs its own OS and software pile. In SaaS the OS and the supporting environment is furnished and maintained for the client who can then run and execute their applications. In SaaS, the CSP executes and organizes the entire software setup and provides a

Corresponding Author: Divya Vadlamudi, Department of Computer Science and Engineering, KL University, Vaddeswarm, A.P, India

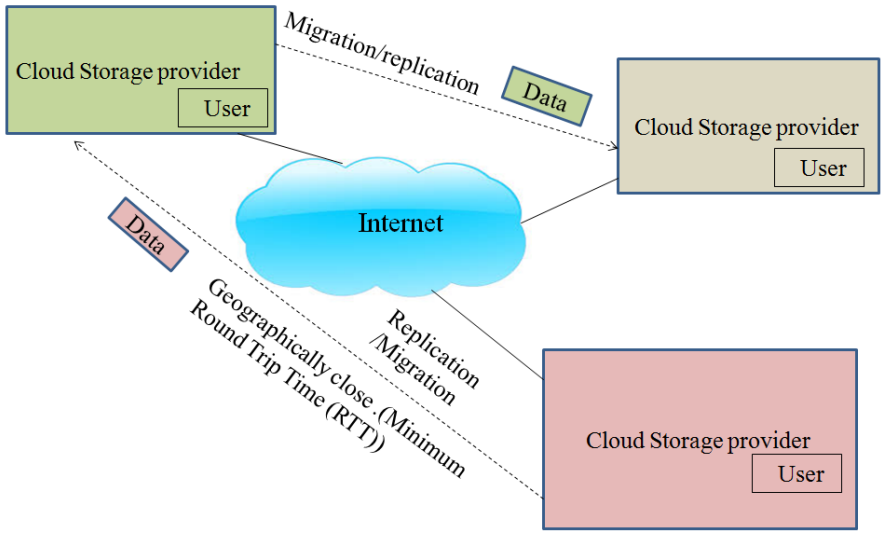

Fig. 1: Simple mechanism of cloud data storage

definite service. Finally there exists four deployment models like private cloud which provides authorized access in an organization, public cloud which is open for public usage and hybrid is an arrangement of two or more cloud servers and community cloud is a mutual setup between many organizations of same community.

Figure 1 illustrates a simple mechanism of cloud data storage which explains the idea under data duplication in cloud in order to ensure high availability of data by Cloud Service Provider (CSP). The CSP maintains duplicate or migrate data to other cloud service provider which found to be nearest based on round trip time and CSP. Many organizations are adopting cloud to outsource their storage and computational needs. Organizations can pass on read and write permissions once storage is leased (So, 2011). The rise in cloud computing exacerbate a brand new front of security challenges in cloud due to lack of efficient data tracking. The incapability to trace the information effectively in cloud environments is changing into one amongst the highest prioritized task for the stake holders. This is due to the following couple of reasons.

The first is short fall of tools on data tracking in clouds. Secondly, current working mechanisms on log generation and maintenance are system centric perspective and that they might not support for the present day's cloud environments.

MATERIALS AND METHODS

Provenance and need of provenance in cloud systems: Provenance furnishes the derivation record of data (Simmhan *et al*., 2005) which acts crucial for rapid adoptability of cloud and helps in enhancing reliability, accountability, transparency and confidentiality of data in the cloud. Provenance is especially crucial within the cloud, because information within the cloud will be shared extensively and anonymously. Provenance is especially crucial in the cloud, as the data in the cloud will be shared extensively and anonymously, without provenance, data consumers have no means to verify validity or identity of data. Hence, provenance can be used for providing security and increase the value of the data in cloud.

Provenance can be outlined as "the place of origin or history of something known in advance as it is a data object stored in the cloud in our scenario". It provides information on the actions taken on data from its genesis onwards. Provenance is a meta-data which helps in verification, audit trails, reproducibility, privacy and security, trust.. Provenance is helpful in answering the questions such as the identifying when an object was created, its main purpose of creation and the details of object origination (Muniswamy *et al*., 2010). In a simple form, we can explain the information like when, why and where clauses on the object creation can be identified. Such data is used for auditing and reconstruction, guaranteeing the reproducibility of data, trust and security, fault identification. Secure provenance is one which maintains log files and Audit files generated by a data item securely which help in reconstructing any data item after a faulty operation which led to unreliable data or loss of data in cloud. This metadata helps in tracing back the creation process of objects including the required functional data. Figure 2 illustrates the scenario raised

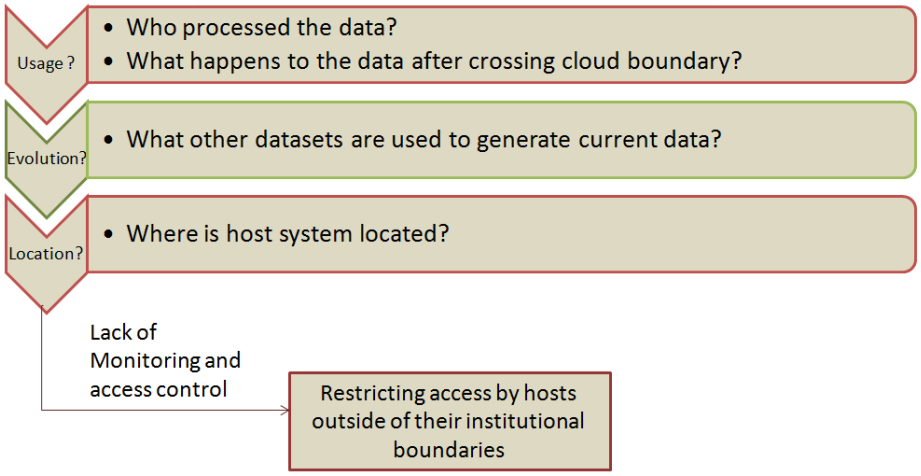

Fig. 2: Lack of access control and data tracking mechanisms

when there is no proper data, provenance tracking and access control mechanisms (Muniswamy and Seltzer, 2010).

This is where we need to understand in keen on the importance of implementing data and provenance tracking, monitoring and access control mechanisms. Data can be reconstructed using meta data called provenance and it also plays a crucial role in ensuring many security concerns and which should be protected from unauthorized access when neglected may cause a serious data loss in the cloud (Simmhan *et al*., 2005; Yoo *et al*., 2011). Hence, there should be a mechanism which ensures secure provenance.

RESULTS AND DISCUSSION

Types of provenance: Provenance Information can be divided into four. They are as follows (Vouk, 2008). Cloud Process provenance: Tracks execution information and control information etc.

- Cloud data provenance: dynamics of file locations, data and data flows etc.
- Cloud workflow provenance: track workflow's evolution and structure
- System provenance: tracks on the complete system information like compiler version, operating system, etc.

Properties of secure provenance: Kiran-Kumar proposed many mechanisms for provenance collection and storage. He described (provenance for cloud) provenance system propertie and their importance:

- Provenance data coupling which indeed explains that object should be coupled with the respective provenance, thus provenance helps in accurately and completely describes the data
- Multi object causal ordering where it conveys the need to maintain the causal relationship between the objects
- Persistence of independent data ensures that he system should be capable enough to retain provenance of an object even if the object was removed
- Efficient query which is very important aspect as the provenance should be comfortably accessible to users who ever desires to access
  and verify the provenance and its corresponding data

Apart from the above properties it is also realized that the properties like Unforgibility and confidentiality are also to be ensured in order to maintain secure provenance (Lu *et al*., 2010; Pichan *et al*., 2015). In the existing provenance system, secure provenance collection and user authentication to reveal the provenance is not ensured well. Research issues like secure provenance collection and surveillance of provenance need to be provisioned to guarantee trust along with safe and secure provenance usage in cloud application. As mentioned provenance collection and storage may require properties (4 properties) specified above but a secure provenance must also ensure the following basic requirements: Unforgeability, is the phenomenon where any adversary

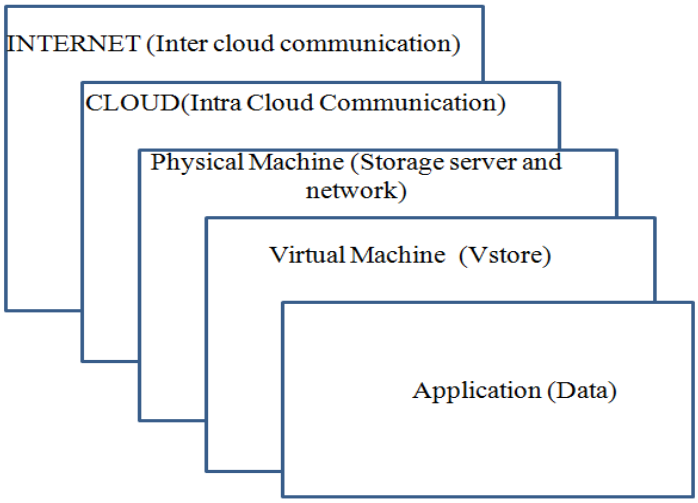

Fig. 3: Granularity of provenance

Table 1: Comparative table on various provenance collection and storage mechanisms

Mechanism	Provenance and data coupling	Maintaining causal relationship between objects	Query	Effect of changes in cloud service or version	Inter Cloud	
					Ensuring unforgebility	Confidentiality
Stand alone Cloud Store (P1) (Muniswamy *et al*., 2010; Zhang *et al*., 2011)	×	√	×	√	×	×
Cloud Store with a cloud database (P2) (Muniswamy *et al*., 2010; Zhang *et al*., 2011)	×	√	√	√	×	×
Cloud store with cloud database and messaging service (P3) (Zhang *et al*., 2011)	√	√	√	√	×	×
Provenance framework	-	-	√	×	×	×
Collecting Provenance via the Xen Hypervisor	√	√	√	√	×	×
Data PROV (Zhang *et al*., 2011)	√	√	√	√	×	√

or antagonist cannot forge a valid provenance record. That is, inability to modify any item in existing provenance record or forging a new record record directly without discovering and thus ensures the attesting the ownership and process history of data stored in the cloud effectively. Conditional privacy preservation, is ensuring information confidentiality at high rates and anonymous authentication in cloud. Any provenance record which is to be genuine requires to be conditional privacy preserving also. That is, only an authorized user has ability and authority to reveal the real distinctiveness or identity recorded in the provenance.

Granularity of provenance: Provenance can be applied at various levels (Zhang *et al*., 2011) like on an virtual machine, a physical machine or on the whole on a cloud. Types of provenance can be derived based on the granularity of provenance. Figure 3 illustrates the granular view of provenance right from application level to Cloud and Internet. The figure explains in keen on various levels where the provenance can be applied.

Existing mechanisms: Earlier some provenance collection and storage mechanisms were proposed and implemented few of them are Provenance Aware Storage System (PASS) (Vadlamudi *et al*., 2015) which collects, preserves, manages automatically and provides provenance search. Provenance collection via Xen hypervisor is a mechanism which collects provenance via a privileged domain (DomO) yet there is no Security concern. Stand alone cloud store (P1) (Muniswamy *et al*., 2010) is a storage scheme which stores object and its provenance as different S3 (Amazon storage) objects. This scheme may not support provenance data coupling, efficient query and security to provenance. The next scheme is Cloud store with a cloud database (P2) which stores file or object in S3 and corresponding provenance in SimpleDB which do not support provenance data coupling and security parameters. Another provenance storage scheme is cloud store with cloud database along with messaging service is same as the Protocol 2 but differs in using cloud messaging service andand transaction in order to ensure provenance data coupling but do not ensure secure data of the source but also its provenance so that we can ensure proper data audit ability and verifiability maintenance. DataPROV (Zhang *et al*., 2011) is another mechanism which overcomes all the above flaws and provides trust in cloud service provider security on cloud

data is essential. Hence, it is recognized that it is so important to encrypt not only multiple cloud service providers which ensures high availability of cloud data to customers.

Open provenance model is a mechanism in which the focus went on designing a model to meet the requirements like allowing provenance information to be exchanged among the systems, to allow tools (developer tools) to operate on provenance and little focus on security of data provenance. Lineage file system also concentrated more on collecting provenance or lineage on file system but did not make sure about the privacy and security of provenance and also for the file system. There are many other approaches which focused on collecting, storing and a little on securing but still a keen focus on provenance.

CONCLUSION

Secure provenance is of predominantly necessary to the flourish of cloud computing. Hence providing confidentiality to the collected provenance along with unforgebility to the provenance is essential in order to provide trusted evidence for scenarios like forensics and thus, can drive cloud computing for wide Consent. Hence, it is realized that a secure provenance collection, storage and authenticity is to be ensured with a novel frame work which is feasible and can assure security and trust for data in Inter cloud data transfer and storage.

REFERENCES

Lu, R., X. Lin, X. Liang and X.S. Shen, 2010. Secure provenance: The essential of bread and butter of data forensics in cloud computing. Proceedings of the 5th ACM Symposium on Information Computer and Communications Security, April 3-16, 2010, ACM, Beijing, China, ISBN: 978-1-60558-936-7, pp: 282-292.

Mell, P. and T. Grance, 2011. The NIST definition of cloud computing recommendations of the national institute of standards and technology. Nist Spec. Publ., 145: 1-7.

Muniswamy, R.K.K. and M. Seltzer, 2010. Provenance as first class cloud data. ACM. SIGOPS. Operating Syst. Rev., 43: 11-16.

Muniswamy, R.K.K., P. Macko and M.I. Seltzer, 2010. Provenance for the Cloud. FAST., 10: 14-15.

Pichan, A., M. Lazarescu and S.T. Soh, 2015. Cloud forensics: Technical challenges, solutions and comparative analysis. Digital Investigation, 13: 38-57.

Rao, B.T., 2016. A study on data storage security issues in cloud computing. Procedia Comput. Sci., 92: 128-135.

Simmhan, Y.L., B. Plale and D. Gannon, 2005. A survey of data provenance techniques. MCS Thesis, Computer Science Department, Indl iana University, Bloomington, Indiana.

So, K., 2011. Cloud computing security issues and challenges. Int. J. Comput. Networks, 3: 1-9.

Vadlamudi, D., M.K. Chaitanya, T. Srikanth, V.B. Venu and U. Joseph et al., 2015. An applicative approach for collecting and fortifying history of data in cloud environment. Int. J. Software Eng. Appl., 9: 11-20.

Vouk, M.A., 2008. Cloud computing-issues, research and implementations. J. Comput. Inform. Technol., 16: 235-246.

Yoo, C.S., 2011. Cloud computing: Architectural and policy implications. Rev. Ind. Organiz., 38: 405-421.

Zhang, O.Q., M. Kirchberg, R.K. Ko and B.S. Lee, 2011. How to track your data: The case for cloud computing provenance. Proceedings of the 2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom), November 29-December 1, 2011, IEEE, New York, USA., ISBN: 978-1-4673-0090-2, pp: 446-453.