A survey on big data analytics for enhanced security on cloud

Article *in* International Journal of Engineering & Technology · April 2018

DOI: 10.14419/ijet.v7i2.21.12397

4 authors, including:

Phani Kumar
PACE Institute of Technology and Science
6 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Kalaivani K
Vels Institute of Science Technology and Advanced Studies
26 PUBLICATIONS **101** CITATIONS
SEE PROFILE

International Journal of Engineering & Technology, 7 (2.21) (2018) 331-334

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A survey on big data analytics for enhanced security on cloud

R. Anandan[1*], S. Phani Kumar[2], K. Kalaivani[3], P. Swaminathan[4]

[1]*Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies(VISTAS), Chennai, India.*
[2]*Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies(VISTAS), Chennai, India.*
[3]*Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies(VISTAS), Chennai, India.*
[4]*Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies(VISTAS), Chennai, India.*
**Corresponding author E-mail:anandan.se@velsuniv.ac.in*

Abstract

Cloud based data storage has become a common activity these days. Because cloud storage offers more advantages than normal storage methods those are dynamic access and unlimited storage capabilities for pay and use. But the security of the data outsourced to the cloud is still challenging. The data owner should be capable of performing integrity verification as well as to perform data dynamics of his data stored in the cloud server. Various approaches like cryptographic techniques, proxy based solutions, code based analysis, homomorphic approaches and challenge response algorithms have been proposed. This survey depicts the limitations of the existing approaches and the requirements for a novel and enhanced approach that ensures integrity of the data stored in cloud enabling better performance with reduced complexity.

Keywords: *Big data, cloud computing, security, storage in cloud.*

1. Introduction

Clouds computing is an internet based computing where resources and data are shared to computers and other devices on a large-scale. Cloud can be provisioned on-demand and rapidly with minimum management. It has an impact on various users that include individual customers, start-ups, SMEs, governments, businesses and enterprises. Cloud computing has especially transformed information technology in a way the resources are delivered and consumed. This is due to high performance, high computing power, and low-cost of service, easy accessibility, scalability and high availability.

Cloud storage is the most widely adopted and the fastest growing service for storing, archiving, sharing, and backup and synchronization of multiple devices. Storage services provides highly available, user-friendly and ubiquitous access to data, inexpensive service with no or minimal maintenance, immediate scaling and pay-per-usage service. Data is accumulated in logical groups that extents numerous servers and locations and delivered over the internet based on a request for a given service level. Various types of cloud storage services are offered by the cloud storage providers: Basic cloud storage service: They are accessible using application program interfaces (APIs). Examples are AmazonS3, Rack space, etc. Advanced cloud storage service: Along with the basic storage service it offers interfaces such as client or web applications and API integration of services into third-party software. Examples like Mozy, Dropbox, etc.

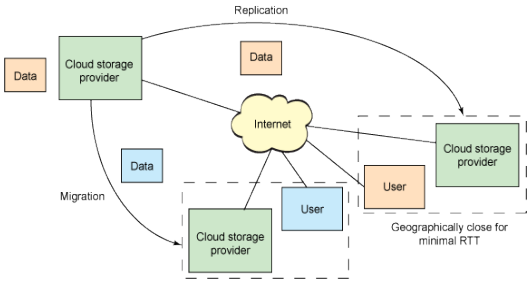

Figure 1: Cloud

2. Applications of cloud storage

Cloud storage is used by individuals, governments, SMEs, organizations and especially IT related enterprises. From an organization perspective the applications include the following:

Backup: Using cloud storage organization can back up business critical data to the cloud that can be made available at short notice and restored to specific time in the past at lowered cost.

Archive: Large volumes of data generated by organizations that is forced to retain a copy can use the cloud at a lower price and higher capacity.

Application data storage: Applications especially business critical applications require temporary and permanent data storage. Cloud storages is used to store the data at a reduced cost and is made easily and readily accessible.

3. State-of-art

D. Bernstein et al.[1], proposed a cloud security show that enables the clients to perform reviewing on the distributed storage. Subsequent to inspecting, the outcomes are utilized to check the rightness of capacity.

What's more, it additionally simultaneously decreases information blunder inside the server. Their outline likewise underpins sheltered and compelling dynamic information control operations. They have accomplished the information respectability and additionally information accessibility in distributed storage. Their work decreases the different assaults, for example, information adjustment server, plotting and get to vitiation kinds of assaults.

R. Dua et al.[2], suggested a conveyed tenable information stockpiling framework as well as the recovery procedure into cloud database utilizing the re encryption systems. Their plan bolsters the encryption and sending methods with incorporated encryption procedure, encoding data and its sending. There are numerous works that discuss concerning cloud information stockpiling and recovery strategies. A. Gholami et al.[4], proposed a secured information stockpiling procedure which suggests for open audit ability of cloud database framework stockpiling with the goal that the client can utilize an outsider examining component to keep up information trustworthiness.

S. Pearson et al.[6] proposed a cloud-based capacity plot which underpins outsourcing of dynamic information, where the proprietor is equipped for not just documenting and getting to the information put away by the cloud specialist co-op. Their plan empowers the approved clients to guarantee that they are accepting the latest variant of the outsourced information.

F. Liu et al.[7] proposed a property based information sharing plan to implement a fine-grained information get to control by misusing the normal for the information sharing framework. F. Liu et al.[7] composed a proficient blending based PPDP convention for giving security in information correspondences on cloud.

Demchenko et al. [8] proposed another convention for security safeguarding review for cloud information stockpiling. Despite the fact that this model decreases a lot of cost and gives information security, it sets aside a ton of time for progressively inspecting the information put away in the cloud database.

N. Santos et al. [9] proposed an open examining plan for recovering code-based distributed storage framework. In their model, they gave different levels of benefits in light of clients.

S. Zhu et al.[10] , proposed another strategy to clarify the way toward offering information to different clients in a protected, proficient and adaptable way in cloud information stockpiling. They have actualized an open key cryptosystem to deliver same size information figure messages so a compelling assignment of unscrambling rights for figure writings is conceivable on client ask. Utilizing the cryptosystems which influences utilization of cryptography to total key system, they have demonstrated the strategy to pack the mystery keys. This stores information in a protected way. By this approach, the various levelled key administration process can just lessen spaces for the situation that all clients having the key offer a similar arrangement of benefits. This framework has the confinement of use of the model to just constrained figure writings classes.

K. W. Hamlen et al. [11] proposed another limit and quality based encryption conspire called for furnishing viable capacity openly cloud with get to control. Their exploratory outcomes demonstrate that their plan is strong and secure.

S. Fischer-Hubneret al. [12] proposes an automated approach to selecting cloud storage providers that meets the user's requirement. Their approach relies on multiple aspects such as specific features, performance and cost described using XML schema. The result recommends the list of storage providers, gives the estimates of storage cost and performance, cost saving and migration recommendations. (Chang et al. 2012) presents a mathematical model to address the issue of selecting cloud storage providers to maximize the benefits with a given budget. Two algorithms are proposed: One algorithm to find the maximum failure probability with given budget and the other to maximum validness with given budget.

Y. Huang et al. [13] presents a methodology to study cloud storage services of five storage providers namely Dropbox, Microsoft SkyDrive, Google Drive, LaCiaWuala and Amazon Cloud Drive. An investigation of the providers using file synchronization capability, differences on client software and placement of data centre and the consequences of the design on performance is presented. Y. Huang proposes an automated selection of cloud infrastructure. The criteria considered include deployment cost, infrastructure location, application clients and interaction among applications to select the cloud providers. To handle the uncertainties in the customer's subjective preferences, Y. Huang proposed a fuzzy inference to express the fuzziness or vagueness. To promote the truth-telling, a game theory based selection strategy is proposed for the selection of cloud storage services. The criteria taken for evaluation are QoS attributes and cost. However, only one customer and one service can be evaluated at a time. And thus it is time consuming and inefficient.

D. Bernstein. [1] presents "Trust as a facilitator" that discusses the advantages and barriers for adopting cloud computing from the perspective of a cloud service customer. A unified trust-aided framework leveraging trust and reputation to predict the future behaviour of the cloud service providers is presented.

J. Cappos et al. [15] introduce a credibility based trust management in the cloud environment. The credibility model detects malicious (Self promoting and slandering attack) and non-malicious feedbacks and also can distinguish feedback from experience and armature customers. Majority consensus and feedback density are used as trust parameters to find malicious feedbacks from the online reputation systems.

M. Shankarwar et al. [16] presents a detail survey of trust management of cloud service providers that are classifies into four different trust management categories namely Reputation, Recommendation, Policy and Prediction. An analytical framework for cloud trust management is presented that includes three layers namely trust feedback, assessment and distribution layer to compare trust management prototypes based on a number of attribute or criteria.

J. G. Politz et.al [17] argues that the verification of the trust worthiness of the cloud service providers must be produced with evidences without getting into the details of the infrastructure's complexity and dynamism. A framework is presented with two levels and uses remote attestation. Assurance that the provider behaves in a way as expected by the customer using chain of trust between the users the providers is used to select trustworthy service providers. Policy based trust that is derived from formal auditing and attribute based trust is used as evidences to evaluate trust

J. G. Politz et.al [17].For attribute trust the competency, goodwill and integrity of the sources of trust against the attributes: security, privacy and performance. The sources of the cloud attribute can be user's and cloud brokers' observation, opinion of peers, data from social networks, attribute certification, and statements from the providers or the assessments of the cloud service auditors. A trust evaluation system based on building trust relationships with the cloud service providers using fuzzy logic's sugeno fuzzy-inference. J.G. Politz proposed. Several dimensions are considered as annotation of trust that includes availability, scalability, security and usability. QoS values have been used in literature to objectively assess the cloud service providers. It is not possible to obtain all the QoS values as the collecting them is time consuming, The cloud service providers give different representation of the same parameter and it is an expensive affair.

A, Katalet.al [18] focuses on predicting the QoS values that are missing during evaluation as it is not possible to have full assessment dataset. They also combine the predicted QoS values with subjective trust based on estimation of customer satisfaction. A personalized evaluation framework for cloud service trustworthiness is presented based on utility theory and collaborative filtering recommendation.

K, Chitharanjan et.al [19] present a decision support for personalized provider selection by focusing on the customers' expectation and preferences in a multi-attribute trustworthy based evaluation. An item based collaborative filtering approach is employed for missing values prediction and Pearson correlation

coefficient is employed to find similarities between the cloud providers. For trust aware service selection firstly the utility of the attribute is determined to identify a provider's trustworthiness. Secondly, the customers satisfaction is estimated and aggregated using a customer-service matrix and decision on the trustworthy cloud service provider is estimated.

P.R Anisha et al. [20] proposes a multi-dimensional trust evaluation using ratings and users history. Evidential reasoning aggregates both the reputation and perception based trust to select trustworthy cloud service provider. Three key requirements are provided to reduce uncertainty and thereby increase trustworthiness. The requirements are categorized as key operational performance requirements (adaptability, scalability, flexibility, resilience, continuity), key QoS requirements (Availability, usability, reliability, quick response, consistency) and key security and privacy requirements (transparency, audit ability, access and accuracy, controllability, accountability).

P.R Anisha et.al [20] presents a framework that makes use of two parameters namely trustworthiness and competence. Trustworthiness of the service provider is estimated from the feedback on the service providers or from direct interaction. Competence is computed based on the service level agreement guarantees. The risk due to interacting with the service provider is estimated using utility and importance to the trustworthiness and competence. A relationship between the trustworthiness, competence of the service provider and the perceived risk is given.

4. Cryptographic mechanisms

Cryptography is gotten from Greek word. It has 2 sections: 'crypto' signifies "shrouded, mystery" and 'graphy' signifies "composing". It is an investigation of systems for secure correspondence within the sight of outsiders to keep up data securities, for example, information respectability, privacy, verification, and non-disavowal. It is a craftsmanship to change the messages to influence them to secure and resistant against security assaults. The specialty of ensuring data by changing into an incomprehensible arrangement, called figure content or then again decode the message into plain content. The figure content is just comprehended by somebody who just knows how to unscramble it. The data is scrambled utilizing an encryption calculation, which indicates how the message is to be encoded. Any interloper that can see the figure content ought not have the capacity to decide about the first message. Just an approved gathering can translate the figure content which requires a mystery decoding key.

Sorts of cryptography

There are two sorts of cryptography:

1. Secret key cryptography or Symmetric-key cryptography: In SKC, the sender and the beneficiary know a similar mystery code, which is known as key. With a similar key message are scrambled by the sender and unscrambled by the collector. It can be of 2 types: Stream Buffer, Block Buffer. Stream Buffer: Stream cushion encodes the digits of a message each one in turn. Stream Cipher capacities are utilized on a flood of information one at time by working on it by bits. It comprises of two segments: 1) a key stream generator and 2) blending capacity. Blending capacity utilizes XOR capacity, and key stream generator is unit in stream encryption algorithm. Block figure: In Block figure, it takes various bits and after that scramble them as a solitary unit. Information is encoded/decoded if information is in the types of pieces. In straightforward words, the plain content is separated into pieces which are utilized to create pieces of figure content cushioning the plaintext in squares. 64 bits squares have been normally utilized.
2. Public key cryptography or Asymmetric-key cryptography: Asymmetric key (or open key) encryption is utilized to explain the issue of key conveyance. In PKC, two keys are utilized; private keys and open keys. For encryption open key is utilized and for unscrambling private key is utilized. Open key is known to open and private key is known to the client.

Different cryptographic algorithms

Data encryption standard

It was planned in 1970's by IBM and was confirmed in 1977 by the National Bureau of Standards (NBS) for business utilize. It is a piece figure that works on 64-bit squares utilizing a 56-bit key and 8 rounds. In spite of the fact that DES has been around long back yet no genuine shortcoming has been recognized. The greatest drawback of DES is the 56 bit key estimate.

Advanced encryption standard

It was outlined by Vincent Rijmen and Joan Daemen and was presented in 1998. The calculation can utilize flighty key length and square length. The key length can incorporate 128, 192, or 256 bits and square length can be of 128, 192, or 256 bits.AES is an exceedingly productive and secure calculation. The disadvantage lies in its handling as it requires more handling.

Rivest cipher

Ronald Rivest built up this calculation and in this way, the name of the calculation was put after Ronald's Rivest name. It gives a progression of RC calculations including RC1, RC2, RC3, RC4, RC5 and RC6.

Blowfish

It was created by Bruce Schneie and was first distributed in the year 1993. This piece figure has 8 rounds, having the piece measure is of 64 bits and the key length can differ from 32 to 448 bits. Blowfish was proposed as a substitute was DES. This calculation is fundamentally speedier than other calculations and the key quality is brilliant. Blowfish calculation is able just for applications where the key for the most part continues as before.

The Public Key Cryptography (PKC) utilizes one (open) key for encryption and another (private) key for decoding. The PKC calculations that are being used today are:

RSA

RSA remains for Ron Rivest, Adi Shamir and Leonard Adleman. RSA was named after the mathematicians who created it. RSA was first distributed in 1997. Variable measure key and encryption square is utilized as a part of RSA. Principle favorable position of RSA calculation is upgraded security and comfort. Utilizing Public Key Encryption is likewise an favorable position of this calculation. RSA needs in encryption speed.

Diffie-Hellman

This calculation was presented in1976 by Diffie-Hellman. The Diffie-Hellman calculation stipends two clients to set up a shared mystery key and to convey over a shaky correspondence channel. One way validation is free with this kind of calculation. The greatest confinement of this sort of calculation is correspondence made utilizing this calculation is itself helpless against man in the center assault.

These algorithms are efficiently used for encrypting the data at user side for providing security and then it will be sent to cloud storage. If a user want to receive data then decrypt method is performed and sent to receiver.

Comparison Table

Parameters	AES	DES	Rivest Cipher	Blowfish	RSA
Key Length	128,192, 256	64 (56 Usable)	0 To 2040 Bits Key Size (128 Suggested)	Variable Key Length I.E. 32-448	Key Length Depends On No. Of Bits In The Module
Rounds	10,12,14	16	1 To 255(64 Suggested)	16	1
Block Size (Bits)	18	64	34,64,128(64 Suggested)	64	Variable Block Size
Attacks Found	Key Recovery Attack, Side Channel Attack	Exclusive Key Search, Linear Cryptanalysis, Differential Analysis	Co-Relation Attack, Timing Attack	No Attack Is Found To Be Successful Against Blowfish	Brute Force Attack, Timing Attack
Level Of Security	Excellent Security	Adequate Security	Secure	Highly Secure	Good Level Of Security
Encryption Speed	Faster	Very Slow	Slow	Very Fast	Average

5. Conclusion

Distributed storage has certainly advanced giving an adequate degree to a few utilize cases to cost viably use and understand their goal utilizing the capacity administrations which we have endeavoured to overview in this paper. While there are issues of non-consistency crosswise over cloud merchants, there is a prerequisite to give uniform UIs and consistent coordination with the standard work area and server registering. Also, since a cloud framework is a disseminated framework, storerooms might be outlined like the appropriated record framework. This will make ready for clear and simple incorporation with current programming application. In a comparable way, key-esteem stockpiling needs to help improved highlights like help for exchanges crosswise over various gatherings of substances. Google DataStore supports custom record for elements, Windows Azure Table and SimpleDB of Amazon need it. Square based stockpiling like Azure Drive must help simultaneousness for both read and compose mode get to. We trust that such deficiencies will soon be finished and another type of cutting edge storerooms will rise. In this paper we introduces the condition of-workmanship for the distinctive distributed storage instruments and delivers the points of interest and disadvantages of all the mechanisms.

References

[1] Bernstein D, "Containers and Cloud: From LXC to Docker to Kubernetes", *IEEE Cloud Computing*, Vol.1, No.3, (2014), pp.81-84.

[2] Dua R, Raja A & Kakadia D, "Virtualization vs containerization to support paas", *IEEE International Conference on Cloud Engineering (IC2E)*, (2014), pp.610–614.

[3] Tholeti BP, "Hypervisors, virtualization, and the cloud: Learn about hypervisors, system virtualization, and how it works in a cloud environment", *IBM developerWorks*, (2011).

[4] Gholami A, Laure E, Somogyi P, Spjuth O, Salman N & Dowling J, "Privacy-preservation for publishing sample availability data with personal identifiers", *Journal of Medical and Bioengineering*, Vol.4-2, (2014), pp.117–125.

[5] Gholami A, Lind AS, Reiche J, Litton JE, Edlund A & Laure E, "Design and implementation of the advanced cloud privacy threat modeling", *International Journal of Network Security & Its Applications (IJNSA)*, (2016).

[6] Pearson S, "Privacy, security and trust in cloud computing" *Privacy and Security for Cloud Computing*, (2013), pp.3–42.

[7] Liu F, Tong J, Mao J, Bohn R, Messina J, Badger L & Leaf D, *NIST Cloud Computing Reference Architecture: Recommendations of the National Institute of Standards and Technology (Special Publication 500-292)*, USA: CreateSpace Independent Publishing Platform, (2012).

[8] Demchenko Y, Membrey P, Grosso P & de Laat C, "Addressing big data issues in scientific data infrastructure", *International Conference on Collaboration Technologies and Systems (CTS)*, (2013), pp.48-55.

[9] Santos N, Gummadi KP & Rodrigues R, "Towards trusted cloud computing", *Proceedings of the Conference on Hot Topics in Cloud Computing*, (2009).

[10] Zhu S & Gong G, "Fuzzy authorization for cloud storage", *IEEE Transactions on Cloud Computing*, Vol.2, (2014), pp.422–435.

[11] Hamlen KW, Kagal L & Kantarcioglu M, "Policy enforcement framework for cloud data management", *IEEE Data Eng. Bull.*, Vol.35, No.4, (2012), pp.39-45.

[12] Fischer-Hubner S, Angulo J & Pulls T, "How can cloud users be supported in deciding on, tracking and controlling how their data are used?.", *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, (2013), pp.77-92.

[13] Huang Y & Goldberg I, "Outsourced private information retrieval", *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, (2013), pp.119–130.

[14] Microsoft, "Data and insights" http://www.microsoft.com/enterprise/it-trends/big-data/, 2015. Accessed July 2015.

[15] Cappos J & Torres S, "PolyPasswordHasher: Protecting Passwords In The Event Of A Password File Disclosure", *Technical report*, (2014).

[16] Shankarwar M & Pawar A, "Security and privacy in cloud computing: A survey", *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*, (2014), pp.1–11.

[17] Politz JG, Guha A & Krishnamurthi S, "Typed-based verification of Web sandboxes", *Journal of Computer Security*, Vol.22, No.4, (2014), pp.511–565.

[18] Katal A, Wazid M & Goudar RH, "Big data: Issues, challenges, tools and Good practices", *Sixth International Conference on Contemporary Computing (IC3)*, (2013), pp.404-409.

[19] Chitharanjan K & Kala Karun A, "A review on hadoop-HDFS infrastructure extensions", *IEEE Conference on Information & Communication Technologies (ICT)*, (2013), pp.132-137.

[20] Anisha PR, Kishor Kumar Reddy C, Srinivasulu Reddy K & Surender Reddy S, "Third Party Data Protection Applied To Cloud and Xacml Implementation in the Hadoop Environment With Sparql", *Journal of Computer Engineering (IOSRJCE)*, Vol.2, No.1, (2012), pp.39-46.