IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Efficient Map/Reduce secure data using Multiagent System

H. Kousar[1*], B.R.P. Babu[2]

[1*]Dept. of CSE, East Point College of Engineering and Technology, Bangalore, India
[2] Dept. of CSE, V.T.U, Belgaum, India

[*]*Corresponding Author: hkheenakousar73@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 22/May/2018, Published: 21/May/2018

***Abstract*—** Today is an era of digital world where technology and information play a very important role in our lives. Huge amount of data is generated, more than in terms of zetabytes and this data is been transmitted over an internet. Data is from Internet search, mobile devices, Internet of Things which include variety of objects like tablet, sensors, smart phones, digital cameras, tablets, etc., business transactions, Government organization, next generation radio astronomy telescopes, traffic data etc. This result in the raise of big data application. Current technologies i.e. Hadoop is unable to process data and secure the data within a given elapsed time. In this paper we proposed Multiagent Map Reduce Secure Model (MMRSM) using cloud. MMRSM supports both batch processing of data and Streaming of data. The proposed system supports both data security and efficient resource management of processing streaming data with minimum cost and time.

***Keywords*—** Multiagent, Hadoop, Internet of Things, Streaming of data Computing

I. INTRODUCTION

Cloud Computing [1]. a model for access on-demand to configurable shared computing resources such as storage, servers, applications, software, services and networks which can be provisioned easily as needed. To gain efficient utilization of resources cloud computing provides an aggregate interfaces and allows scaling to solve huge problems. Based on the application requirement the software can be configured as needed. Cloud computing provides best of technology to access high reliable, storage and high performance cluster without need to purchase and maintain hardware. It also provides an environment for scientific applications to analyze the climate data, social media data, business applications while lowering operation cost and conserving energy. Cloud computing not only reduces the cost but also smaller companies can make use of wide array of applications. However there are many challenges to handle large volumes of historical and streaming sensor data which ranges from structured to unstructured, with respect to performance, cost and collaborative work environments.

An agent is an autonomous entity, number of agents form Multiagent system which interact with each other by exchanging messages across a network. Through cooperating, coordinating and negotiating the agents interact with each other in order to achieve the design objectives. Normal cloud users served by agents for cloud service provider service always negotiate to help other agents to complete the tasks [5].

Big Data [2][3] is a collection of large complex datasets(structures and unstructured) which is difficult to process in the available database management system. Big Data is originated from different technologies such as internet, computer networks, sensors, social media and cluster computing [4]. Big Data is characterized by 5Vs i.e. volume, velocity, variety, veracity and value. Volume: Growing explosively of data beyond the capability of handling large data sets. Velocity: Fast transmission and generation of data across the internet as exemplified by data collections form massive sensors from micro to macro and data transmission of sensor data to super computers and processing of data. Variety: Data in diverse form in which structural data and model are archived. Veracity: refers to accuracy, trustworthiness and diversity of quality. Value: refers to decision support and specific research applications which improve the work, lives and prosperity. The challenges with Big data are to search, visualize, share, capture, storing the data, analyzing and processing the data. Analysis of data benefitted in various domains such as environmental studies, medical, social networking, science, astronomy, telecommunication, IT business etc. IOT generates huge amount of data, Environment sensors monitors contents of humidity, gases present in air and accordingly generates the data. Mobile apps generate vast amount of data in the form of video, text, images, mails etc.

Streaming data is nothing but the data that is continuously generated by hundreds of thousands of data sources, includes

wide variety of data such as log files generated by the customers using web applications or mobile, e-commerce purchases, in-game player activity, geospatial services, telemetry from connected devices or instrumentation in data centres. This data has to be processed sequentially and incrementally on a record-by-record basis or over sliding time windows and used for a wide variety of analytics using correlations, sampling, filtering , aggregations. The information derived from this analysis gives companies visibility into many aspects of business and customer activity such as server activity, service usage(for billing/metering), geo-location of devices, people and physical goods and respond promptly to emerging situations

From the online properties billions of click stream recorded from the media publishers, aggregates and enriches the data with demographic information about users and optimizes content placement on its site, delivery relevancy and better experience to its audience. An online gaming company collects streaming data about player -game interactions and feeds the data to its gaming platforms. it then analyzes the data in real time, offers incentives and dynamic experience to engage its players.

Section I contains the introduction of cloud computing, Big Data, Multiagent system , Section II contain the related work of secure Map Reduce Multiagent system, Section III contain the Architecture of Secure Multiagent Map Reduce, Section IV contain the Results of Comparison between RSA & AES algorithm with and without Multiagent system, section V concludes research work with future directions.

II. RELATED WORK

For storing the data and query processing in cloud N. Cao, C. Wang [12] proposed privacy preserving search system i.e Prism for cloud computing. It provides parallelism features, Map Reduce efficiency. No need to modify underlying system for the usage of Prism. It does not incur overhead. Drawback of Prism is not suitable for providing security to public cloud.

Airavat, proposed by Roy, S. Setty, A. Kilzer, V. Shmatikov, and E. Witchel[13]. is a Map Reduce based system which provides privacy and strong security for sensitive data for distributed computations. It is an integration of differential privacy and mandatory access control. It prevents leakage of data. The disadvantage of this it is not suitable for providing security to the streaming data.

Sedic proposed by K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan[14]. partitions the data and move the sensitive data on the private cloud and sanitized data to the public cloud. It is implemented on Hadoop to evaluate both synthesized and real computing jobs on the cloud. Sedic protects privacy data in Map Reduce operations. It extracts the data automatically from the Combiner's from Reduce function to process the data in the public cloud.

Philip Derbekoa , Shlomi Dolevb, Ehud Gudesb , Shantanu Sharma[15]. proposed the survey different security and privacy challenges and requirements considering variety of adversarial characteristics and capabilities of Map Reduce. It also provides the review of existing privacy protocols and security for Map Reduce and their overhead issues.

Upper bound privacy leakage constraint based approach proposed by X. Zhang, C. Liu, S. Nepal, S. Pandey, and J. Chen,[16] identifies the intermediate data sets to be encrypted or not. The cost of preserving privacy cost of intermediate data sets can be reduced significantly. The disadvantage of this method is more complicated, processing of efficient data is quite challenging.

Adilah Sabtu, Nurulhuda Firdaus Mohd Azmi, and Siti Sophiayati Yuhaniz [17]. proposed security to the big data and privacy protection of unstructured data. They provide securtiy and privacy protection access control for the processing of Map Reduce by using Whitelist

Comparison between Batch processing and Stream Processing

Batch processing is to used to compute arbitrary queries over different sets of data, which usually compute results that are derived from all the data it encompasses and enables deep analysis of large data sets. Map Reduce -based systems like Amazon EMR are examples of platforms that supports batch jobs. Stream processing requires ingesting a sequence of data and updating metrics incrementally, reports and summary statistics in response to each arriving data record. It is best suited for real time monitoring and response functions. Many organizations are building a hybrid model by combining the 2 approaches and maintain a real time layer and the batch layer.

	Batch Processing	Stream Processing
Data Scope	Queries or processing over all or most of the data in the data set	Queries or processing over data within a rolling time window or on just most recent data record.
Data Size	Large batches of data	Individual records or micro batches consisting of a few records
Performance	Latencies in minutes to hours	Requires latency in the order of seconds or milliseconds.
Analyses	Complex Analytics	Simple response functions, aggregates and rolling metrics.

III. METHODOLOGY

A. Architecture of Multiagent System

Figure 1. shows Multi-Agent System (MAS) architecture supports Map/Reduce streaming and batch data to facilitate confidentiality, correctness assurance, availability and integrity of users' data in the cloud. Our security framework consists of two main layers as agent layer and CDS layer. Security to data is provided by means of Multiagent system consists of 4 types of agents i,e Data Confidentiality Agent(DConA), Data Correctness agent(DCorA), Data Availability Agent(DAA), Data Integrity Agent(DIA

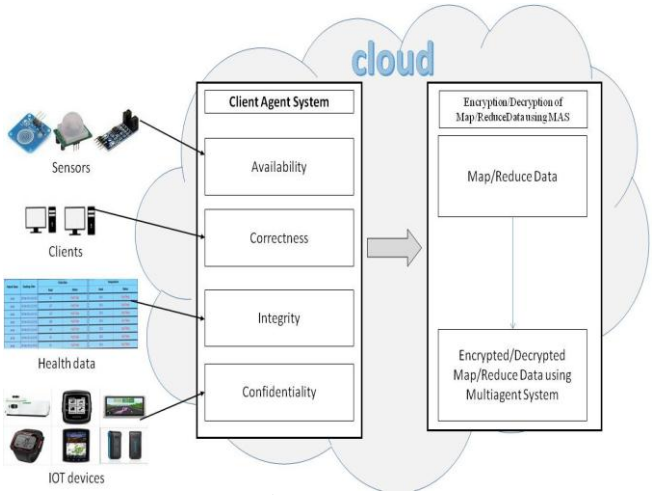

Figure 1. Architecture of Multiagent Map/Reduce Secure data.

In distributed processing of large scale data in fault tolerant and efficient manner on private, public or hybrid cloud a special type of programming called MapReduce is used. To solve large scale of problems in a distributed computation an efficient MapReduce is used. e.g. different types of join operations, pattern matching, clustering, log analysis, matrix multiplication and analysis of social networks. The most important essential concern with respect to MapReduce is security and privacy of data when a computation is executed on the cloud. To execute MapReduce job in public or hybrid cloud confidentiality of data computations, authentication of mappers-reducers, correctness and integrity of data are required.

The data obtained after processing Map Reduce computation must be secured. The security to Mapreduce data can be provided by applying the best cryptographic algorithm to the data. There are different types of encryption algorithm in order to provide security. The two important encryption algorithms is Symmetric & Asymmetric. The Algorithm included in the study is AES & RSA. The performance analysis are calculated with Multiagent and without Multiagent.

B.Analysis of RSA & AES algorithm using Multiagent System

One of the best public key cryptography is RSA[6]. It is named after the scientist. It is the algorithm which is suitable for both encryption as well as signing. RSA uses 2 keys i.e. private and public key. Messages encrypted with public key are decrypted with the private key. Because of large size of both stream and non-stream data there exists a bottleneck .The data size can be further reduced and also ensures security after applying Map Reduce operation by using Multiagent system. Multiagent system[7] reduces the bottle neck by executing the same RSA or DES algorithm in parallel.

RSA or AES algorithm divides the ciphertext or plaintext into several packets with the same length, then encryption and decryption are done with multiple- agents running the RSA or AES algorithm for the elements or packets assigned to it. The question is how agents will know which element for that an additional index field is assigned to control offset of an element. In this way using Multiagent system increases the speed of execution time of both RSA & AES.Multiagent system are simulated using virtual machines. Here agents are created virtually based on the size of the file and deleted as of not required. In this paper we analyzed using 4 VM. Hence there will be no storage space require as such. Cost incurred also will be less.

c. Implementation

Various work is carried to [8] to proposed a homomorphic encryption for securing data in cloud, but consumes more time. Different security algorithms proposed in [9] to solve various concerned on security issues on cloud computing environment. In [10] various problems faced by provider of cloud and solution to overcome this problem is given. Microsoft Windows Azure[11] used as a platform for an application development. It provides emulator computing to run, debug, test and check the performance of code in cloud. Human effort and resources are saved and speed of application enhanced. Java and windows azure SDK is used to design the application to test the different parameters of RSA and AES for comparison.

IV. RESULTS AND DISCUSSION

Different values are taken by running an application on cloud environment. Different reading of encryption and decryption with and without Multiagent system with common key size and varying file size. Table 1.shows the encryption and decryption with and without Multiagent system using RSA

Table 1. RSA: Encryption & Decryption time with and without Multiagent System

Key Size(bits)	File Size(MB)	Encryption Time(ms)		Decryption Time(ms)	
		Without MAS	*WithMAS*	*Without MAS*	*With MAS*
512	200	210	110	205	105
512	400	415	250	435	248
512	650	550	325	568	415
512	880	750	455	775	556

Graph in Figure 2 & 3 is plotted between File size and Encryption and decryption time using RSA algorithm with and without Multiagent system. It shows that with different file size the computation time and power increased as varying the file size.

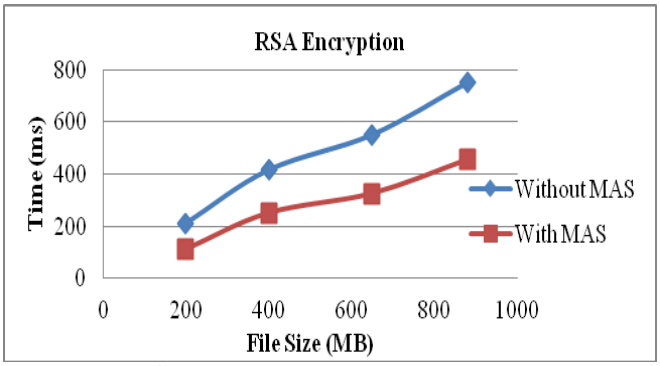

Figure 2. Graph of RSA Encryption comparison with and without MAS.

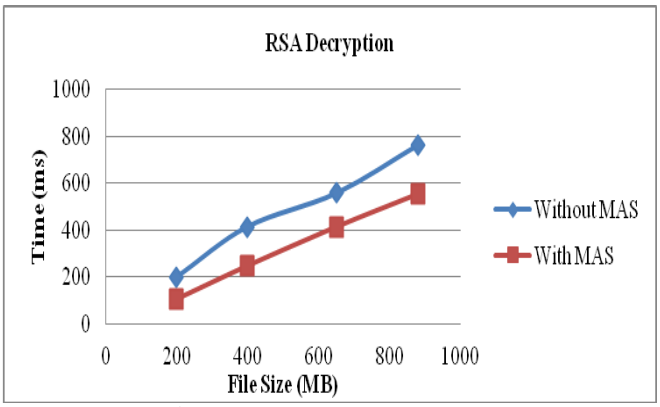

Figure 3. Graph of RSA Decryption comparison with and without MAS.

Table 2.shows the encryption and decryption with and without Multiagent system using AES algorithm.

Table 2. AES: Encryption & Decryption time with and without Multiagent System

Key Size(bits)	File Size(MB)	Encryption Time(ms)		Decryption Time(ms)	
		Without MAS	*WithMAS*	*Without MAS*	*With MAS*
256	200	150	80	165	98
256	400	350	185	385	214
256	650	450	279	492	291
256	880	650	455	692	499

Graph in Figure 4 & 5 is plotted between File size and Encryption and decryption time using AES algorithm with and without Multiagent system. It shows that with different file size the computation time and power increased as varying the file size.

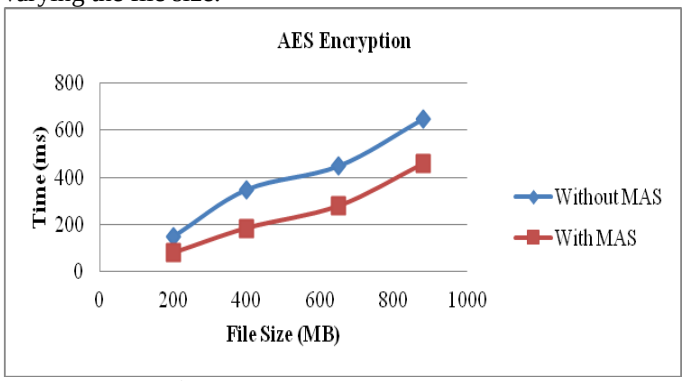

Figure 4. Graph of AES Encryption comparison with and without MAS.

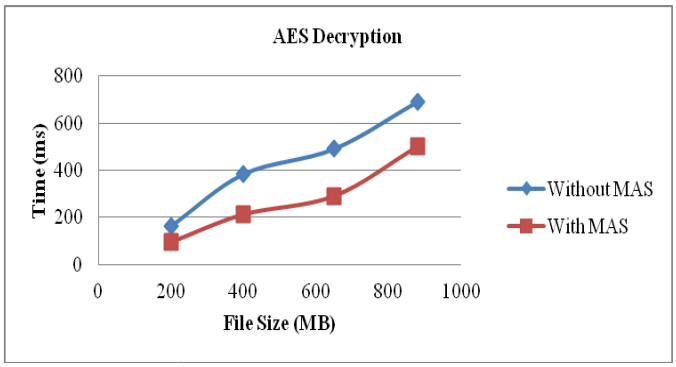

Figure 5. Graph of AES Decryption comparison with and without MAS

V. CONCLUSION AND FUTURE SCOPE

In this paper we secured the Map reduce data by analyzing with the RSA and AES with 128 bit algorithm by varying the file size. We also analyzed that AES is best compared to RSA as the file size increased there is a reduction of time and cost. AES is more efficient. The further work can be carried

out by extending the size of the key from 128 bit to 198 or 256 bit so the data will be more secured.

REFERENCES.

[1] Gartner cloud computing definition, http://www.gartner.com/itglossary/cloud-computing/

[2] Alexandros Labrinidis, H. V. Jagadish, Challenges and Opportunities of Big Data, Journal proceedings of the VLDB Endowment, Volume 5 Issue 12, August 2012, pages 2032 - 2033.

[3] Big Data, http://en.wikipedia.org/wiki/Big_data

[4] Randal E. Bryant, Randy H. Katz, Edward D." Lazowska, Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society" Version 8: December 22, 2008

[5] A. M. Talib, R. Atan, R. Abdullah and M. A. A. Murad, "*Towards New Access Data Control Technique Based on Multi Agent System Architecture for Cloud Computing in Software Engineering and Computer Systems* Part II," In: V. Snasel, J. Platos and E. El-Qawasmeh, Eds., *Springer Series*: *Communications in Computer and Information Science* 189, Springer-Verlag, pp. 268-279.

[6] R. Rivest, A. Shamir, L. Adleman, "*A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*,"Communications of the ACM, vol. 21, iss. 2, pp. 120–126, February 1978.

[7] Sulaiman, R., D. Sharma, W. Ma and D. Tran, 2007. *A Multi-agent security framework for e-health services.* Knowl. Intell. Inform. Eng. Syst., 4693: 547-554. DOI: 10.1007/978-3-540-74827-4_69

[8] J. Surbiryala, C. Li, and C. Rong, "A framework for improving security in cloud computing," in 2nd IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA 2017). IEEE, 2017.

[9] Kaur M, Mahajan M. (2013). "*Using encryption Algorithms to enhance the Data Security in Cloud".* International journal of communication and computer technologies.

[10] Arora R, Parashar A. "*Secure User Data in Cloud Computing Using Encryption*", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 4, Jul-Aug 2013, pp.1922-1926

[11] Padhy R, P Patra, M. R., &Satapathy, S. C. "*WINDOWS AZURE PAAS CLOUD : AN Overview".* International Journal of Computer Application., Issue 2, Volume 1, February 2012.

[12] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "*Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data*," Proc. IEEE INFOCOM, pp. 829-837, 2011.

[13] I. Roy, S. Setty, A. Kilzer, V. Shmatikov, and E. Witchel. *Airavat: Security and privacy for MapReduce.* Technical Report TR-10- 09, UT-Austin, 2010.

[14] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "*Sedic: Privacy-Aware Data Intensive Computing on Hybrid Clouds,"* Proc. 18th ACM Conf. Computer and Comm. Security (CCS '11), pp. 515-526, 2011.

[15] Philip Derbekoa , Shlomi Dolevb, Ehud Gudesb , Shantanu Sharma "*Security and Privacy Aspects in* MapReduce on Clouds: A Survey" Elsevier Computer Science Review, May 4, 2016

[16] X. Zhang, C. Liu, S. Nepal, S. Pandey, and J. Chen, "*A Privacy Leakage Upper-Bound Constraint Based Approach for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud*," IEEE Trans. Parallel and Distributed Systems, to be published, 2012.

