



IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 Issue: IV Month of publication: April 2016 DOI:

www.ijraset.com

Call: 🛇 08813907089 🕴 E-mail ID: ijraset@gmail.com

An Online Based Approach for Health Care Using Clustering Algorithm with Map Reduce For Scalable Data

R.Sasi Kannan¹, C.Venu Gopal², R.Viswanathan³, S.Athinarayanan⁴ ^{1,2,3}B.Tech Student, ⁴Assistant Professor, Dept. of IT Prathyusha Engineering College, Poonamallee, Tamilnadu, India.

Abstract: The world is moving on the speed of the internet. With the increase of internet user's in past years has given rise to multiple online things like booking tickets, shopping. Shopping through internet has paved way to order medicines through online. So, this method is one such online communications of patients and doctors. Effective and authenticated communication is the best way for people to use any kind of internet participation. Here we provide a concept of data anonymization were the details of every single user whether it is a patient, doctor or a third-party user is hidden from another person. The third party user can be split into 3 and they are region, age, disease through which the classification can be done. This third-party user can be any of the government organization of any medicines manufacturing company's or an university taking a survey. Key Terms— Big Data; Cloud Computing; MapReduce; Data Anonymization; Proximity Privacy.

I. INTRODUCTION

CLOUD computing and big data, two disruptive trends at present, pose a significant impact on current IT industry and research communities [1, 2]. Today, a large number of big data applications and services have been deployed or migrated into cloud for data mining, processing or sharing. The salient characteristics of cloud computing such as high scalability and pay-asyou-go fashion make big data cheaply and easily accessible to various organizations through public cloud infrastructure. Data sets in many big data applications often contain personal privacy-sensitive data like electronic health records and financial transaction records. As the analysis of these data sets provides profound insights into a number of key areas of society (e.g., healthcare, medical, government services, e-research), the data sets are often shared or released to third party partners or the public. The privacy-sensitive information can be divulged with less effort by an adversary as the coupling of big data with public cloud environments disables some traditional privacy protection measures in cloud [3, 4]. This can bring considerable economic loss or severe social reputation impairment to data owners. As such, sharing or releasing privacy-sensitive data sets to third-parties in cloud will bring about potential privacy concerns, and therefore requires strong privacy preservation.

S.NO	TITLE	PROS	CONS	YEAR	AUTHOR	PUBLISHED IN JOURNAL OR
						CONFERENCE
1.	A VIEW OF CLOUD	DATA CONFIDENTIA	LESS VIRTUALIS	2009	M.ARMBUS T	JOURNAL OF ENGINEERING
	COMPUTING	LITY AND	ED		A.D.JOSEPH	SCIENCE AND
		SCALABLE	RESOURCES			TECHNOLOGY
		STORAGE	AND PAY			
			FOR USE			
			LICENSE			
			MODEL			
	SECURITY AND	SECURITY, PRI	LESS DATA	2010	TAKABI	INTERNATIONAL
2.	PRIVACY	VACY AND	CENTRIC		J.B.D.JOSHI	JOURNAL OF
	CHALLENGES	DATA	SECURITY			ADVANCEMENT IN
	IN CLOUD	CONFIDENTIA				ENGINEERING
	COMPUTING	LITY				AND
	ENVIRONMENT					TECHNOLOGY
3.	ADDRSSEING	TRUSTED	AVAILABILI	2012	D.ZISSIS	INTERNATIONAL
	CLOUD	THIRD	TY OF		D.LEKKUS	JOURNAL OF
	COMPUTING	PARTY,SECURI	DATA IS			COMPUTER
	SECUURITY	TY	LOW AND			APPLICATION AND
	ISSUES	IDENTIFICATI	NO			RESEARCH
		ON OF	PRIVACY IN			
		THREATS	IHIRD			
4.	A PRIVACY	DATA SET	HIGH COST	2013	X.ZHANG	IEEE
	LEAKAGE	ENCRYPTION	DATA SET		C.LIU	TRANSACTIONS
	UPEER BOUND		ENCRYPTIO		S.PANDEY	ON PARALLEL
	CONSTRAINT		Ν			AND
	BASED					DISTRIBUTIVED
	APPROACH					SYSTEMS
	FOR COST					VOLUME:24
	EFFECTIVE					
	PRIVACY					
	PRESERVING					
	OF					
	INTERMEDIAT					
	E DATA SETS					
	IN CLOUD					
5.	A SECURE	STORING	PROXY	2012	L.H.SIAO	IEEE
	ERASURE	DATA IN	RENCRYPTI		W.G.TZENG	TRANSACTIONS
	CODE-BASED	THIRD PARTY	ON			ON PARALLEL
	CLOUD	AND DATA				AND
	STORAGE	CONFIDENTIA				DISTRIBUTIVED
	SYSTEM WITH	LITY				SYSTEMS
OFT A	SECURE DATA	-1				VOLUME:23 2
<u>5E1: Al</u>	HUCHRISVAR BIRESErve		MINOD	2014		
6.	PRIVACY	PRESERVING	MINOR	2014	HMOOD.A	INTERNATIONAL
	PRESERVING	ROTH	CLUSTER	1	FUNG	I JOUKNAL OF

				1	/	1
	MEDICAL	INFORMATION	QUALITY			TECHNOLOGY
	REPORTS	UTILITY AND				AND RESEARCH
	PUBLISHING	PRIVACY				
	FOR CLUSTER					
	ANALYSIS					
7.	ANONYMITY	ANONYMIZATI	INTEGRATE	2011	BENJAMI	INTERNATIONAL
	MEETS GAME	ON	D DATA		C.M.FUNG	JOURNAL OF
	THEORY:SECU	TECHNIQUE IS	WITHOUT		MOURAD	MODERN
	RE DATA	USED	REVEALING		DEBBAI	ENGINEERING
	INTEGRATION		MORE			AND RESEARCH
	WITH		DETAILED			
	MALICIOUS		INFORMATI			
	PARTICIPANT		ON			
8.	K-ANONYMITY	DATA	REIDENTIFI	2002	L.SWEENEY	INTERNATIONAL
	MODEL FOR	ANONYMITY	CATION IS			JOURNAL OF
	PROTECTING	AND DATA	DIFFICULT			TRENDS AND
	PRIVACY	PRIVACY				TECHNOLOGY
9.	DISRIBUTED	MULTIPLE	LESS	2009	P.JURCZYK	INTERNATIONAL
	ANONYMIZATI	DISTRIBUTED	PRIVACY		L.XIONG	JOURNAL OF
	ON:ACHIEVING	AND PRIVATE				COMPUTERL
	PRIVACY FOR	DATABASES				APPLICATION
	BOTH DATA					
	SUBJECTS AND					
	DATA					
	PROVIDERS					
10.	SEDIC:PRIVAC	MAP	LARGE	2011	K.ZHANG	INTERNATIONAL
	Y AWARE	REDUCING IS	AMOUNT		X.ZHOU	JOURNAL
	DATA	USED	OF			SECURITY
	INTENSIVE		COMPUTAT			
	COMPUTING		ION IS			
	ON HYBRID		REQUIRED			
	CLOUDS					

II. CONCLUSIONS

In this paper, local-recoding anonymization for big data in cloud has been investigated from the perspectives of capability of defending proximity privacy breaches, scalability and time-efficiency. We have proposed a proximity privacy model, -dissimilarity, by allowing multiple sensitive attributes and semantic proximity of categorical sensitive values. Since the satisfiability problem of - dissimilarity is NP-hard, the problem of big data local recoding against proximity privacy breaches has been modeled as a proximity-aware clustering problem. We have proposed a scalable two-phase clustering approach based on MapReduce to address the above problem time-efficiently. A series of innovative MapReduce jobs have been developed and coordinated to conduct data-parallel computation. Extensive experiments on realworld data sets have demonstrated that our approach significantly improves the capability of defending proximity attacks, the scalability and the time-efficiency of localrecoding anonymization over existing approaches

REFERENCES

- S. Chaudhuri, "What Next?: A Half-Dozen Data Management Research Goals for Big Data and the Cloud," Proc. 31st Symp. Principles of Database Systems (PODS'12), pp. 1-4, 2012.
- [2] L. Wang, J. Zhan, W. Shi and Y. Liang, "In Cloud, Can Scientific Communities Benefit from the Economies of Scale?," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 2, pp. 296-303, 2012.
- [3] X. Wu, X. Zhu, G.-Q. Wu and W. Ding, "Data Mining with Big Data," IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 97-107, 2014.
- [4] X. Zhang, C. Liu, S. Nepal, S. Pandey and J. Chen, "A Privacy Leakage Upper Bound Constraint-Based Approach for CostEffective Privacy Preserving of Intermediate Data Sets in Cloud," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 6, pp. 1192-1202, 2013.
- [5] B.C.M. Fung, K. Wang, R. Chen and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Comput. Surv., vol. 42, no. 4, pp. 1-53, 2010.
- [6] L. Sweeney, "K-Anonymity: A Model for Protecting Privacy," Int'l J. Uncertain. Fuzz., vol. 10, no. 5, pp. 557-570, 2002.
- [7] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam, "L-Diversity: Privacy Beyond kAnonymity," ACM Trans. Knowl. Disc. Data, vol. 1, no. 1, Article No. 3, 2007.
- [8] N. Li, T. Li and S. Venkatasubramanian, "Closeness: A New Privacy Measure for Data Publishing," IEEE Trans. Knowl. Data Eng., vol. 22, no. 7, pp. 943-956, 2010.
- [9] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi and A.W.C. Fu, "UtilityBased Anonymization Using Local Recoding," Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data (KDD'06), pp. 785-790, 2006
- [10] K. LeFevre, D.J. DeWitt and R. Ramakrishnan, "WorkloadAware Anonymization Techniques for Large-Scale Datasets," ACM Trans. Database Syst., vol. 33, no.3, pp.1-47, 2008.
- [11] G. Aggarwal, R. Panigrahy, T. Feder, D. Thomas, K. Kenthapadi, S. Khuller and A. Zhu, "Achieving Anonymity Via Clustering," ACM Trans. Algorithms, vol. 6, no. 3, Article No. 49, 2010.
- [12] T. Wang, S. Meng, B. Bamba, L. Liu and C. Pu, "A General Proximity Privacy Principle," Proc. IEEE 25th Int'l Conf. Data Engineering (ICDE'09), pp. 1279-1282, 2009.
- [13] T. Iwuchukwu and J.F. Naughton, "K-Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization," Proc. 33rd Int'l Conf. Very Large Data Bases (VLDB'07), pp. 746-757, 2007.
- [14] X. Zhang, L.T. Yang, C. Liu and J. Chen, "A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using Mapreduce on Cloud," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 2, pp. 363-373, 2014.
- [15] X. Zhang, C. Liu, S. Nepal, C. Yang, W. Dou and J. Chen, "A Hybrid Approach for Scalable Sub-Tree Anonymization over Big Data Using Mapreduce on Cloud," J. Comput. Syst. Sci., vol. 80, no. 5, pp. 1008–1020, 2014.











45.98



IMPACT FACTOR: 7.129







INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 🕓 (24*7 Support on Whatsapp)