May – June 2020
ISSN: 0193-4120 Page No. 2321 - 2328

Data Classification Using K_NN Classifier and Data Security Using RSA and Biometric Encryption

[1]D.Divya Prathyusha , [2]G.Abhishek, [3]P.Sai Teja , [4]B. Tirapathi Reddy
Department of Computer Science and Engineering, K L University, Vaddeswaram, Guntur.
[1]prathyushadarsi@gmail.com, [2]abhishekgvish@gmail.com, [3] saitejapotla9@gmail.com,
[4]tirapathireddyb@kluniversity.in

Article Info
Volume 83
Page Number: 2321 - 2328
Publication Issue:
May - June 2020

Article History
Article Received: *11August 2019*
Revised: *18November 2019*
Accepted: *23January 2020*
Publication:*10May2020*

Abstract

Cloud Computing[4][17] Is One Of The Subject Undergoing Intense Study And Security Is An Important Feature Of Cloud Computing[4]. In Cloud, End User's Data Is Stored On Distant Servers Which May Be Handled By Others And Can Be Acquired Through Internet Connection. There Are Many Ways That The Data May Be Weaken. Hence Confidentiality[3] Of Data Becomes A Serious Issue. This Is One Of The Important Features Of Security. In This Paper, We Do A Study Of The Issues Related To Confidentiality In Cloud Computing[4] An Enhance The Confidentiality[3] And Thus Ensuring The Cloud Security. Security Is The Most Important Aspect In Present Cloud Computing. We Need To Provide The Security Without Losing The Confidentality. And We Proposed An Idea For The Basic Encryption[13][2] Standard Techniques Works To Secure The Data In Sever From Attacks. In This Paper We Want To Explain Some Techniques And Mechanism Which Is Mostly Using In Sever Encryption[13][2] Still. And Coming To User End Security We Need To Use Secure Techniques To Protect The Data In Most Efficient Way. In The User End Security We Are Explain A Biometric Encryption. K-Nn Classifier[15]. In This Paper We Will Analyze The Techniques That We Are Explaining And Provide Some Sort Of Solution To Provide More Confidentality[3][1][14]. The Major Objective Of This Paper Is To Provide Confidentiality Which Is One Of The Major Characteristics Of Security That Cloud Needed.

I. INTRODUCTION

Confidentiality[3] Safe Guard The Data Which Is Present In The Cloud And It Can Be Retrieved Only By The Authorized Party. It Is A Major Threat Of Cloud Computing. Confidentiality Plays A Prime Role In The Protection Of Organization Or Individual Data, Secure Protocols Can Be Applied At Various Different Levels Of The Cloud. Confidentiality Is Compromising Due To Dishonest Cloud Service Providers. Confidentiality Can Be Secured Through Better Encryption[2][13] Methods .Confidentiality Is Derived Through Encipher Techniques. But Encipher Alone May Not Provide The Security. A Technique Uses Both Encipher And Obfuscation To Defend The Confidentiality As It Considered That Encipher Alone Cannot Provide Security. In This Paper We Are Looking To Find A Solution For The Cloud Data Confidentiality[3][13][2]To Provide Security[5]. The Cloud Computing Is One Of The Leading Technology That Have Been Using Still. Cloud Computing Is Still Lagging Protecting The Data In It. Security Is The Major Expect Of The Cloud Now A Days. The Existed Techniques Are Securing The Data Up To Some Extent Not In The Complete Way. So Thus Way We Come Up Some Modifications In The Present Using Techniques. In This Paper We Are Discuss About That Techniques And The Modifications. The Techniques Are Masking[1], Knn-Clasifiler And Biometric Encryption These Are One Of The Present Techniques Providing The

Confidentiality To The Cloud Computing Existed Data. And We All Know That Cloud Has Been Providing The Service And The Security Comes Under The Identity As A Service. And We Perform Some Operations On It And Get Some Results Which Prove That More Secure Then The Past Ones.

II. Related Work

Confidentiality Through Encipher And Obfuscation. It Uses Both Encipher And Masking To Safe Guard The Confidentiality As It Considers Encipher Alone Cannot Provide Security. Same With The Case Of Masking As Reverse Engineering Attacks Or Brute Force Methods Can Break This. Here Obfuscation Is Integrated With Encryption. Masking Uses Mathematical Functions Root () And Floor ().Encipher Can Be Applied To Alphabets And Alphanumeric Type Of Data And Masking To Numeric Data. The Data That Needs To Be Stored Should Be Encipher And Masking.

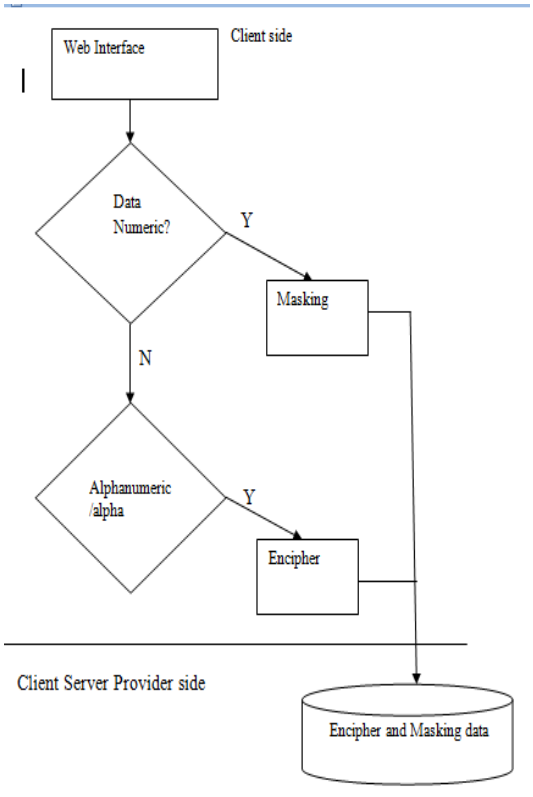

Figure1: Data Obfuscation And Encryption

Nulling Out Or Deletion :It Is One Of The Simple Way For Masking. In This We Will Mask A Null Value To A Particular Field. This Will Prevent The Visibility Of The Elements In The Data. It Reduces The Data Integrity And Maintained In The Set Of Data .But It Is Not Practical Value May Fail At Any Application Validation At The Front End Of The Software.

Masking[13] Out: Scrambling The Data Is One Of The Simplest Ways To Protect The Data For The Unwanted Users. It Is The Extension Of Its Previous Technique Nulling Out .But It Does Not Allows For Full Masking. This Technique Is Commonly Used For The Credit Card System .Customer Uses His Card Mostly For The Billing Purposes. Now The Shop Keeper Or The Merchant Will Bill The Item By Swapping It In The Machine. Then They Quote A Reference To The Credit Card With Last 4 Digits In The Card Xxxx Xxxx 4444. As A Merchant He Can Only See The 4 Digits In The Card. Once The Billing System Enters The Customer's Details For Payment, The Full Number Is Disclosed To The Payment Gateway Systems. It Comes Under Dynamic Data Masking.

Types Of Attacks Known Plaintext Attack On Rsa: Known Plaintext Attack Is An Attack Which Is Having Both The Unencrypted Text (Plain Test) And Encrypted Text (Cipher Text) Of The Data Available. This Data Is Used To Determine The Secret Key And Encode And Decode The Available Data. In Olden Days These Ciphers Are Very Capable To The Attack, Where As Present Ciphers Are Lesser Capable For The Attacks.

The Chosen Plaintext Attack On Rsa : The Hacker Has The Potential To Access The Selected Random Unencrypted Text (Plaintext) And See The Encrypted Text (Cipher Text). The Aim Of This Attack Is To Gain The Additional Information Or Data That Will Reduce The Security Or Complexity Of The Cryptogram. The Secret Key Can Be Acquired Which Reduces The Splits In The Cipher. In Some Cases Of The Chosen Plaintext Attack, A

May – June 2020
ISSN: 0193-4120 Page No. 2321 - 2328

Minimum Amount Of Unencrypted Text (Plaintext) Should Be Familiar By The Attacker.

Known Cipher Text Attack On Rsa: Cipher Text Attack Is An Model For Attacking The Cryptanalysis In Which The Attacker Has The Ability To Access A Set Of Cipher Text. In This The Attacker Will Have Some Knowledge About The Plain Text (Unencrypted Text) Or The Language It Was Written Or The Encryption Which Was Used .The Protocols Data And Message Are The Part Of Plain Text .

The Chosen Cipher Text Attack On Rsa: A Chosen Cipher Text Attack Is An Model For Attacking The Cryptanalysis In Which The Cryptanalysis Is Able To Gather The Data By Deciphering The Chosen Cipher Text. The Main Objective Of This Is To Gain The Information That Reduces The Security Of The Encipher Scheme.

III. Proposed System

Data Masking[13]: Data Security[5] In The Cloud The Process Of Encryption[2] Occurs On Sender Side At The Socket Layer. The Figure Also Shows Us That The Data Is Free Of Invader Attack Because Mask Is Applied Only To Data There By Hiding Original Data. Once Masking [12] Is Done It Is Added Up With Padding Because Of Which Data Is Added Up With More Security. All These Gives Double Data Encryption Of Which Data Is Transferred More Securely To Receiver.

At This Receiver End, The Exact Reverse Process Of Sender Takes Place. Here The Diagram Illustrate The Data Is Doubly Decrypted Since Was Doubly Encrypted From Senders Side. Firstly, Padding Of Data Is Removed After Which We Obtain Masked Data And Then Masking Is Also Removed Thus Obtaining An Original Data Form.

Importance Of Data Masking: When Copy Sensitive Data Outside Of Production Environment. Moving The Test Data To Virtual Storage Namely Cloud. Passing Data To End Customers. Grip Off-Shore Development.

Sender Side

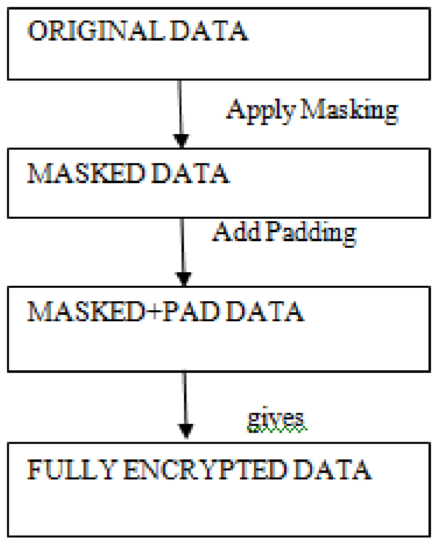

Figure 2: Sender Side Process

Receiver Side

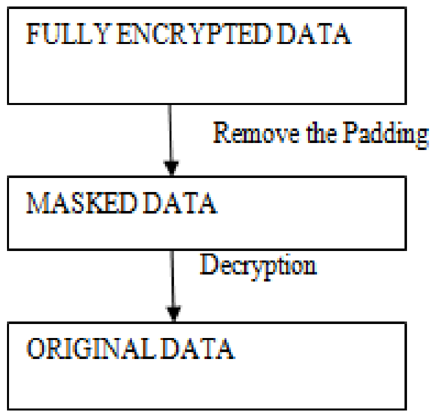

Figure 3: Receiver Side Process

Varieties Of Data Masking

Static Data Masking: For The Outsourced Developers In Different Locations And Different Companies This Masking Is The Only Useful One And Is Used By Many Organizations When They Generate Testing For Module. Duplication Of Databases Is Must In These Scenarios. For Making This Happen We Should Have Static Masking Tools. These Tools Will Make Data A Sensitive One While Passing It Out Of Any Company. Using A Standard Etl Procedure This Masking Will Give A Data

Protection By Generating A Testing Or Offline Database. This Database Can Be Updated On A Daily Or Weekly Basis. This Might Be A Security Risk, Thus Have An End Up With Different Development And Test Issues.

Dynamic Data Masking: As When The Data Is Requested The Data Streams From Database Environment Are Masked Or Transformed And Basically Ddm Is A Technique For Limiting An Unauthorized Data Access. This Masking Will Give An End Result To Those Who Are Not Having Access To Original Data But Are Working In Production Environment. For An Instance, Contractors And Staff Were Always Trying To Update A Database Or Even Sometimes Troubleshoot It So, In These Case It Is Must To Have No Access To Sensitive Data Such As Health Or Some Private Card Numbers So On And So Forth. Here The Information Is Twisted Or Changed So That These Staff Will Working On Data Which Is Harmless As They Modify The Database.

Data Masking[13] And The Cloud:Most Of The Organizations Are Developing Their Applications In Cloud These Days. Iaas, Paas, Saas Are The Ones Which Are Allowed By To Organizations By Cloud Solution. We Can Create Test Data In Different Ways And Shift It To Cloud. In Sdlc Data Masking Now Is A Part As Development Environments.

Data Masking Techniques[13]

Substitution: This Strategy Is The Best One Of Applying Masking And Be Able To Store Authentication Of Data Records[13]. This Method Does Replace Contents Of Column Of Data With Information Which Seems Similar But Is Far More Different From Real Details. For Instance, Surnames In A Customer Db Could Be Stabilized By Replacing Last Names With These Surnames From Random List. Substitution Data Will Be Hard But Data Masking[13] Should Have Datasets Of Required Items. For Example To Replace Surnames By Substitution, A List Of Last Names Be Available. Then To Replace Mobile Numbers, A List Of Phone Numbers Be There. This Method Has To Be Applied For All Such Fields In Db Structure.

Shuffling: This Technique Is Similar To Substitution Except That It Is From A Column Itself. Here The Data Is Shuffled Randomly Within Columns. It Is Best For Small Data Sets. If This Method Is Determined Then The Data Can Be Easily Shuffled. For Example If This Method Went Down To Table Swapping The Column Data In Every Group Of Two Rows It Won't Take Much Stress In Reverting Things Back To Their Original State. It Is Not That Effective When Used For Small Amount Of Data.

Encryption[2]: It Is The Most Typical Method For Solving Masking Problem. This Will Generally Mix Up The Total Data. This Won't Leave Data Which Is Really Realistic And Can Make Data Large Sometimes. It Also Kills The Look And Feel Of Data Along With Its Formatting. The Encrypted Data Is Not Meaningful And It Looks Like A Binary Data. When Changing Encrypted Varchar Fields It Generally Leads To Character Set Issues. Some Types Of Encryption May Also Put Constraints On Format. This Says That Fields Must Be With Suitable Padding Character Which Must Be Taken Off While Decrypting It.

Biometric Encryption[12][11]

A Biometric [11] Is Determined As A Unique, Computational, Biological Characteristic Or Feature For Automatically Verifying Or Recognizing The Human Being. Biological Attribute Has Become The Fields For Biometrics. Now-A-Days, Biometric Machinery Are Used To Examine Human Attributes For Security Purposes. Physical Biometric Patterns Are Examined For Security Causes Are Eye, Face, Voice, Hand And Fingerprint. Biometric Identification Contains Of Two Levels: Enrollment Level And Verification Level. During The Enrollment Level, A Trial Of The Classified Biometric Is Obtained. Some Unique Features Or Characteristics Of This Trial Are Then With Drawn

May – June 2020
ISSN: 0193-4120 Page No. 2321 - 2328

To Form A Biometric Pattern For Successive Contrast Purposes. During The Verification Level, An Updated Biometric Trial Is Acquired. As In Enrollment, Features Of This Biometric Trial Are Withdrawn. These Characteristics Are Then Differentiating With The Previously Obtained Biometric Pattern[11]. It Is Easy To Differentiate Among The Main Intension Of The Biometric Process Identification And Authentication. The Process For Biometric Identification Is To Match An Individual To Large Set Of Users Of The System And In The Process Of The Biometric Authentication[1][14] Simply Verifies Claimed Is Valid Or Not. Biometric Encipher Is A Method That Is Used To Protect The Biometric Identification. Biometric Encipher Is Different From Normal Password Encipher Because Here The Biometric Image Is Combined With The Randomly Generated Key Using Be Binding Algorithm To Create Biometrically Encipher Key. While Decipher, The Biometrically Encipher Key Is Combined With Biometric Image Using Be Retrieval Algorithm To Get The Key Retrieved.

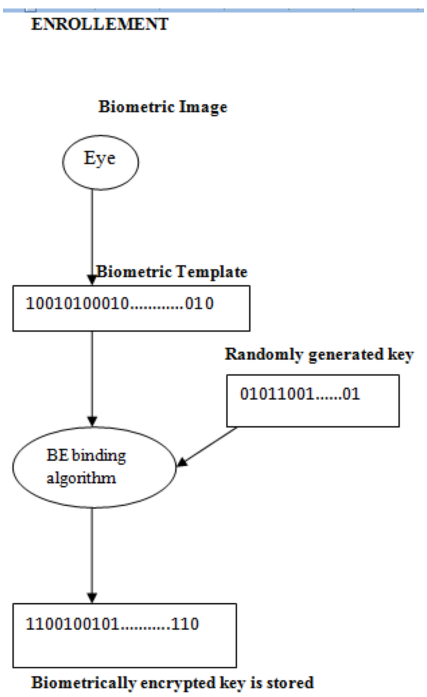

Figure 4: Encryption Phase

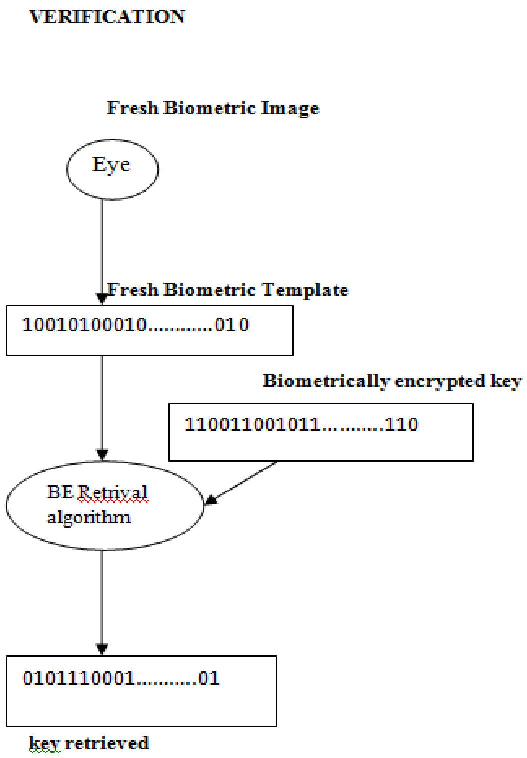

Figure 5: Verification Phase

K-Nn Classifier[15][19]

Knn Algorithm Classification [14] Is An Algorithm And It Is One Of The Most Used Learning Algorithms.In K-Nn Classifier The Data To Be Stored On The Needs Of Security Like What Kind Of Data Require The Security And What Data Do Not Require The Security. This Is Done By Using K-Nn (K-Nearest Neighbour) Classification. This Data Is Divided Into Two Ways, Non-Sensitive And Sensitive. Rsa Is Appeal To The Sensitive Data And For Non Sensitive Data It Is Stored As It Is In Virtual Machine-Nn Machine Learning Is Used To Segregate The Sensitive Data From The Non-Sensitive Data. K-Nn Is Used In A Designed Simulation Environment For Accuracy Purpose, The Value Of K Is Maintained To 1. After Separating Sensitive From Non-Sensitive Data, Sensitive Data Is Passed To Rsa Algorithm For Encryption[2]. Non Sensitive Data Is Directly Allocated To A Vm (Virtual Machine)[8] Without Any Encryption Techniques. Sensitive Data Is That Data Which Is Very Important For Individual Or Organizations Like Personal Data, Financial Records, Business

ISSN: 0193-4120 Page No. 2321 - 2328

Material, Legal, Medical, Government Data Etc. Non Sensitive Data Is The Data Used In Public For Marketing Information, Announcements Etc. K-Nn Is One Of The Supervised Algorithm Which Depends Up On The Instance Based Learning[19], Where To Categorize New Unclassified Data Sets Into User Specified Classes 'K', A Set With The Data Which Is Trained Is Stored. K-Nn[15] Calculates The Distance Between New Input And All Of Training Data And Then Sorts The Distance To Determine The K Th Minimum Distance. Finally It Determines The Class Of The New Input Based On The Majority Vote.

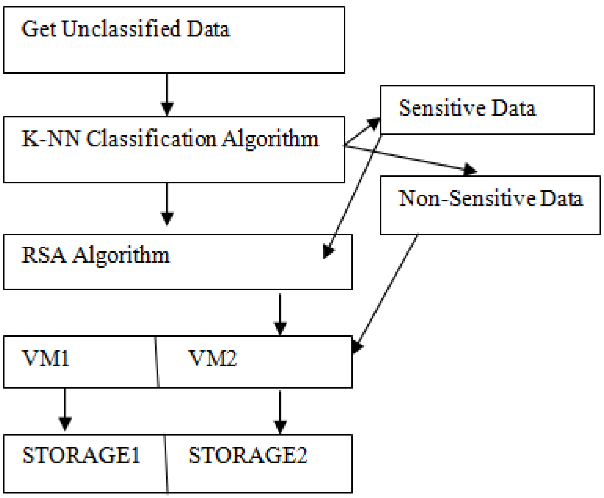

Figure 6: K-Nn Classifier Algorithm

RSA Algorithm[20]

1 Select P And Q Both Should Be Prime And $P \neq Q$.

2 Calculate N, $N=P*Q$

3 Calculate $\Phi (N) = (P-1) (Q-1)$

4 Select Integer E Where $Gcd(\Phi(N),E)=1 : 1<E< \Phi(N)$

5 Calculate D Where $D=E^{-1}(Mod\ (\Phi\ (N))$

6 Public Key Pu= {E, N}

7 Private Key Pr= {D, N}

Encryption

Plaintext $M<N$

Cipher Text $C=M^e$ Mod N

Decryption

Cipher Text C

Plaintext $M= C^d$ Mod N

III. Conclusion

In This Paper We Find The Solutions For The Cloud Data Confidentiality[3] To Provide Security. In This Paper We Discussed About The Techniques And The Modifications. The Techniques We Used Are Masking, Knn-Classifier[15][19] And Biometric Encryption[12][11] These Are One Of The Present Techniques Providing The Confidentiality To The Cloud Computing Existed Data. In This Encryption Needs To Be Integrated With Some Other Methods To Provide Better And Stronger Security And Also Discussed The K-Nn Classification Method For Classification Of Data As Sensitive And Non-Sensitive And Then Applying Encryption Only On Sensitive Data Using Rsa.Rsa Algorithm Is Applied For This . Biometric Encryption[12][11] Is A Method That Is Used To Protect The Biometric Identification. Chosen Cipher Attack[20] Is Not Possible For More Random Values. The Above Techniques We Can Provide The Confidentality[3][1]. We Can Try To Provide More Security To Algorithms Like Attribute Based Encryption.

IV. Future Scope

The Future Work Are Strengthening Confidentality[3] [2] In Cloud Network[16] Communication, Data Location And Applying Searchable Encryption Schemes To Create Secure Cloud And Implement Through Robust Security. In Further Rsa[20] Can Be Secured More By Implementing High Bit Size. To Ensure Backward And Forward Secrecy, A Rekey Process Must Be Triggered After Each Membership Change (Join Or Leave) In The Group. It Consists In Generating A

New Pair Of Keys And Distributing It To The Members Including The New One In Case Of A Join Or To The Residual Members In Case Of A Leave. Confidentiality[3] Is Ensured With Rsa Algorithm.

Acknowledgements

We Thank Mr. B. Tirapathi Reddy (Professor) From Koneru Lakshmaiah Educational Foundation (Klu) For His Guidance By Giving Valuable Suggestions And Never-Ending Support Which Made Us To Develop This Publication.

References

1. Tirapathi Reddy, B. M.V.P Chandra Sekhara Rao. "Privacy-Preserving Proof of Ownership for Data in Cloud Storage Systems International Journal of Engineering and Technology, , Vol.7 (2.8) 13-17. 2018.
2. Survey On Cloud Computing And Data Masking Techniques Priya Dhir Research Scholar Phd.In Computer Application (Rimt University) Sushil Garg Rimt-Maec, Mandi Gobindgarh, Pb.
3. Confidentiality Issues In Cloud Computing And Countermeasures: A Survey Yusuf Haider M1, Siva Selvan2 .Department Of Computer Science And Engineering Manipal Institute Of Technology (Manipal University) Manipal, India.
4. Cloud Computing: Overview And Research Issues Divya Kapil School Of Computing, Graphic Era Hill University, Dehradun,India.Sonu Kumar Dept. Cse, Soet, Hnbg University, Uttarakhand, India,Parshant Tyagi Goldplus Glass Industry Limited, Roorkee, India.Mr. Vinay Prasad Tamta, Dept. Cse, Soet, Hnbg University, Uttarakhand, India.
5. Providing User Security Guarantees In Public Infrastructure Clouds Nicolae Paladi, Christian Gehrmann, And Antonis Michalas .
6. Towards Trusted Cloud Computing Nuno Santos Krishna P. Gummadi Rodrigo Rodrigues
7. Cloud-Trust—A Security Assessment Model For Infrastructure As A Service (Iaas) Clouds.Dan Gonzales, Member, Ieee, Jeremy M. Kaplan, Evan Saltzman, Zev Winkelman, And Dulani Woods .
8. Dropping Activation Outputs With Localized First-Layer Deep Network For Enhancing User Privacy And Data Security Hao Dong, Chao Wu , Zhen Wei, And Yike Guo .
9. Https://En.Wikipedia.Org/Wiki/Data_Masking
10. Https://Www.Imperva.Com/Blog/2017/07/Static -Versus-Dynamic-Data-Masking.
11. Biometric Encryption Colin Soutar, Danny Roberge‡, Alex Stoianov, Rene Gilroy, And B.V.K. Vijaya Kumar ,Bioscrypt Inc.(Formerly Mytec Technologies Inc.),5450 Explorer Drive, Suite 500,Mississauga, Ont,L4w 5m1
12. Biometric Encryption: Technology For Strong Authentication, Security And Privacy,Ann Cavoukian, Alex Stoianov And Fred Carter,Office Of The Information And Privacy Commissioner, Toronto, Ontario, Canada,{Commissioner, Alex.Stoianov, Fred.Carter}@Ipc.On.Ca
13. Data Masking, Encryption, And Their Effect On Classification Performance: Trade-Offs Between Data Security And Utility Juan C. Asenjo Nova Southeastern University, Asenjo.Juanc@Gmail.Com.
14. Privacy And Confidentiality Issues In Cloud Computing Architectures David Jiménez Martínez.
15. An Enhanced K-Nearest Neighbor Algorithm Using Information Gain And Clustering;CharuGupta ; Kratika Goyal ; Dharna Gureja
16. Cryptography and Network Security Principles and Practices, Fourth Edition By William Stallings Publisher: Prentice Hall,Pub Date: November 16, 2005,Print ISBN-10: 0-13-187316-4,Print ISBN-13: 978-0-13-187316-2,eText ISBN-10: 0-13-187319-9,eText ISBN-13: 978-0-13-187319-3
17. Cloud Computing: A Practical Approach Anthony T. VelteToby J. Velte, Ph.D Robert

May – June 2020
ISSN: 0193-4120 Page No. 2321 - 2328

Elsenpeter New York Chicago San Francisco,Lisbon London Madrid Mexico City,Milan New Delhi San Juan,Seoul Singapore Sydney Toronto

18. J. Daugman, “High confidence visual recognition of persons by a test of statistical independence”,IEEE Trans. on Pattern Analysis and Machine Intelligence 15, 1148-1161, (1993)
19. Machine Learning ,Tom M. Mitchell ,Publisher: McGraw-Hill Science/Engineering/Math; (March 1, 1997)
20. Common Attacks on RSA and its Variants with PossibleCountermeasures,Auqib Hamid Lone*, Prof. Moin Uddin,Department of Computer Science & Engineering,,Jamia Hamdard, New Delhi, India