INCREASING DATA ANONYMITY USING PRIVACY TECHNIQUES AND ADVANCED ENCRYPTION STANDARD

THAMER KHALIL ESMEEL

Master of Science

UNIVERSITI MALAYSIA PAHANG

MAKLUMAT PANEL PEMERIKSA PEPERIKSAAN LISAN

(only for Faculty of Computer's student)

Thesis ini telah diperiksa dan diakui oleh
This thesis has been checked and verified by

Nama dan Alamat Pemeriksa Dalam : Assoc. Prof. Dr. Noraziah Binti Ahmad
Name and Address Internal Examiner Faculty of Computing, Universiti Malaysia Pahang

Nama dan Alamat Pemeriksa Luar : Prof. Dr. Abdul Azim Abd Ghani
Name and Address External Examiner Fakulti Sains Komputer dan Teknologi Maklumat, Universiti Putra Malaysia

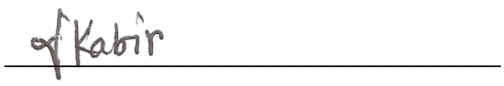
Disahkan oleh Penolong Pendaftar IPS :
Verified by Assistant Registrar IPS

Tandatangan : Tarikh :
Signature *Date*

Nama :
Name

SUPERVISOR'S DECLARATION

We hereby declare that we have checked this thesis and in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science in Software Engineering.

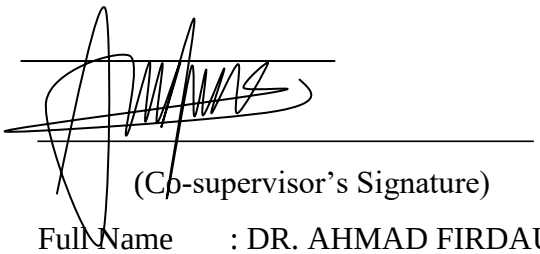


(Supervisor's Signature)

Full Name : DR. MUHAMMAD NOMANI KABIR

Position : SENIOR LECTURER

Date : 22 FEBRUARY 2021



(Co-supervisor's Signature)

Full Name : DR. AHMAD FIRDAUS BIN ZAINAL ABIDIN

Position : SENIOR LECTURER

Date : 22 FEBRUARY 2021

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

(Student's Signature)

Full Name : THAMER KHALIL ESMEEL

ID Number : MCS18001

Date : 21 FEBRUARY 2021

INCREASING DATA ANONYMITY USING PRIVACY TECHNIQUES AND ADVANCED ENCRYPTION STANDARD

THAMER KHALIL ESMEEL

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Science

Faculty of Computing

UNIVERSITI MALAYSIA PAHANG

APRIL 2021

ACKNOWLEDGEMENTS

All praise to Allah and (Azza wa a jall) and peace and blessings be upon our Prophet Muhammad, the Messenger of Allah.

First, my sincere thanks to supervisor Dr Muhammad Nomani Kabir and Co-supervisor Dr Ahmad Firdaus bin Zainal Abidin for their guidance and encouragement and constant support to me during the writing of this thesis, and my appreciation to them for being patient and tolerance in dealing with my mistakes during the research.

Second, I would like to express my gratitude towards my brothers and sisters and all my family members of my wife, son and daughter for all forms of support that they have provided me to complete this research.

Third, I would like to thank the Ministry of Higher Education and Scientific Research, University of Mosul, College of Arts, Iraq for allowing me to study at Universiti Malaysia Pahang.

Last but not least, I would like to thank all those who helped directly or indirectly to make this thesis successful and supported me during my research.

ABSTRAK

Keselamatan data dan privasi data telah menjadi bidang yang perlu diberi perhatian sejak akhir-akhir ini. Dataset sering terdiri dari bidang data sensitif dimana pendedahan ia mampu membahayakan individu yang berkaitan dengan data tersebut. Dengan kemajuan teknologi terkini, penyerang mampu membangunkan kaedah baru dari semasa ke semasa untuk mendapatkan akses ke maklumat sensitif dari kad pengenalan bank, kad warganegara dan pengundi. Untuk menyelesaikan masalah ini, teknik privasi menghalang pengenalan orang tersebut melalui peningkatan anonimitas individu dalam set data untuk melindungi maklumat sensitif. Tesis ini menangani masalah melindungi data pengguna sensitif dalam set data dari pelanggaran privasi dan keselamatan dan aktiviti jahat oleh penyerang dengan menggunakan teknink bersepadu antara teknik privasi dan sekuriti Advanced Encryption Standard (AES). Objektif kajian adalah untuk membangunakan teknik terpadu diantara teknik privasi dan teknik keselamatan untuk melindungi data dalam set data. Untuk melaksanakan teknik bersepadu antara teknik privasi dan sekuriti untuk melindungi data pengguna dalam set data. Untuk validasi teknik bersepadu dengan membandingkan keputusan ketepatan sebelum dan selepas menggunakan teknik privasi dalam set data menggunakan teknik perlombongan data. Kajian metodologi mengandungi tiga fasa. Fasa pertama menentukan set data yang sesuai dan menyiasat teknik dan memahami teknik dan rangka kerja berkaitan dengan privasi dan sekuriti. Fasa kedua melibatkan perlaksanaan teknik bersepadu oleh teknik privasi bersama AES. Fasa ketiga melibatkan penilaian hasil klasifikasi dari pelaksanaan eksperimen proses teknik yang dicadangkan. Metodologi kajian ini terdiri daripada tiga fasa. Fasa pertama menentukan set data yang sesuai dan menyiasat serta memahami teknik-teknik dan rangka kerja berkaitan dengan privasi dan sekuriti. Fasa kedua pula melibatkan perlaksanaan teknik bersepadu oleh teknik privasi dengan AES. Fasa ketiga melibatkan penilaian keputusan klasifikasi selepas bereksperimen dengan pelaksanaan proses teknologi yang dicadangkan. Dalam penyelidikan ini, teknik keselamatan yang digunakan adalah AES, dan teknik privasi adalah privasi berbeza, k-Anonymity, Sampel-uniqueness; dan teknik perlombongan data (Naive bayes, J48 dan Neural network) dengan menggunakan Weka Experimenter Environment (Weka). Pelbagai ujian telah dilakukan menggunakan Weka untuk mengklasifikasikan data untuk memeriksa kegunaan data untuk tujuan analisis. Hasil klasifikasi sebelum dan sesudah menggunakan teknik privasi. Didapati bahawa perbezaan privasi pada kolum jantina set data memberikan keputusan terbaik dalam ketepatan dan mengungguli teknik k-Anonimiti dan teknik sample-uniqueness. Walau bagaimanapun, teknik k-Anonimiti dan sampel-uniqueness pada kolum umur dan kolesterol menunjukkan hasil yang lebih baik dalam keputusan daripada teknik privasi pembezaan. Selain itu, diperhatikan bahawa untuk ketepatan klasifikasi pada Weka, Naïve bayes memberikan hasil yang lebih baik daripada Neural network dan J48. Untuk mengelakkan set data dari penyerang luaran, AES digunakan untuk mengenkripsi set data yang dilalui melalui teknik privasi. Untuk meningkatkan keselamatan, fail yang dienkripsi kunci umum yang terlibat dengan penyulitan adalah dibahagikan kepada lima fail dan kemudian setiap subfile dan subkunci dilampirkan dan disimpan di lima pelayan. Hasil ujian menunjukkan keputusan yang positif dengan menggunakan teknik bersepadu menggunakan teknik privasi dengan teknik keselamatan untuk melindungi data pengguna sensitif dalam set data. Untuk kajian akan datang, kajian boleh difokuskan pada peningkatan ketepatan klasifikasi skema perlindungan dengan memeriksa lebih banyak teknik privasi dan menerapkannya pada data awan.

ABSTRACT

Data security and data privacy have been an important area in recent years. Dataset often consists of sensitive data fields, exposure of which may jeopardize individuals associated with the data. With the technological advancement, the attackers have been developing new methods from time to time to gain access to sensitive information from the bank, national and voter identification cards, etc. To resolve this issue, privacy techniques hinder the identification of the person through an increase of the anonymity of the individual in the dataset to protect sensitive information. This thesis addresses the problem of protecting sensitive user data in the dataset from privacy and security violations and malicious activities by attackers of the system using the integrated technique between privacy techniques with a security technique *Advanced Encryption Standard* (AES). The research objectives are to design an integrated technique between privacy techniques with a security technique for protecting data in the dataset. To implement the integrated technique between privacy techniques with security techniques to protect sensitive user data in the dataset. To validate the integrated technique by comparing the accuracy results before and after applying privacy techniques in the dataset using the data mining techniques. The research methodology consists of three phases. The first phase constitutes determining the suitable dataset and investigating the techniques and understanding of techniques and frameworks related to security and privacy. The second phase involves implementing the integrated technique of privacy techniques with AES. The third phase involves evaluating the classification results after experimenting with the implementation of the processes of the proposed technique. In this research, the security technique that was used is AES, and privacy techniques were differential privacy, k-Anonymity, Sample-uniqueness; and the data mining techniques: Naive Bayes, J48 and Neural Network were used under Weka. Many tests were conducted using Weka Experimenter Environment to classify the data to check the usefulness of the data for analysis purposes. Classification results before and after using privacy techniques were compared. It was found that differential privacy on the gender column of the dataset provides the best results in accuracy and outperforms the k-Anonymity technique and sample-uniqueness technique. However, k-Anonymity and sample-uniqueness techniques on the age and cholesterol columns show better results in accuracy than the differential privacy technique. Besides, it was observed that for classification accuracy on Weka Experimenter Environment, Naïve Bayes presented better results than Neural Network and J48. To avert the dataset from external attackers, AES was used to encrypt the dataset passed through the privacy technique. To increase the safety, the encrypted file the general key involved with the encryption was split into five files and then each subfile and the subkey were attached and stores in five servers. The test results demonstrate the success of using the integrated technique using privacy techniques with the security technique to protect sensitive user data in the dataset. The future work can be focused on enhancing the classification accuracy of the protection schemes by examining more privacy techniques and applying them to big data and carrying out experiments on the big data in the cloud.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

=	Equality
≠	Inequality
>=	Greater than or equal to
<=	Less than or equal to
<	Less than
>	Greater than
×	Multiplication
−	Subtraction
+	Addition
÷	Division
/	Division
%	Percentage
&	Ampersand
Φ	Phi
?	Question mark
XOR	Exclusive OR
( )	Parentheses

LIST OF ABBREVIATIONS

AES	Advanced Encryption Standard
AUC	Area under the curve
CSV	Comma-Separated values
DES	Data Encryption Standard
DSA	Digital Signature Algorithm
ECC	Elliptic Curve Cryptography
EEG	Electroencephalogram
FP	Functional Tree
FPGA	Field Programmable Gate Array
FT	False positives
IB	InfiniBand
ICN	Information Centric Networking
LCA	Latent Class Analysis
LLP	Learning from label proportion
LMT	Logistic model trees
LR	Linear regression
NIST	National Institute of Standards and Technology
PD	Parkinson’s disease
ROC	Receiver Operating Characteristic
RSA	Rivest Shamir Adleman
SCIS	Smartphone-based Compression-Induced Imaging System
SLDC	Shipment Lead time with Delivery Confirmation
SLECR	Shipment Lead time with Empty Container Return
SSL/ TLS	Secure Sockets Layer/Transport Layer Security
SVM	Support Vector Machine
TXT	Text File

LIST OF APPENDICES

REFERENCES

Abani, N. (2018). Caching Strategies for Private and Efficient Content Retrieval in Information-Centric Networks. (Doctor of Philosophy), University of California. Retrieved from https://escholarship.org/uc/item/2w98v2p9

Abbas, O., Mustafa, M. E., & Ibrahim, S. B. (2015). The Role of Data Mining in Information Security. International Journal of Computer (IJC), 17(1), 1-20. Retrieved from https://www.researchgate.net/profile/Osman_Abbas22/publication/295907254_The_Role_of_Data_Mining_in_Information_Security/links/295907256e293212808ae295907265dd295907254cbac295907251fa.pdf

Abood, O. G., & Guirguis, S. K. (2018). A survey on cryptography algorithms. International Journal of Scientific and Research Publications (IJSRP), 8(7), 410-415. doi:http://dx.doi.org/10.29322/IJSRP.8.7.2018.p7978

Abraham, N. T. (2018). A Study of the Characteristics of a Differential Privacy Implementation. (10808383 M.S.), Purdue University, Ann Arbor. Retrieved from https://search.proquest.com/dissertations-theses/study-characteristics-differential-privacy/docview/2054017307/se-2?accountid=29391 ProQuest Dissertations & Theses Global database.

Abusalim, B. Y. (2015). An Efficient Approach For Data Encryption Using Two Keys. (Master's Thesis), Islamic University-Gaza. Retrieved from https://library.iugaza.edu.ps/thesis/117024.pdf

Adetunji, A., Ayinde, A., & Akanbi, C. (2014). Predictive modelling for early maize planting months using j48 mining algorithm. 5(2), 67-72. doi:https://doi:10.5251/ajsir.2014.5.2.67.72

Agarwal, S., Bharti, P., & Pathak, R. K. (2019). Implementation of DES Algorithm in Python. International Journal of Science and Research (IJSR), 8(12), 402-405. doi:https://doi:10.21275/ART20203106

Aggarwal, C. C., & Philip, S. Y. (2008). A general survey of privacy-preserving data mining models and algorithms Privacy-preserving data mining (pp. 11-52): Booktitle. doi:https://doi.org/10.1007/978-0-387-70992-5_2

Akgün, B., & Öğüdücü, Ş. G. (2015). Streaming linear regression on Spark MLlib and MOA. Paper presented at the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. doi:https://doi:10.1145/2808797.2809374

Amin, M. N., & Habib, M. A. (2015). Comparison of different classification techniques using WEKA for hematological data. American Journal of Engineering Research, 4(3), 55-61

Andrews, M., Wilfong, G., & Zhang, L. (2015). Analysis of k-anonymity algorithms for streaming location data. Paper presented at the 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). doi:https://doi:10.1109/INFCOMW.2015.7442434

Awaji, M. H. (2018). Evaluation of Machine Learning Techniques for Early Identification of At-Risk Students. (Dissertation, Doctor of Psychology (PhD)), College of Engineering and Computing, Nova Southeastern University (NSU) Florida.

Brown, E. E. (2019). Adaptable Privacy-preserving Model. (Dissertation, Doctor of Psychology (PhD)), College of Engineering and Computing, Nova Southeastern University (NSU) Florida.

Chan, M., Elsherbini, H., & Zhang, X. (2016, 19-21 Sept. 2016). User density and spatial cloaking algorithm selection: Improving privacy protection of mobile users. Paper presented at the 2016 IEEE 37th Sarnoff Symposium. doi: https://doi.org/10.1109/SARNOF.2016.7846722

Chanyaswad, T. (2018). Privacy-Preserving Machine Learning via Data Compression & Differential Privacy. (doctor of philosophy), Princeton university. Retrieved from https://dataspace.princeton.edu/bitstream/88435/dsp01p2676z32x/1/Chanyaswad_princeton_0181D_12796.pdf

Chatterjee, A., Dhanotia, J., Bhatia, V., & Prakash, S. (2018, 9-11 Feb. 2018). Virtual optical encryption using phase shifted digital holography and RSA algorithm. Paper presented at the 2018 3rd International Conference on Microwave and Photonics (ICMAP). doi:https://doi.org/10.1109/ICMAP.2018.8354560.

Daddala, B., Wang, H., & Javaid, A. Y. (2017, 27-30 June 2017). Design and implementation of a customized encryption algorithm for authentication and secure communication between devices. Paper presented at the 2017 IEEE National Aerospace and Electronics Conference (NAECON). doi: https://doi.org/10.1109/NAECON.2017.8268781

Day, S. C. (2018). A Natural Language Processing and Machine-Learning Based Approach to Authorship Attribution of Tweets. (10841641 Ph.D.), North Carolina Agricultural and Technical State University, Ann Arbor. Retrieved from https://search.proquest.com/dissertations-theses/natural-language-processing-machine-learning/docview/2100700558/se-2?accountid=29391 ProQuest Dissertations & Theses Global database.

Devasia, T., Vinushree, T., & Hegde, V. (2016). Prediction of students performance using Educational Data Mining. Paper presented at the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE). IEEE. https://doi.org/10.1109/SAPIENCE.2016.7684167.

Dichou, K., Tourtchine, V., & Rahmoune, F. (2015, 13-15 Dec. 2015). Finding the best FPGA implementation of the DES algorithm to secure smart cards. Paper presented at the 2015 4th International Conference on Electrical Engineering (ICEE). doi:https://doi.org/10.1109/INTEE.2015.7416749.

Dongpo, Z. (2018, 2018/10). Big Data Security and Privacy Protection. Paper presented at the 8th International Conference on Management and Computer Science (ICMCS 2018). doi: https://doi.org/10.2991/icmcs-18.2018.56.

Elliot, M. (2000). DIS: A new approach to the measurement of statistical disclosure risk. Risk Management, 2(4), 39-48. . doi:https://doi.org/10.1057/palgrave.rm.8240067

Feldhofer, M., Dominikus, S., & Wolkerstorfer, J. (2004, 2004//). Strong Authentication for RFID Systems Using the AES Algorithm. Paper presented at the Cryptographic Hardware and Embedded Systems - CHES 2004, Berlin, Heidelberg. doi:https://doi.org/10.1007/978-3-540-28632-5_26

Golle, P. (2006). Revisiting the uniqueness of simple demographics in the US population. Paper presented at the Proceedings of the 5th ACM workshop on Privacy in electronic society, Alexandria, Virginia, USA. doi:https://doi.org/10.1145/1179601.1179615.

Hathaliya, J. J., & Tanwar, S. (2020). An exhaustive survey on security and privacy issues in Healthcare 4.0. Computer Communications, 153, 311-335. doi:https://doi.org/10.1016/j.comcom.2020.02.018

Hathikal, S., Chung, S. H., & Karczewski, M. (2020). Prediction of ocean import shipment lead time using machine learning methods. SN Applied Sciences, 2(7), 1272. doi:https://doi.org/10.1007/s42452-020-2951-5

Kak, A. (2015). Lecture Notes on “Computer and Network Security”. Lecture, Purdue University. Retrieved from https://engineering.purdue.edu/kak/compsec/NewLectures/Lecture8.pdf

Kala, R., Shukla, A., & Tiwari, R. (2009, 19-21 May 2009). Fast Learning Neural Network Using Modified Corners Algorithm. Paper presented at the 2009 WRI Global Congress on Intelligent Systems. doi:https://doi.org/10.1109/GCIS.2009.429.

Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news. International Journal of Innovative Science, Engineering & Technology, 2(2), 438-446.

Kaur, A. (2017, 21-23 Feb. 2017). A hybrid approach of privacy preserving data mining using suppression and perturbation techniques. Paper presented at the 2017

International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). doi:https://doi.org/10.1109/ICIMIA.2017.7975625

Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science, 57, 500-508. doi:https://doi.org/10.1016/j.procs.2015.07.372

Lobato, E. M. (2017). The Anonymity Engine, Minimizing Quasi-Identifiers to Strengthen k-Anonymity. (Master's Thesis), University of Colorado.

Mancuhan, K. (2017). Data Classification for l-diversity. (10636716 PhD), Purdue University, Ann Arbor. Retrieved from https://search.proquest.com/dissertations-theses/data-classification-i-l-diversity/docview/2017216603/se-2?accountid=29391 ProQuest Dissertations & Theses Global database.

McDonough, J. R. (2018). Utilizing Data Mining Techniques and Ensemble Learning to Predict Development of Surgical Site Infections in Gynecologic Cancer Patients. (Master's Thesis), Binghamton University. Retrieved from https://orb.binghamton.edu/cgi/viewcontent.cgi?article=1054&context=dissertation_and_theses

McMillan, A. (2018). Differential Privacy, Property Testing, and Perturbations. (Doctor of Philosophy), University of Michigan. Retrieved from http://hdl.handle.net/2027.42/143940

Meisner, E. (2003). Naive Bayes Classifier example: Generic. Retrieved from https://www.inf.u-szeged.hu/~ormandi/ai2/06-naiveBayes-example.pdf.

Mivule, K., & Turner, C. (2011). Applying Data Privacy Techniques on Tabular Data in Uganda. ArXiv, abs/1107.3784. Retrieved from https://arxiv.org/abs/1107.3784

Muhammad, A. M. (2016). Advances in Clustering based on Inter-Cluster Mapping. (Doctor of Philosophy), Western Sydney University (Australia). Retrieved from https://pdfs.semanticscholar.org/b7a9/fd0de3f7813a8e41a501af3d83cb5f7d0b92.pdf

Nguyen, A. (2019). Understanding Differential Privacy. Web Page. Retrieved from https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

Pandit, B. (2018). Thesis-generating knowledge base of common behaviour and workflow patterns for secure systems. (Master's Thesis), East Carolina University. Retrieved from http://hdl.handle.net/10342/6732

Pelaez, K. (2019). Latent class analysis and random forest ensemble to identify at-risk students in higher education. Journal of educational data mining, 11, 1. Retrieved from https://par.nsf.gov/servlets/purl/10173401

Pham, M. H. (2018). Signal detection of adverse drug reaction using the adverse event reporting system: literature review and novel methods. (Master's Thesis), University of South Florida. Retrieved from https://scholarcommons.usf.edu/etd/7218

Qinghai, L., Hong, S., & Yingpeng, S. (2015). Privacy-preserving data publishing for multiple numerical sensitive attributes. Tsinghua Science and Technology, 20(3), 246-254. doi:https://doi.org/10.1109/TST.2015.7128936

Rodriguez-Garcia, M., Batet, M., & Sánchez, D. (2017). A semantic framework for noise addition with nominal data. Knowledge-Based Systems, 122, 103-118. doi:https://doi.org/10.1016/j.knosys.2017.01.032

Rogalewicz, M., & Robert, S. (2016). Methodologies of Knowledge Discovery from Data and Data Mining Methods in Mechanical Engineering. Management and Production Engineering Review, 7(No 4), 97-108. doi:https://doi.org/10.1515/mper-2016-0040

Saad, H. (2018). An Integrated Framework of Data Mining and Process Mining to Characterize Quality and Production Processes. (13421028 PhD), State University of New York at Binghamton, Ann Arbor. Retrieved from https://search.proquest.com/dissertations-theses/integrated-framework-data-mining-process/docview/2179565582/se-2?accountid=29391 ProQuest Dissertations & Theses Global database.

Schriner, J. (2018). Building Test Anonymity Networks in a Cybersecurity Lab Environment. (Master's Thesis), City University of New York (CUNY). Retrieved from https://academicworks.cuny.edu/jj_etds/80/

Sewaiwar, P., & Verma, K. K. (2015). Comparative study of various decision tree classification algorithm using WEKA. International Journal of Emerging Research in Management &Technology, 4(2015), 2278-9359. Retrieved from https://www.semanticscholar.org/paper/Comparative-Study-of-Various-Decision-Tree-Using-Sewaiwar-Verma/2277fd2277b2824fa44053551eb44053558f44053551a44053592a44053550bc44053590abe44053984#citing-papers.

Shakil, K. A., Anis, S., & Alam, M. (2015). Dengue disease prediction using weka data mining tool. arXiv preprint arXiv:1502.05167, Retrieved from https://arxiv.org/ftp/arxiv/papers/1502/1502.05167.pdf.

Sharma, A., & Sahay, S. K. (2016). An effective approach for classification of advanced malware with high accuracy. arXiv preprint arXiv:1606.06897, Retrieved from https://arxiv.org/ftp/arxiv/papers/1606/1606.06897.pdf.

Sharma, A., & Sahay, S. K. (2018). An investigation of the classifiers to detect android malicious apps Information and Communication Technology (pp. 207-217): Springer. doi:https://doi.org/10.1007/978-981-10-5508-9_20

Singandhupe, A. R. (2017). Securing a UAV Using Features from an EEG Signal. (Master's Thesis), University of Nevada, Reno. Retrieved from http://hdl.handle.net/11714/2750

Sriramoju, S. B. (2017). Analysis and Comparison of Anonymous Techniques for Privacy Preserving in Big Data. International Journal of Advanced Research in Computer and Communication Engineering, 6(12), 2278-1021. Retrieved from https://www.researchgate.net/profile/Shoban_Sriramoju/publication/322096247_IJARCCE_Analysis_and_Comparison_of_Anonymous_Techniques_for_Privacy_Preserving_in_Big_Data/links/322096245a322096244af377458515f322096246b320531216/IJARCCE-Analysis-and-Comparison-of-Anonymous-Techniques-for-Privacy-Preserving-in-Big-Data.pdf.

Triguero, I., Maillo, J., Luengo, J., García, S., & Herrera, F. (2016, 15-18 Dec. 2016). From Big Data to Smart Data with the K-Nearest Neighbours Algorithm. Paper presented at the 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). doi:https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.177

Tyler, J. (2016). Don't be your own worst enemy: protecting your organisation from inside threats. Computer Fraud & Security, 2016(8), 19-20. doi:https://doi.org/10.1016/S1361-3723(16)30063-X

Wang, Z. (2018). Smartphone-Based Compression-Induced Imaging System Data Security. (Master's Thesis), Temple University. Libraries. doi:http://dx.doi.org/10.34944/dspace/3772

Witten, I. H., Frank, E., & Mark, A. (2016). Data Mining: Practical machine learning tools and techniques. Books (pp. 654): Morgan Kaufmann. doi: https://doi.org/10.1016/C2015-0-02071-8

Yan, X. (2018). Privacy Preserving Bag Preparation for Learning from Label Proportion. (10981539 M.S.), Illinois Institute of Technology, Ann Arbor. Retrieved from https://search.proquest.com/dissertations-theses/privacy-preserving-bag-preparation-learning-label/docview/2189095105/se-2?accountid=29391 ProQuest Dissertations & Theses Global database.

Yao, S. (2018, 8-9 Dec. 2018). An Improved Differential Privacy K-Means Algorithm Based on MapReduce. Paper presented at the 2018 11th International Symposium on Computational Intelligence and Design (ISCID). doi:https://doi.org/10.1109/ISCID.2018.10133.

Zimmeck, S. (2017). Using machine learning to improve internet privacy. (Doctor of Philosophy), Columbia University. doi: https://doi.org/10.7916/D8862N5F.

Zyskind, G., Nathan, O., & Pentland, A. (2015, 21-22 May 2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. Paper presented at the 2015 IEEE Security and Privacy Workshops. doi:https://doi.org/10.1109/SPW.2015.27