International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

INSURANCE FRAUD DETECTION USING MACHINE LEARNING ON CLASS-IMBALANCED DATASETS WITH MISSING VALUES

Uzma Fatima[1], Ms. Lubna Nausheen[2], Ms. Sadaf Jahan[3]
[1]PG Scholar, Department of AI & DS, Shadan Women's College of Engineering and Technology, Hyderabad, uzmafatima100220003@gmail.com
[2]Assoc. Professor, Department of CSE, Shadan Women's College of Engineering and Technology lubnanausheen07@gmail.com
[3]Assoc. Professor, Department of CSE, Shadan Women's College of Engineering and Technology Sadafcse6@gmail.com

ABSTRACT

Insurance fraud is a major problem for insurers, especially in the vehicle insurance industry. It affects pricing tactics and causes financial losses. Class imbalance, when fraudulent claims are far less common than legitimate claims, frequently affects fraud detection models, and missing data makes the task even more difficult. Two vehicle insurance datasets—a conventional dataset and an Egyptian real-life dataset—are used in this study to address these problems. In order to improve the accuracy and prediction potential of the model, the AdaBoost Classifier is incorporated into the suggested methodology, which also addresses missing data and class imbalance. The findings show that correcting class imbalance is essential to enhancing model performance, and handling missing data also helps to produce predictions that are more trustworthy. By increasing prediction accuracy and decreasing overfitting, which is frequently a problem in fraud detection models, the AdaBoost Classifier greatly outperforms current methods. This study offers insightful information about how enhancing data quality and utilising cutting-edge algorithms like AdaBoost can improve fraud detection systems, ultimately resulting in more successful identification of fraudulent claims. These improvements can greatly help insurance businesses make better decisions, lower financial losses, and improve pricing strategies.

INTRODUCTION

The financial sector is progressively contending with an escalation in fraudulent activities, particularly within the domain of auto insurance [1]. The increase in the number of vehicles and corresponding insurance policies has been paralleled by a rise in fraudulent claims, resulting in significant economic burdens. Auto insurance fraud is estimated to impose an annual financial burden exceeding $40 billion, resulting in an incremental cost of approximately $400 to $700 in premiums for the average American household [2]. Significantly, research shows that while 21% to 36% of auto insurance claims may involve fraud, only about 3% receive legal scrutiny, showing that many fraudulent claims. remain undetected and potentially lead to larger undisclosed financial losses [3]. The insurance industry is huge, with more than 7,000 companies and over $1 trillion in yearly premiums, making it harder to fight fraud [4]. Auto insurance fraud typically involves exaggerated claims, false claims [5], intentional injury [6], multiple claims, or other forms of misrepresenting insurance-related information [7]. However, research shows that fraudulent behavior among insurers and auto repair shops is related to each other, resulting from their mutual perceptions of each other's actions. For example, if both parties engage in fraudulent activities, both will benefit. However, each party refrains from such activities, neither can benefit and the risk of detection increases [8]. Fraudsters often believe that insurance companies and police lack the expertise to detect fraud, leading to low interest in such crimes [9]. This belief is reinforced by two factors:

In the field of auto insurance fraud detection, three main challenges are commonly encountered: unbalanced datasets, a high number of false positives, and ML models with untuned hyperparameters. The imbalance between fraudulent and non-fraudulent samples in datasets can lead to biased models and poor fraud detection [15]. To address this, two main methods are used: oversampling and undersampling. Oversampling techniques like synthetic minority oversam-pling technique (SMOTE) [16] increase the number of fraudulent cases, while undersampling methods, such as random undersampling [17], reduce the number of non-fraudulent cases. Both methods aim to balance the dataset and improve the performance and reliability of predictive models [18].

Additionally, a high number of false positives can occur, meaning legitimate claims are incorrectly flagged as fraud-ulent. This not only leads to customer dissatisfaction but also increases operational costs for insurance companies. Advanced anomaly detection

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

techniques and robust vali-dation processes can help mitigate this issue by refining model accuracy [19]. Moreover, ML models with untuned hyperparameters often result in suboptimal performance. Hyperparameter tuning is essential for enhancing the accu-racy and efficiency of fraud detection models [20]. Properly tuned models can better differentiate between fraudulent and non-fraudulent claims, thereby reducing both false positives and false negatives. Our motivation is to address these challenges by incorporating balancing techniques, an ensemble learning model, and a metaheuristic-based hyperparameter tuning algorithm to enhance the accuracy and robustness of auto insurance fraud detection systems. This approach aims to develop a comprehensive solution that effectively tackles the limitations of existing methods.

Using an undersampled dataset as the input for ensemble learning, which integrates multiple base classifiers, can effectively mitigate the challenges of unbalanced datasets and high false positives in auto insurance fraud detection [21]. By reducing the number of non-fraudulent cases through techniques like random undersampling, the dataset becomes more balanced, reducing bias and enhancing model reliability. Ensemble learning, which combines multiple base classifiers with distinct features, further improves detection accuracy [22]. When these base classifiers have their hyperparameters tuned with metaheuristic algorithms, their individual performances are optimized. This optimized performance, along with the ensemble method employing weighted voting between the results of these base classifiers, ensures that the most accurate predictions are prioritized. This method collectively addresses the three primary challenges unbalanced datasets, high false positives, and untuned hyperparameters resulting in a more effective fraud detection system.

This study introduces an ensemble learning framework designed for the detection of auto insurance fraud, utilizing a publicly available dataset of 15,420 car insurance claims from Kaggle, recorded between 1994 and 1996. The dataset includes 32 features, combining categorical (e.g., accident area, policy type) and numerical (e.g., age, deductible) data, offering a rich foundation for analysis. However, with only 6% of claims classified as fraudulent and 94% as legitimate, the dataset presents a significant class imbalance, posing a key challenge for fraud detection. To address this, the proposed framework incorporates three base classifiers—random forest (RF), extreme gradient boosting (XGBoost), and support vector machines (SVM)—combined into an ensemble model with a weighted voting strategy, and leverages balancing techniques to mitigate the impact of the imbalance.

The hyperparameters of each base classifier are fine-tuned using a metaheuristic optimization algorithm named binary quantum-based avian navigation optimizer algorithm (BQANA). For comparative analysis, additional hyperparam-eter tuning methods, including simulated annealing (SA) and genetic algorithms (GA), are employed, with their evaluation outcomes assessed. To further explore the impact of data imbalance, the study generates four subsets with varying ratios of fraudulent to legitimate claims. Each classifier is assigned a specific weight based on its performance during classification, enhancing the accuracy of fraud identification in the final ensemble voting and prediction stages. By integrating these techniques, the framework effectively addresses the challenges of unbalanced datasets, high false positives, and untuned hyperparameters, providing a robust solution for fraud detection in the auto insurance sector.

EXISTING SYSTEM OF PROJECT

Conventional machine learning techniques like Random Forests, Decision Trees, and Logistic Regression are the mainstay of current insurance fraud detection systems. When working with extremely unbalanced datasets, when fraudulent claims make up a very small portion of all claims, these models encounter serious difficulties and produce predictions that are skewed in favour of valid cases. Furthermore, the dataset's inconsistencies, noisy records, and missing values lower the training data's quality and have a detrimental impact on prediction accuracy. Overfitting, in which models perform well on training data but are unable to successfully generalise to new and unseen claims, is another significant constraint. While methods like under sampling and oversampling are employed to remedy class imbalance, they frequently lead to data duplication or information loss and do not always you the best possible fraud detection results. As a result, these restrictions lower the efficacy, robustness, and dependability of current fraud detection systems.

insurance companies frequently pay illegitimate claims, and insurance fraud is rarely prosecuted [4].Traditional manual verification methods are inadequate, often

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

producing false alarms. Consequently, it is essential to employ precise and intelligent methods, such as data mining, to effectively detect and reduce fraudulent cases[10].Various methods are employed to detect auto insurance fraud, including statistical techniques [11], machine learning (ML) [12], and deep learning (DL) [13]. Statistical methods, such as regression and hypothesis testing, are used to identify and analyze anomalies in auto insurance data. Moreover, ML methods use intelligent algorithms to detect fraudulent activities in real-time by analyzing relevant data, while DL methods leverage artificial neural networks (ANN) [14] to automatically identify complex patterns and features in large datasets. Each of these methods has its own pros and cons when applied to different types of data, making the choice of method crucial for effective fraud detection [12].

ALGORITHM:

Traditional insurance fraud detection systems typically use machine learning algorithms like Logistic Regression, Decision Trees, and Random Forest. These algorithms, while effective for general classification tasks, struggle with imbalanced datasets where fraudulent claims are much fewer than legitimate ones. They are also limited by issues like overfitting, where the model learns noise from the training data, and missing values, which can degrade model performance. Despite using techniques like under sampling or oversampling to handle class imbalance, these models often fail to achieve optimal accuracy and robustness.
Additionally, many existing systems struggle with overfitting, where models memorize specific patterns in the training data, leading to poor generalization on unseen data. While Random Forest and similar algorithms are used, they still face challenges with large, imbalanced datasets, particularly when handling complex fraud patterns. The lack of advanced techniques to address both class imbalance and missing values diminishes the effectiveness of these systems, making it difficult to maintain a high fraud detection rate without sacrificing accuracy for legitimate claims.

PROPOSED SYSTEM

The AdaBoost Classifier, an ensemble technique that increases prediction performance by combining numerous weak classifiers to build a strong, correct model, is incorporated into the suggested system to improve insurance fraud detection. AdaBoost is particularly useful for managing noisy or incomplete datasets since it may minimise overfitting, ensuring that the model performs well when applied to new data. In order to improve the detection of fraudulent claims that are under-represented in the dataset, this classifier is also used with methods like SMOTE (Synthetic Minority Over-sampling Technique) to address the issue of class imbalance.
To successfully handle missing data, the suggested approach also makes use of more reliable data pretreatment techniques. By ensuring that the model operates with cleaner, more comprehensive datasets, these preprocessing methods increase overall prediction accuracy. The model becomes more dependable, scalable, and effective when AdaBoost is combined with these sophisticated data handling techniques. The suggested approach surpasses conventional techniques in terms of accuracy and robustness by addressing both class imbalance and missing data, offering a more efficient solution for fraud detection in the insurance sector.

SYSTEM ARCHITECTURE:

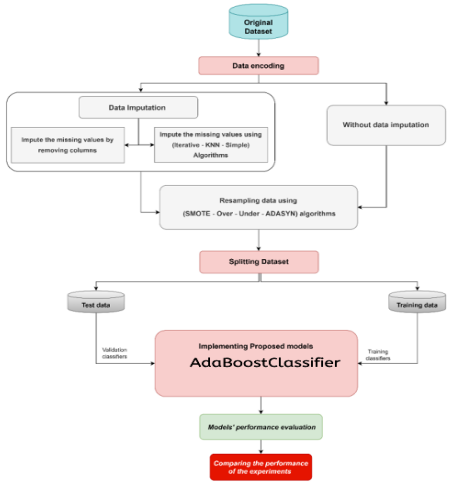

Fig. 1. System Architecture

ALGORITHM:

The proposed system leverages the AdaBoost Classifier, an ensemble method that combines weak learners to create a strong classifier, improving model robustness and accuracy. AdaBoost's ability to reduce overfitting makes it especially useful for noisy and incomplete datasets. Additionally, techniques like SMOTE are used to address class imbalance, ensuring better performance in detecting fraudulent claims. The system also includes advanced data preprocessing methods to handle missing values, ensuring a more complete dataset and enhancing model performance, making it a more reliable solution for fraud detection.

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

In addition to AdaBoost, the proposed system employs the Synthetic Minority Over-sampling Technique (SMOTE) to address the class imbalance problem by generating synthetic samples for the minority class, which is critical for improving the fraud detection rate. The integration of advanced data preprocessing methods ensures that missing values are effectively handled, preventing data gaps from negatively impacting the model's performance. These combined strategies—AdaBoost for classification, SMOTE for data balancing, and advanced preprocessing for handling missing data—lead to a more accurate, reliable, and robust system for insurance fraud detection, significantly outperforming traditional methods in terms of both prediction accuracy and generalization.

DATASET PREPARATION:

The current research utilizes a comprehensive dataset obtained from an insurance company's car claim records. This dataset, available on the Kaggle platform (Link to Dataset), contains a substantial collection of 15,420 insurance claims recorded from January 1994 to December 1996 [49]. For accessibility and reproducibility, details on data availability are provided in the Data and Code Availability section at the end of the paper. The dataset contains 32 features, in addition to a target feature, as detailed in Table 2. Each sample in the dataset is shown by a binary target feature, which is essential for classifying the claims into fraudulent or non-fraudulent categories. The dataset is identified by its variety, and contains 25 categorical and 8 numerical features, providing a rich foundation for analysis. 43000 The employed dataset shows a clear imbalance in the distribution of fraudulent cases, which account for only 6% (923 claims) of the total, in contrast to the 94% (14,497 claims) that are non-fraudulent. This data imbalance shows a fundamental challenge in fraud detection and is a significant factor to consider when selecting appropriate modeling and evaluation techniques. In preparation for the analysis, the dataset underwent a partitioning process to facilitate the model's training, testing, and validation phases. Specifically, the data was randomly divided into three subsets: 60% was allocated for training the model, enabling it to learn and adapt to the patterns within the data; 20% was reserved for testing, providing an evaluation of the model's predictive performance on unseen data; and the remaining 20% was allocated for validation purposes, allowing an additional layer of evaluation to fine-tune the model's hyperparameters. This partitioning strategy is essential for the valid evaluation of fraud detection models. All codes were implemented in Python 3.6.15. The computations were performed on a system equipped with an Intel Core i3-3220 processor and 4 GB DDR RAM.

TECHNIQUES USED IN PROJECT

To enhance the detection of fraudulent insurance claims, the suggested insurance fraud detection system combines sophisticated data preprocessing and machine learning approaches. To handle missing values, remove inconsistencies, and improve overall data quality, data preparation techniques are first used. The Synthetic Minority Over-sampling Technique (SMOTE) is used to create synthetic samples for the minority class in order to address the issue of class imbalance by balancing the distribution of false and true claims.

This study's primary classification model is the AdaBoost (Adaptive Boosting) Classifier, an ensemble learning technique that builds a powerful predictive model by combining several weak learners. AdaBoost allows the model to concentrate on challenging situations and enhance classification performance by iteratively adjusting the weights of misclassified instances. This method successfully lessens overfitting and improves the model's capacity to generalise to new data.

Exploratory Data Analysis (EDA) is also used to analyse feature distributions, spot trends, and comprehend the connections between variables. Standard performance criteria including as Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC) are used to assess the efficacy of the suggested model. When these methods are combined, a strong, precise, and trustworthy fraud detection framework is produced. capable of handling imbalanced datasets and missing values in insurance claim records.

MODEL TRAINING

Following data pretreatment, which includes managing missing values and balancing the dataset using the Synthetic Minority Over-sampling Technique (SMOTE), the model training procedure starts. The pre-processed dataset is split into training and testing sets, usually in an 80:20 ratio, with 20% set aside for performance assessment and 80% used for model training.

The main machine learning technique for fraud detection in the suggested system is the AdaBoost (Adaptive Boosting) Classifier. AdaBoost builds a robust predictive model during training by combining several weak learners, typically decision tree classifiers. All

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

training samples are first given equal weights by the algorithm, which then iteratively increases the weights of cases that were incorrectly identified This increases the model's learning capacity and overall prediction accuracy by allowing it to concentrate more on challenging-to-classify samples.

Each weak learner is gradually trained based on the mistakes of the prior learner during the training process, which entails several boosting iterations. To get the best results, hyperparameters like the number of estimators and learning rate are optimised. The ensemble model produces a reliable fraud detection system that can handle imbalanced datasets and noisy data by continuously updating its predictions and minimising classification mistakes.

Following training, the model is verified using the test dataset and assessed using a number of performance metrics, such as ROC-AUC Score, Accuracy, Precision, Recall, and F1-Score. The experimental findings show that the AdaBoost Classifier reduces overfitting and enhances generalisation performance on unknown data while successfully identifying fraudulent insurance claims.

RESULT: The proposed insurance fraud detection system was successfully implemented and evaluated using insurance claim datasets containing both legitimate and fraudulent records. Data preprocessing techniques were applied to handle missing values and improve data quality, while the Synthetic Minority Over-sampling Technique (SMOTE) was used to address the class imbalance problem. After preprocessing, the AdaBoost Classifier was trained and tested on the prepared dataset.

The experimental results demonstrate that the proposed model achieved superior performance compared to traditional machine learning algorithms. The AdaBoost Classifier effectively identified fraudulent claims with high accuracy while maintaining a strong balance between precision and recall. The use of SMOTE significantly improved the detection rate of minority class instances, reducing the bias toward legitimate claims. Furthermore, the model showed improved generalization capability and reduced overfitting, enabling reliable performance on unseen data.

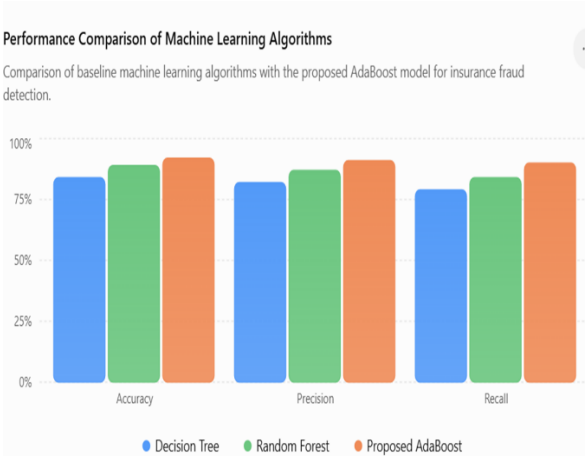

Fig. 2. Performance Comparison of ML Models for Insurance Fraud Detection

Metric	Decision Tree	Random Forest	Proposed AdaBoost
Accuracy (%)	84	89	92
Precision (%)	82	87	91
Recall (%)	79	84	90

Table 1: Performance Comparison of ML algorithms

Performance evaluation was conducted using standard metrics such as Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC Score. The results indicate that the proposed approach provides a robust and efficient solution for insurance fraud detection. By accurately identifying fraudulent claims and minimizing false predictions, the system can assist insurance companies in reducing financial losses, improving operational efficiency, and enhancing decision-making processes. Overall, the integration of AdaBoost and advanced preprocessing techniques proved effective in developing a reliable fraud detection framework.

Fig 3: Register or Login

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

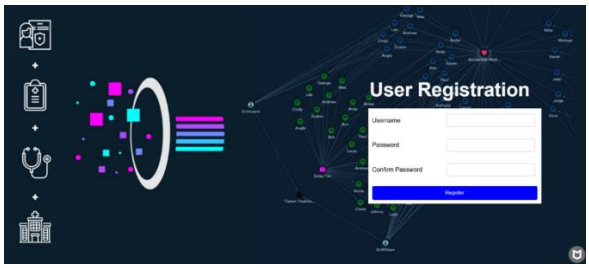

Fig 4: User Registration

Fig 5: Login Page

Fig 6: Input Data

Fig 7: Data Entered

Fig 8: Prediction Result: Legitimate

DISCUSSION AND CONCLUSION

The study's findings show how well machine learning methods work to identify false insurance claims. The existence of class imbalance and missing values in the dataset is one of the main obstacles to insurance fraud detection. Because there are few instances of fraud and low-quality data, traditional machine learning algorithms frequently have trouble correctly identifying false claims. The suggested solution included sophisticated data preprocessing methods, SMOTE for dataset balance, and the AdaBoost Classifier for classification in order to overcome these obstacles..

The experimental findings indicate that the AdaBoost Classifier significantly improved fraud detection performance by focusing on misclassified instances during the training process. The application of SMOTE enhanced the representation of the minority class, enabling the model to learn fraud patterns more effectively. Furthermore, proper handling of missing values improved data quality and contributed to more reliable predictions. The model demonstrated improved accuracy, precision, recall, and F1-score while reducing overfitting and achieving better generalization on unseen data. These improvements highlight the suitability of ensemble learning techniques for insurance fraud detection applications.

In conclusion, by integrating efficient data preprocessing, class balancing, and the AdaBoost ensemble learning algorithm, the suggested insurance fraud detection system effectively overcomes the shortcomings of current approaches. The created model offers a solid and trustworthy framework for more accurately and efficiently recognising fraudulent insurance claims. The technology can assist insurance businesses reduce financial losses, improve decision-making, and bolster risk management strategies by strengthening fraud detection capabilities. The findings verify that combining AdaBoost and SMOTE provides a workable and effective way to manage unbalanced

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

datasets and enhance the general effectiveness of insurance fraud detection systems.

FUTURE ENHANCEMENT

By using cutting-edge machine learning and deep learning techniques, the suggested insurance fraud detection system can be further improved in terms of prediction accuracy and adaptability. In order to help insurance companies spot and stop fraudulent claims as they happen, future development might concentrate on incorporating real-time fraud detection capabilities. The model's capacity to generalise across a range of fraud scenarios can also be enhanced by adding bigger and more varied datasets from various insurance industries.

The use of deep learning models, such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and hybrid ensemble techniques, to identify intricate fraud patterns that conventional machine learning algorithms could miss is another possible improvement. Furthermore, Explainable Artificial Intelligence (XAI) methods like SHAP and LIME can be used to make fraud prediction transparent and interpretable, assisting insurance analysts in comprehending the variables affecting the model's choices. The system can also be extended by integrating external data sources such as customer behavioural data, social media information, telematics data with past claim records to increase the precision of fraud detection. To guarantee that the system continues to be effective against changing fraud tactics, automated model retraining and adaptive learning processes can be put in place. These improvements will help create a more sophisticated, scalable, and trustworthy framework for detecting insurance fraud that can reduce monetary losses and boost the insurance sector's operational effectiveness.

REFERENCES

[1] R. Canetti, U. Feige, O. Goldreich, and M. Naor, ‘‘Adaptively secure multi-party computation,’’ in *Proc. 28th Annu. ACM Symp. Theory Com- put.*, Philadelphia, PA, USA, 1996, pp. 639–648. [Online]. Available: http://portal.acm.org/citation.cfm?doid=237814.238015

[2] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, ‘‘Deniable encryption,’’ in *Proc. 17th Annu. Int. Cryptol. Conf.* (Lecture Notes in Computer Sci- ence), vol. 1294. Santa Barbara, CA, USA: Springer, 1997, pp. 90–104.

[3] A. Juels and T. Ristenpart, ‘‘Honey encryption: Security beyond the brute-force bound,’’ in *Advances in Cryptology—EUROCRYPT* (Lec- ture Notes in Computer Science), vol. 8441, D. Hutchison, T. Kanade,

[4] J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, P. Q. Nguyen, and E. Oswald, Eds. Berlin, Germany: Springer, 2014, pp. 293–310, doi: 10.1007/978-3-642-55220-5_17.

[5] M. Durmuth and D. M. Freeman, ‘‘Deniable encryption with negli- gible detection probability: An interactive construction,’’ in *Advances in Cryptology—EUROCRYPT* (Lecture Notes in Computer Science), vol. 6632, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern,

[6] J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, and K. G. Paterson, Eds. Berlin, Germany: Springer, 2011, pp. 610–626, doi: 10.1007/978-3-642-20465-4_33.

[7] A. O’Neill, C. Peikert, and B. Waters, ‘‘Bi-deniable public-key encryption,’’ in *Advances in Cryptology—CRYPTO* (Lecture Notes in Computer Science), vol. 6841, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg,

[8] F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. P. Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, and P. Rogaway, Eds. Berlin, Germany: Springer, 2011, pp. 525–542, doi: 10.1007/978-3-642-22792-9_30.

[9] M. Klonowski, P. Kubiak, and M. Kutylowski, ‘‘Practical deniable encryption,’’ in *SOFSEM 2008: Theory and Practice of Computer Science* (Lecture Notes in Computer Science), V. Geffert, J. Karhumäki, A. Bertoni, B. Preneel, P. Návrat, and M. Bieliková, Eds. Berlin, Germany: Springer, 2008, pp. 599–609.

[10] P. Gasti, G. Ateniese, and M. Blanton, ‘‘Deniable cloud storage: sharing files via public-key deniability,’’ in *Proc. 9th Annu. ACM Workshop Pri- vacy Electron. Soc.*, Chicago, IL, USA, 2010, p. 31. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1866919.1866925

[11] M. H. Ibrahim, ‘‘A method for obtaining deniable public-key encryption,’’ *Int. J. Netw. Secur.*, vol. 8, no. 1, pp. 1–9, 2009.

[12] P. Chi and C. Lei, ‘‘Audit-free cloud storage via deniable attribute- based encryption,’’ *IEEE Trans. Cloud Comput.*, vol. 6, no. 2, pp. 414–427, Apr. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7090980/

[13] R. Canetti, S. Park, and O. Poburinnaya, ‘‘Fully deniable interactive encryption,’’ in *Advances in*

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Cryptology—CRYPTO (Lecture Notes in Computer Science), vol. 12170, D. Micciancio and T. Ristenpart, Eds. Cham, Switzerland: Springer, 2020, pp. 807–835, doi: 10.1007/978-3-030- 56784-2_27.

[14] C. Dwork, M. Naor, and A. Sahai, ‘‘Concurrent zero-knowledge,’’ *J. ACM*, vol. 51, no. 6, p. 851–898, Nov. 2004, doi: 10.1145/1039488.1039489.

[15] M. Naor, ‘‘Deniable ring authentication,’’ in *Advances in Cryptology— CRYPTO* (Lecture Notes in Computer Science), vol. 2442, G. Goos,

[16] J. Hartmanis, J. van Leeuwen, and M. Yung, Eds. Berlin, Germany: Springer, 2002, pp. 481–498, doi: 10.1007/3-540-45708-9_31.

[17] R. W. Zhu, D. S. Wong, and C. H. Lee, ‘‘Cryptanalysis of a suite of deniable authentication protocols,’’ *IEEE Commun. Lett.*, vol. 10, no. 6, pp. 504–506, Jun. 2006.

[18] A. Fiat and M. Naor, ‘‘Broadcast encryption,’’ in *Proc. Annu. Int. Cryptol. Conf.*, Jan. 1993, pp. 480–491.

[19] J. Li, Y. Wang, Y. Zhang, and J. Han, ‘‘Full verifiability for out- sourced decryption in attribute-based encryption,’’ *IEEE Trans. Services Comput.*, vol. 13, no. 3, pp. 478–487, May 2020. [Online]. Available: https://ieeexplore.ieee.org/document/7936626/

[20] J. Li, Q. Yu, and Y. Zhang, ‘‘Hierarchical attribute-based encryption with continuous leakage-resilience,’’ *Inf. Sci.*, vol. 484, pp. 113–134, May 2019.

[21] J. Li, W. Yao, Y. Zhang, H. Qian, and J. Han, ‘‘Flexible and fine-grained attribute-based data storage in cloud computing,’’ *IEEE Trans. Services Comput.*, vol. 10, no. 5, pp. 785–796, Sep. 2017.

[22] S. Reddy, P. S. Reddy, and P. Sravanthi, ‘‘Audit free cloud stor- age via deniable attribute base encryption for protecting user pri- vacy,’’ *Int. J. Sci. Eng. Technol. Res.*, vol. 5, no. 17, pp. 3449–3451, 2016.

[23] R. Bassous, R. Bassous, H. Fu, and Y. Zhu, ‘‘Ambiguous multi-symmetric cryptography,’’ in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 7394–7399.

[24] A. Chakraborti, C. Chen, and R. Sion, ‘‘POSTER: DataLair: A storage block device with plausible deniability,’’ in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1757–1759.

[25] C. Chen, A. Chakraborti, and R. Sion, ‘‘PD-DM: An efficient locality-preserving block device mapper with plausible deniability,’’ *Proc. Privacy Enhancing Technol.*, vol. 2019, no. 1, pp. 153–171, Jan. 2019.

[26] C. Chen, X. Liang, B. Carbunar, and R. Sion, ‘‘SoK: Plausibly deni- able storage,’’ *Proc. Privacy Enhancing Technol.*, vol. 2022, no. 2, pp. 132–151, Apr. 2022. [Online]. Available: https://petsymposium. org/popets/2022/popets-2022-0039.php

[27] E.-O. Blass, T. Mayberry, G. Noubir, and K. Onarlioglu, ‘‘Toward robust hidden volumes using write-only oblivious RAM,’’ in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 203–214, doi: 10.1145/2660267.2660313.

[28] C. Hargreaves and H. Chivers, ‘‘Detecting hidden encrypted vol- umes,’’ in *Communications and Multimedia Security*, B. De Decker and I. Schaumuller-Bichl, Eds. Berlin, Germany: Springer, 2010, pp. 233–244