ISSN: 2754-6705

Journal of Mathematical & Computer Applications

SCIENTIFIC Research and Community

Review Article

Open Access

Data Obfuscation Techniques for Realtime Data Pipelines

Girish Ganachari

USA

ABSTRACT

Obfuscation safeguards healthcare, banking, manufacturing, and telecommunications real-time data. While prone to incursions, breaches, and attacks, real-time data pipelines acquire sensitive data and make speedy choices. Obfuscating sensitive data permits inquiry but prohibits access. Obfuscation methods are examined to improve data privacy, reduce breach risk, and comply with legislation. It involves load processing, demand management, and data utility. The study emphasises the need for ongoing research and development to adapt obfuscation tactics to new threats and data situations.

***Corresponding author**

Girish Ganachari, USA.

Received: May 06, 2022; **Accepted:** May 13, 2022, **Published:** May 21, 2022

Keywords: Data Obfuscation, Real-Time Data Pipelines, Data Privacy, Data Security, Healthcare Data, Financial Data, Industrial Control Systems, Regulatory Compliance

Introduction

Industry, finance, healthcare, and telecommunications organisations use real-time data pipelines to analyse and transmit data and make decisions. Data processing makes these pipelines susceptible to hackers, breaches, and unauthorised access. Live data pipelines must encrypt sensitive data to avoid security breaches and unauthorised access while maintaining data structure integrity for allowed use. This research shows how data obfuscation technologies improve privacy, breach minimisation, and regulatory compliance in real-time pipelines across multiple industries. Research addresses data usefulness, scalability, and computational complexity. This study also explores developing obfuscation tactics to fight new assaults and react to data changes.

Applications of Data Obfuscation in Real-time Data Pipelines

Companies conceal important data for rapid processing and analysis. These sections examine healthcare, banking, and industrial control system data obfuscation and security and privacy issues.

Healthcare

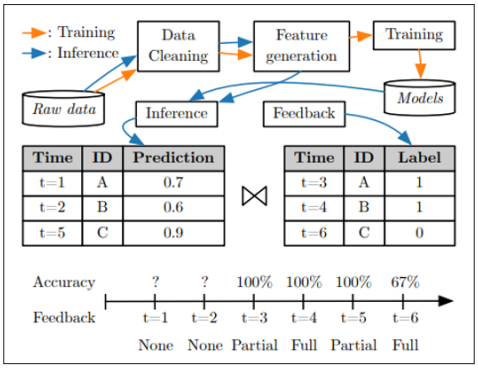

Figure 1: A Generic end-to-end ML Pipeline

Medical data management companies require high security and privacy. Clinicians must use telemedicine and health analytics to obscure patient data while analysing it for privacy. Software-obfuscated medical equipment transmit patient data in IoMT [5]. A breach might compromise patient privacy and safety, making data security crucial. Obfuscating software preserves IoMT functioning and protects patient data. ETL-based biodata privacy protects sensitive patient data during integration and analysis via extensive obfuscation [17]. Online patient monitoring and consulting apps require data obfuscation for privacy, integrity, and transmission. Regulatory compliance and public trust in telemedicine need this method.

B. Financial Services

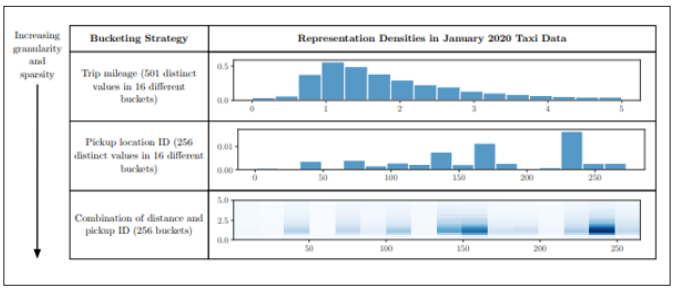

Figure 2: Bucketing Strategies, Normalized to Show Bucket Density

Financial services need real-time data processing for compliance, fraud detection, and transaction monitoring. Financial data security and investigation speed need obfuscation. The use of computers that implement encryption in order to guarantee the secure management of sensitive financial information [7]. Obfuscation and encryption secure transaction data and simplify real-time analysis. Financial firms must obfuscate sensitive data extensively [9]. Obfuscation protects financial data and detects fraud live. The system protects sensitive data and lets institutions monitor transactions in real time to avoid fraud and losses.

Citation: Girish Ganachari (2022) Data Obfuscation Techniques for Realtime Data Pipelines. Journal of Mathematical & Computer Applications. SRC/JMCA-E101. DOI: doi.org/10.47363/JMCA/2022(1)E101

Industrial Control Systems

Manufacturers, power plants, and water distribution networks require industrial controls. Industrial control demands real-time data processing. Obfuscating operational data allows real-time management and protects it from online threats. Mamoud et al. [20] propose blockchain and obfuscation-based water distribution network data aggregation. Operation data influences water distribution system reliability. For critical infrastructure data, Mahmoud et al. conceal real-time monitoring and control system data. Hiding operational data in water distribution systems may increase security and dependability. Electrosense tracks spectrum. Rajendran et al. [21] say it obfuscates spectrum data for fast processing and analysis. Obfuscation protects huge, real-time monitoring systems, says Electrosense. The network conceals spectrum data from inexpensive sensors for privacy. This method protects data and monitors and analyses spectrum in real time for spectrum management. Industrial control systems must hide operating data from hackers. Obfuscating industrial control system operating data avoids cyberattacks and ensures security. Critical infrastructure failures may harm the economy and public safety.

Benefits of Data Obfuscation Techniques

Data obfuscation strategies are vital for real-time data pipeline security and privacy due to their benefits. Better data privacy, fewer breaches, and legal compliance are benefits.

Enhanced Data Privacy

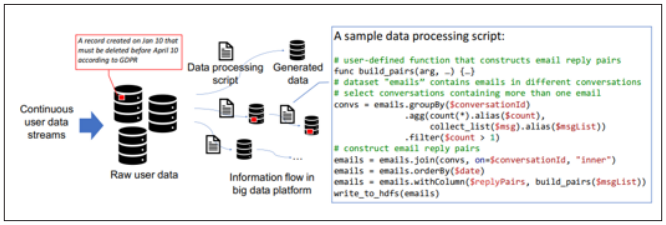

Figure 3: A Motivating Scenario for Fine-Grained Taint Tracking

Protecting individually identifiable information using data obfuscation technology is beneficial. Data obfuscation hides sensitive data from unauthorised outsiders. Healthcare, banking, and telecoms companies that handle sensitive data need this. Shankar and Parameswaran [1] say machine learning pipeline data privacy requires obfuscation. To train and enhance machine learning models, these pipelines examine sensitive data. Without obfuscation, this material is vulnerable to privacy breaches and illicit access.

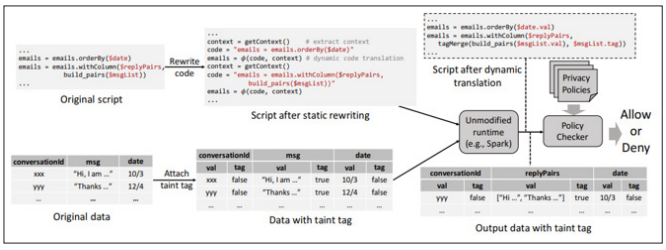

Figure 4: TaintStream Workflow

Obfuscation safeguards sensitive data against unwanted access. Obfuscating patient data in telemedicine and health data analytics apps protects it. Medical histories, treatment data, and personal identifiers must be protected. Obfuscation may enhance patient care, efficiency, and data security [5]. Banks conceal transaction data to avoid fraud and unauthorised access. Banks save account numbers, IDs, and transaction data. Financial institutions may obscure sensitive data to prevent breaches. Maintaining customer confidence and financial system security is crucial [7].

Reduced Risk of Data Breaches

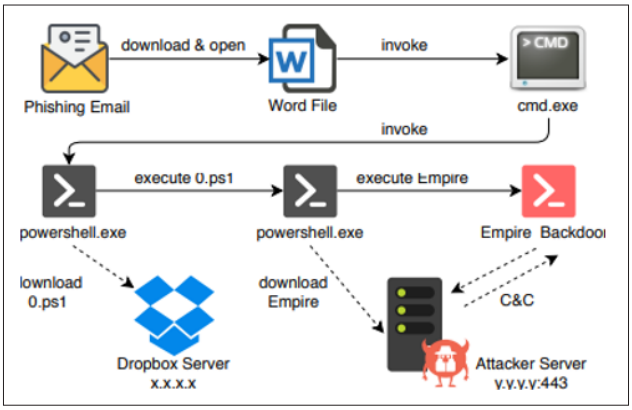

Figure 5: The Kill Chain of the Dde Script-Based Attack

Obfuscation of data helps to limit the number of data breaches by concealing sensitive information [5]. Even unauthorised access is pointless without a mechanism to interpret obscured data. Financial services, healthcare, and industrial control systems need real-time data creation and delivery. Qiu et al. [6] show advanced gradient obfuscation. We want processing efficiency and adversarial attack mitigation. These technologies make private data reading and reverse-engineering challenging. Thus, even if a bad person gets data, they cannot get valuable insights [7]. To maintain patient confidentiality and trust in telemedicine and health data analytics, healthcare data breaches must be reduced. Healthcare providers may obscure sensitive patient data during transmission and processing to protect privacy. Patient data security is necessary for telemedicine to be legal and trustworthy.

Regulatory Compliance

Obfuscating data ensures privacy compliance and regulatory compliance. Under rigorous restrictions, several industries must protect sensitive data. Obfuscation may help companies meet these criteria and avoid fines. Taint Stream aids HIPAA and GDPR compliance. In large data systems, Taint Stream masks data transit [2]. These rules necessitate robust data privacy and security measures for personal data. Legality is needed to protect patient privacy and restore healthcare credibility [7]. Healthcare providers may comply with HIPAA by obfuscating patient data during transmission and processing. Obfuscating patient data helps healthcare providers comply with laws and avoid fines [5]. Banks must follow standards to protect transaction data and confidence. Financial institutions must follow GDPR data protection requirements. Financial institutions may escape legal and financial penalties by obfuscating data.

Challenges in Implementing Data Obfuscation

Maintaining Data Utility

The preservation of the utility of data is made more difficult by data obfuscation. Data privacy and analytical and decision-making value must be balanced. Wang et al. [3] researched malware detection and underlined the need for real-time data processing-friendly obfuscation. While preserving data-driven judgements, advanced encryption should conceal sensitive data. This equilibrium is important in healthcare and finance, where data is crucial for decisions.

Computational Overhead

The process of obfuscation is both complicated and computationally costly. Computing may delay real-time data pipelines, which need quick processing. Qiu et al. [6] study advanced gradient-obscuring algorithms. Maintaining processing efficiency while controlling hostile attacks is the goal. Improving real-time obfuscation is emphasised. In real-time data-processing applications like industrial control systems and financial transactions, excessive computer overhead may slow performance.

Scalability

As the quantity of data continues to increase at a fast rate, scalability becomes troublesome. Obfuscation and speed are needed to handle huge datasets. Big data environments need scalable obfuscation methods to handle growing data quantities and dynamically execute access control rules, according to Anisetti et al. [9]. Scalability lets obfuscation strategies be applied equally to big datasets, assuring security without delaying data processing. Telecommunications and big data analytics need real-time database processing and protection.

Case Studies and Implementation

Healthcare

obscuring the Internet of Things in the medical field [5]. Obfuscation secures telemedicine and analytics health data, according to their research. Patient data is protected during real-time transmission and review using obfuscation. This strategy protects privacy while letting professionals improve therapy. Big data allows healthcare predictive maintenance [25]. Privacy-enhancing data obfuscation is explained here. These technologies improve healthcare dependability and efficiency using predictive maintenance analysis.

Financial Services

Real-time processing systems encrypt the data that pertains to financial transactions [7]. This encryption technology speeds up banking and safeguards data. Financial organisations may protect transaction data and systems via obfuscation and encryption. Striim's real-time ETL needs obfuscation [10]. Real-time data integration needs obfuscation for security and integrity. Financial institutions may safely utilise and evaluate sensitive data.

Industrial Control Systems

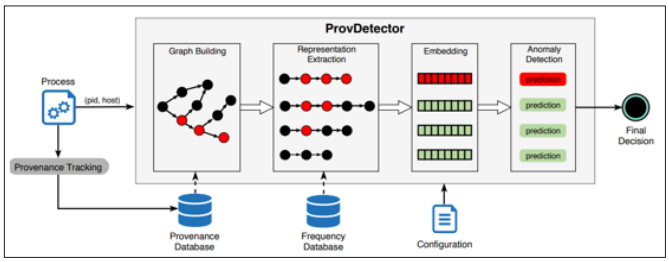

Figure 7: The Overview of PROVDETECTOR

This study analyses blockchain and obfuscation for water distribution system data security [20]. Obfuscation may improve data security in real-time monitoring and control systems, emphasising the necessity of data privacy in critical infrastructure. Water distribution systems may hide operating data to preserve vital data. This strategy ensures critical infrastructure stability and security since even a little interruption may be fatal. Electrosense monitors spectrum. It obfuscates spectrum data for quick processing and analysis, argue Rajendran et al. [21]. Obfuscation protects user privacy and data in real-time surveillance and industrial control systems.

Future Directions

Enhancing Robustness

Obfuscation tactics are being enhanced against increased threats. Obfuscation of the graphics processing unit (GPU) is used in order to protect against attacks [4]. This suggests a move towards hardware-based data security. Resilience requires obfuscation techniques that can survive complex assaults and preserve critical data. For robust security, obfuscation must develop with threats.

Integrating Artificial Intelligence

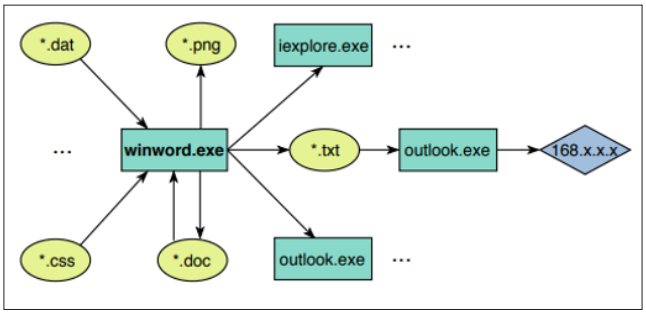

Figure 6: The provenance graph of a benign instance of MS Word

Using artificial intelligence and machine learning to obfuscate data is a successful method. These approaches adapt obfuscation to context and content. Artificial intelligence can discover data access and use trends and irregularities, updating encryption measures to increase data security. AI can make obfuscation smarter and more responsive, protecting sensitive data.

Standardization and Best Practices

Solving real-time data obfuscation requires academic-business collaboration. Optimising methods and standardising frameworks would help diverse fields use obfuscation. Standardisation may make obfuscation more popular and effective under security standards. Data security and leakage may improve with obfuscation and industry best practices. This cooperation will also spur innovation in obfuscation approaches to fit business demands.

Conclusion

Security and privacy need data obfuscation in real-time data pipelines. Multiple sectors employ this technology to safeguard privacy, eliminate security concerns, and comply with legislation. Scalability, processing overhead, and data usefulness must be handled to optimise its potential. Real-time data protection obfuscation technologies will advance via research and collaboration.

References

1. Shankar S, Parameswaran A (2021) Towards observability for production machine learning pipelines. ar Xiv preprint ar Xiv 13557.
2. Yang C, Li Y, Xu M, Chen Z, Liu Y, et al. (2021) TaintStream: fine-grained taint tracking for big data platforms through dynamic code translation. In Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering 806-817.
3. Wang Q, Hassan WU, Li D, Jee K, Yu X, et al. (2020) You Are What You Do: Hunting Stealthy Malware via Data

Provenance Analysis. In NDSS.
4. Chakraborty A, Xie Y, Srivastava A (2018) GPU obfuscation: attack and defense strategies. In Proceedings of the 55th Annual Design Automation Conference 1-6.
5. Kavitha D, Subramaniam C (2017) Security threat management by software obfuscation for privacy in internet of medical thing (IoMT) application. Journal of Computational and Theoretical Nanoscience 14: 3100-3114.
6. Qiu H, Zeng Y, Zheng Q, Zhang T, Qiu M, (2020) Mitigating advanced adversarial attacks with more advanced gradient obfuscation techniques. arXiv preprint arXiv:2005.13712.
7. Breuer PT, Bowen JP, Palomar E, Liu Z (2017) October. Encrypted computing: Speed, security and provable obfuscation against insiders. In 2017 International Carnahan Conference on Security Technology (ICCST) 1-6.
8. Chang SW, Onken CA, Wolf C, Luvaul L, Möller A (2021) SkyMapper optical follow-up of gravitational wave triggers: Alert science data pipeline and LIGO/Virgo O3 run. Publications of the Astronomical Society of Australia, 38: 024.
9. Anisetti M, Ardagna CA, Braghin C, Damiani E, Polimeno A, Balestrucci A, (2021) Dynamic and scalable enforcement of access control policies for big data. In Proceedings of the 13th International Conference on Management of Digital EcoSystems 71-78.
10. Pareek A, Khaladkar B, Sen R, Onat B, Nadimpalli V, et al. (2018) Real-time ETL in Striim. In Proceedings of the international workshop on real-time business intelligence and analytics 1-10.
11. Ardagna CA, Bellandi V, Ceravolo P, Damiani E, Bezzi M, (2017) A model-driven methodology for big data analytics-as-a-service. In 2017 IEEE international congress on big data (BigData Congress) IEEE 105-112.
12. Hajihassani O, Ardakanian O, Khazaei H (2021) Anonymizing sensor data on the edge: a representation learning and transformation approach. ACM Transactions on Internet of Things 3: 1-26.
13. Verma ME, Bridges RA, Sosnowski JJ, Hollifield SC, Iannacone MD, et al. (2021) CAN-D: A modular four-step pipeline for comprehensively decoding controller area network data. IEEE Transactions on Vehicular Technology 70: 9685-9700.
14. Ouaret S (2021) Architecture Design and implementation of the startup Farmy. ai's data pipeline (Doctoral dissertation, Université Mohamed El Bachir El Ibrahimi de Bordj Bou Arreridj).
15. Kersten M.L, Zhang Y, Nes N, Koutsourakis P (2021) Bridging the Chasm between Science and Reality. In CIDR.
16. Mc Cully C, Volgenau NH, Harbeck DR, Lister TA, Saunders ES, et al. (2018) Real-time processing of the imaging data from the network of Las Cumbres Observatory Telescopes using BANZAI. In Software and Cyberinfrastructure for Astronomy v 10707: 141-149.
17. Prasser F, Spengler H, Bild R, Eicher J, Kuhn KA, et al. (2019) Privacy-enhancing ETL-processes for biomedical data. International journal of medical informatics 126: 72-81.
18. Walth G, Wright SA, Weiss J, Larkin JE, Moore AM, et al. (2016) August. The Infrared Imaging Spectrograph (IRIS) for TMT: data reduction system. In Software and Cyberinfrastructure for Astronomy IV SPIE 9913: 1520-1532.
19. Masci FJ, Laher RR, Rusholme B, Shupe DL, Groom S, et al. (2018) The zwicky transient facility: Data processing, products, and archive. Publications of the Astronomical Society of the Pacific 131: 018003.
20. Mahmoud HHM, Wu W, Wang Y (2019) September. Secure data aggregation mechanism for water distribution system using blockchain. In 2019 25th international conference on automation and computing (ICAC) IEEE 1-6.
21. Rajendran S, Calvo-Palomino R, Fuchs M, Van den Bergh B, Cordobés H, et al. (2017) Electrosense: Open and big spectrum data. IEEE Communications Magazine 56: 210-217.
22. Ding Y, Li Z, Yastremsky D (2021) Real-time face mask detection in video data. arXiv preprint arXiv:2105.01816.
23. Edwards P, Cheng L, Kadam G (2019) Border gateway protocol anomaly detection using machine learning techniques. SMU Data Science Review 2: 5.
24. Awaysheh F, Cabaleiro JC, Pena TF, Alazab M (2019) Big data security frameworks meet the intelligent transportation systems trust challenges. In 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE) 807-813.
25. Çoban S, Gökalp MO, Gökalp E, Eren PE, Koçyiğit A (2018) November. [WiP] Predictive maintenance in healthcare services with big data technologies. In 2018 IEEE 11th Conference on Service-Oriented Computing and Applications (SOCA) 93-98.