

Threat Analysis of Insider Collusion Attacks On Kernel-Based Privacy Preserving Data-Mining

LAVANYA. B, SANDHIYA. M, NIVETHA. S, SUBASRI. S, SUBASHINI. S

¹Assistant Professor of Computer Science and Engineering

Bharathiyar Institute of Engineering for Women (Autonomous), Salem, Tamil Nadu, India

^{2,3,4,5}Student, Department of Computer Science and Engineering

Bharathiyar Institute of Engineering for Women (Autonomous), Salem, Tamil Nadu, India

ABSTRACT-Privacy Preserving Data Mining (PPDM) enables knowledge extraction from large datasets while protecting sensitive information. Organizations such as hospitals, banks, and businesses rely on data mining for decision-making, but insider threats can compromise data privacy. Insider collusion attacks occur when two or more authorized users cooperate to access confidential information illegally.

Kernel-based techniques, particularly Support Vector Machines (SVM), are commonly used for classification and pattern recognition but remain vulnerable if security measures are not applied.

This paper analyzes insider collusion threats in kernel-based PPDM systems. The proposed system incorporates modules for data collection, preprocessing, kernel-based mining, insider threat detection, and secure data output. User activities and system logs are continuously monitored to detect suspicious behavior. Experimental results demonstrate that the system improves data security and prevents insider data leakage.

Keywords: Privacy Preserving Data Mining, Insider Collusion Attack, Kernel Methods, Support Vector Machine, Data Security.

1. INTRODUCTION

Data mining extracts patterns from large datasets and is widely applied in healthcare, banking, and business analytics. Sensitive data creates privacy concerns. PPDM allows data analysis while protecting confidential information through anonymization, encryption, and secure computation. Kernel-based methods such as SVM handle complex datasets effectively, but insider threats remain a significant risk. **Insider collusion attacks** are difficult to detect because insiders already have system access, requiring robust monitoring and security mechanisms.

2. LITERATURE SURVEY

Key contributions in PPDM and insider threat detection include:

- **Agrawal & Srikant (2000):** Early privacy-preserving data mining techniques.

- **Clifton et al. (2002):** Privacy-preserving distributed data mining.
- **Fung et al. (2010):** Anonymization techniques (e.g., k-anonymity).
- **Cortes & Vapnik (1995):** Support Vector Machines for classification and pattern recognition.
- **Salem et al. (2008):** Insider attack detection through user activity monitoring.

While research has addressed privacy protection and insider threats, **insider collusion in kernel-based PPDM** remains underexplored. Enhanced monitoring and detection systems are necessary.



Fig 1.Literature Survey Workflow

3. METHODOLOGY

3.1 Data Collection

Data is acquired from multiple sources including databases, organizational records, and online systems, containing structured and unstructured sensitive information. Secure handling ensures reliability.



Fig 2.Data Collection Workflow

3.2 Data Preprocessing

Preprocessing includes cleaning, integration, transformation, and reduction to ensure high-quality datasets for kernel-based mining while maintaining privacy.

Steps include:

- **Data Cleaning:** Removes errors or incomplete entries.
- **Data Integration:** Combines multiple sources into a unified dataset.
- **Data Transformation:** Converts data into analysis-ready formats.
- **Data Reduction:** Reduces dataset size while retaining key information.



Fig 3.Data Preprocessing Pipeline

3.3 Kernel-Based Data Mining Using SVM

Data is transformed to higher-dimensional space for pattern detection. SVM performs classification without exposing sensitive information. Kernel functions enhance similarity measurement and accuracy.

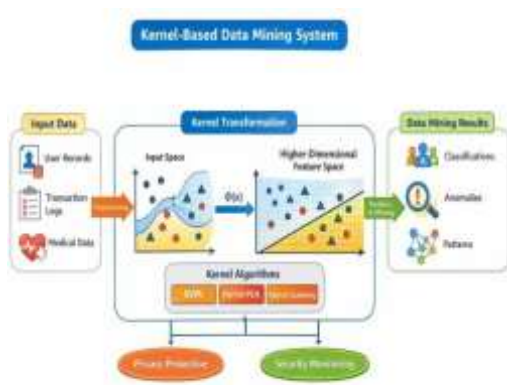


Fig 4.Kernel-Based Mining Workflow

3.4 Insider Collusion Attack Detection

The system monitors user activity, access logs, and collaboration patterns. Abnormal behavior triggers alerts, preventing unauthorized data disclosure.



Fig.5 .Insider Collusion Detection Workflow

3.5 Secure Data Processing & Output

Mining outputs are restricted to authorized information. Encryption and access control ensure confidentiality, and outputs are verified to prevent privacy leakage.

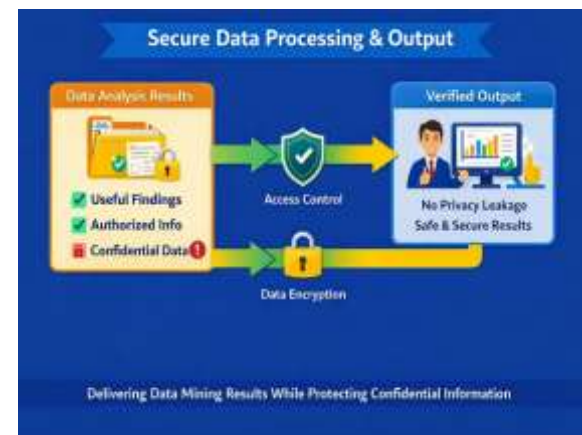


Fig 6. Secure Data Output

4. PROPOSED SYSTEM

The system integrates four modules for secure kernel-based PPDM:

1. **User Module:** Handles registration, authentication, and activity logging. Detects suspicious behavior and collusion attempts.
2. **Data Storage Module:** Stores data securely using privacy-preserving techniques, maintains integrity, and grants access based on permissions.
3. **Privacy Protection Module:** Ensures confidential data is protected during storage and analysis using kernel-based methods.
4. **Threat Detection Module:** Continuously monitors system logs and user access patterns to identify abnormal behavior or collusion, generating alerts.



Fig 7. System Architecture

4.1 User Module

The **User Module** manages registration, authentication, and activity logging to ensure only authorized users access sensitive data. All actions, including login, data upload, and access, are recorded in system logs. These logs help detect suspicious behavior and potential collusion between users. By continuously monitoring user activity, the module acts as the first line of defense against insider threats.



Fig 8. User Module Workflow

4.2 DATA STORAGE MODULE

The **Data Storage Module** securely stores all system data using privacy-preserving techniques. Only authenticated users with proper permissions can access the data. This module ensures data integrity, prevents unauthorized access, and supports privacy-preserving mining operations. Encryption and access controls are applied to maintain confidentiality during storage and retrieval.



Fig 9. Data Storage Module Workflow

4.3 PRIVACY PROTECTION MODULE

The **Privacy Protection Module** ensures confidential data remains secure during storage and analysis. Kernel-based methods allow the system to perform accurate mining without exposing sensitive information. The module restricts unauthorized access and monitors data usage, reducing the risk of insider collusion and safeguarding sensitive records during processing.



Fig 10. Privacy Protection Module Workflow

4.4 THREAT DETECTION MODULE

The **Threat Detection Module** continuously monitors system logs, user activity, and access patterns to identify suspicious behavior or insider collusion attempts. When abnormal activities are detected, the system triggers alerts to administrators for immediate review. This module is critical for early detection of threats and prevention of data leakage. This module plays an important role in preventing data leakage and improving system security.



Fig 11. Threat Detection Module Workflow

5. RESULTS AND DISCUSSION

The proposed system was evaluated for detecting insider collusion attacks in kernel-based privacy-preserving data mining. Key findings include:

- **User Authentication:** Only authorized users could access the system, effectively preventing unauthorized access.

- **Secure Data Storage:** Sensitive information was stored using privacy-preserving techniques, ensuring confidentiality and data integrity.
- **Threat Detection:** User activities and system logs were continuously monitored. Suspicious behaviors, such as repeated access to sensitive data or unusual collaboration, triggered alerts for administrators.

Discussion:

Integrating privacy protection with continuous monitoring significantly reduces the risk of insider collusion attacks. Early detection of abnormal patterns allows organizations to act before data leakage occurs. Kernel-based mining ensures accurate analysis without exposing sensitive information. Overall, the system improves the reliability, security, and privacy of data mining operations.



Figure12: Results & Discussion

6. CONCLUSION

This study addressed insider collusion risks in kernel-based privacy-preserving data mining. By combining user authentication, secure data storage, privacy protection, and threat detection, the system effectively identifies suspicious activities and prevents unauthorized access.

Key Outcomes:

- Improved detection of insider collusion attacks.
- Protection of sensitive information during data mining.
- Enhanced system reliability and security.

Future Work:

- Integration of AI/ML for real-time threat detection.
- Exploration of blockchain for secure and tamper-proof distributed data storage.

The proposed framework provides a secure, privacy-aware environment for analyzing large datasets while minimizing insider threats.

REFERENCES

1. R. Agrawal, R. Srikant, "Privacy-Preserving Data Mining," ACM SIGMOD, 2000.
2. C. Clifton, M. Kantarcioglu, Vaidya, X. Lin, M. Zhu, "Tools for Privacy Preserving Distributed Data Mining," ACM SIGKDD Explorations, 2002.
3. B. C. M. Fung, K. Wang, R. Chen, P. S. Yu, "Privacy-Preserving Data Publishing: A Survey," ACM Computing Surveys, 2010.
4. C. Cortes, V. Vapnik, "Support-Vector Networks," Machine Learning, 1995.
5. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
6. M. Salem, S. Hershkop, S. Stolfo, "Survey of Insider Attack Detection," Insider Attack and Cyber Security Conference, 2008.
7. J. Vaidya, C. Clifton, Y. Zhu, *Privacy Preserving Data Mining*, Springer, 2005.
8. M. Bishop, *Computer Security: Art and Science*, Addison-Wesley, 2003.
9. M. Bishop, S. Engle, "The Insider Threat," IEEE Workshop on Security and Privacy, 2005.
10. A. Shamir, "How to Share a Secret," Communications of the ACM, 1979.
11. R. Chen, B. C. M. Fung, J. Yu, P. S. Yu, "Correlation-Privacy in Data Publishing," IEEE Transactions on Knowledge and Data Engineering, 2012.
12. S. Jha, M. G. Kang, R. Sekar, "Detecting Insider Attacks in Enterprise Systems Using Audit Logs," ACM Transactions on Information and System Security, 2007.
13. L. V. Kale, M. Ripeanu, "Privacy-Preserving Distributed Machine Learning: A Survey," Journal of Big Data, 2018.
14. V. Shmatikov, "Privacy-Preserving Machine Learning: Threats and Solutions," IEEE Security & Privacy, 2015.
15. N. Li, T. Li, S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," IEEE 23rd International Conference on Data Engineering (ICDE), 2007.