Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hybrid Insider Threat Detection Using Adversary Emulation, Endpoint Forensics, and Unsupervised Machine Learning

[1]Manhal Mohamad Basher, [2] Yaseen Hikmat Ismael
1Computer Science / Northern Technical University, Mosul, Iraq 2Computer Science / Mosul, University, Mosul, Iraq
Email: 1 manhalbasher@ntu.edu.iq, 2 hikmat@uomosul.edu.iq

ARTICLE INFO

Received: 02 March 2026

Revised : 08 April 2026

Accepted: 20 April 2026

ABSTRACT

Internal threats are a big security problem. When internal users perform malicious activities, they use their legitimate privileges to perform legitimate-looking activities. There is not much classified data to analyze. Most detection methods use synthetic data or supervised learning, which makes it difficult to detect the actual behavior of the attacker.

This study proposes a hybrid framework for detecting insider threats that combines attack simulation, endpoint forensic analysis, and unsupervised machine learning. Attacks are simulated using the MITRE CALDERA platform, forensic data is collected via Velociraptor, and user behavior is analyzed using Isolation Forest and AutoEncoder models to detect abnormal activity without labeled data.

The analysis results are integrated into a unified risk index ranging from 0 to 100 and classified according to CERT criteria to facilitate interpretation of the results. Experiments demonstrate that the proposed framework can detect many types of internal attacks, including abuse of authentication, depletion of system resources, covert execution, and log tampering, making it a practical and applicable solution in the environments of real- world security.

Keywords: Insider Threats, Caldera Platform, Velociraptor Tools.

I. INTRODUCTION

I.1 Background

Insider threats are considered a critical security issue that faces modern organizations, as malicious users or attackers have legitimate access to systems and resources. Unlike external attackers, internal actors operate within a trusted environment, which reduces the effectiveness of traditional protection mechanisms that rely on securing network boundaries. Specialized literature and reports indicate that incidents resulting from internal threats have serious financial and operational consequences, as well as the challenge of early-stage detection.

I.2Current Business Constraints

Current approaches to detecting internal threats suffer from several methodological and technical limitations that undermine their practical effectiveness. One major drawback is the fact that it is highly reliant on classified data sets, which are hard to come by and often generated artificially. Most methods focus mainly on examining system logs while ignoring the vast digital evidence available at the endpoint. This makes the scope and accuracy of the examination quite limited. It is also difficult to detect slow- running behavioral patterns within the system while also dealing with the complexity of interpreting the results.

Despite the prevalence of the use of unsupervised methods in the detection of insider threats, they face fundamental issues related to the data we possess.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Yuan Wu (2021) was able to show how conventional machine learning models may not be able to distinguish between malicious insider behavior and normal behavior, overwhelmed by the lack and scarcity of the data we possess and the limited number of instances we can label [1]. Similarly, recent works in the realm of supervised learning in the detection of insider threats show how the use of predefined labels may limit the ability of these models to generalize and respond to unknown or unfamiliar situations.

I.3Motivation

To avoid the shortcomings associated with conventional approaches to insider threat detection, we must be open to broad and realistic approaches. This means that the new approaches would be able to monitor insider behavior in a realistic way, utilize all forms of digital evidence available on the endpoint devices, operate even in the absence of classified information, and offer understandable risk assessments.

I.4 Contributions

This study offers several concrete academic and practical contributions, which can be summarized as follows:

Designing a hybrid framework for detecting insider threats that combines adversary behavior simulation, digital forensic evidence extraction from endpoints, and the application of unsupervised machine learning techniques.

Employing the MITRE CALDERA platform to generate realistic digital insider attack scenarios.

Extracting session-based digital behavioral forensic features using the Velociraptor tool to enhance the accuracy of behavioral analysis.

Integrate Isolation Forest and Autoencoder models to detect anomalous patterns within behavioral data.

Develop a unified model for integrating risk indicators that produces classifications that are interpretable and actionable by security analysts, based on CERT-approved risk classification criteria[17].

II. RELATED WORK

II.1Anomaly-based detection

This is based on the fact that it learns the patterns of normal behavior for users, then identifies behavior that is abnormal. It is considered suitable for deployment in the real world due to a number of reasons; attacks are considered to be rare compared to normal behavior, it is difficult to get labeled data on internal threats, and the behavior associated with attacks is complex.

Anomaly-based detection methods are broadly classified under three categories: Single-Class Classification, Behavior Prediction, and Autoencoder Reconstruction. Under the Single-Class Classification category, the One-Class Conditional Random Field (OCCRF) model is used to learn the normative patterns of user behavior, with the Joint Loss function used to enhance the discrimination between the normal and anomaly behaviors. Rashid et al. (2016) used a Hidden Markov Model (HMM) to model the sequence of normal behavior features [4], while Lin et al. (2017) used deep belief network feature representations to detect internal threats. Other studies have used one-class support vector machines (OCSVM) to identify anomalous behavior [5].

With significant advances in deep learning techniques, researchers have begun employing deep learning models to develop more effective anomaly detection schemes. This involves modeling the prediction of next user behavior and comparing it to actual behavior, with any noticeable deviation from the predicted results being considered an indicator of anomalous behavior. For example, Villarreal- Vasquez used an LSTM network to model system event sequences and predict the probability of the next event, considering events with low probability to be anomalous [6]. Yuan Wu adopted a similar approach. Deep Autoencoders were also used to reconstruct behavioral data, with behavior yielding high reconstruction error considered anomalous [7].

II.2 Abuse-based detection and classification- based detection

Abuse-based detection relies solely on modeling threat behaviors and aims to identify malicious activities by comparing test data with threat- associated behavior patterns. According to the current literature, applications of this approach in the field of insider threats are limited to a small number of studies, mainly due to the scarcity of

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

labeled threat- behavior data, as well as the high diversity of insider attack patterns, which limit the generalizability of models.

In contrast, many detection techniques resort to classification methods, which are based on modeling both normal and malicious behaviors. These methods provide the ability to incorporate precise information about threat behavior, enabling a direct relationship between the model and insider threat activities, thereby reducing false alarm rates compared to anomaly-based detection methods. However, these methods face challenges related to training data imbalance, given the scarcity of labeled threat cases. For example, Chattopadhyay et al. (2018) used random forest (RF) and multilayer perceptron (MLP) algorithms, applying random undersampling to address the data imbalance problem [8]. Azaria et al. (2014) developed seven hybrid algorithms that combine support vector machines (SVMs), multinomial Naive Bayes (MultinomialNB), and semi-supervised learning techniques to address class imbalance and under-classification [9].

Furthermore, Lee et al. (2021) used various semi- supervised learning methods like label propagation, label spreading, and self-training to identify insider threats and handle data imbalance[10]. In addition, Alshahari and Suweil (2021) used seven prominent machine learning algorithms to detect data leakage. SMOTE was used to handle class imbalance problems [11]. Wu and Li (2021) used neural networks and random forests with feature selection to improve the detection of threatening activities [12].

Although carefully designed classification methods can perform well when applied to labeled test sets, their effectiveness remains limited when faced with incomplete data or unknown threat behaviors that did not appear in the training data.

II.3 Hybrid Detection

Hybrid detection involves the use of a combination of anomaly monitoring and abuse detection. This is a mix of the two detection methods with the aim of improving the detection of insider threats while reducing false alarms. Though this approach is good in theory, it is yet to be fully implemented in practice with regards to insider threats. This problem has been addressed through a new approach by Mohammed Nasser Almuhaigani et al., which is a hybrid model that uses machine learning to detect insider threats by examining normal and threat behaviors [13].

II.4 Unsupervised Detection

The unsupervised methods do not require any kind of labeling and use methods such as Isolation Forest and clustering. Gavai et al. (2015) employed Isolation Forest in detecting changes in a user's behavior compared to their past behavior and other users [14]. Soh et al. (2019) employed this method in detecting possible threats by considering emotions [15]. The clustering methods have been employed in detecting bad nodes in graphs by considering vectors of node embedding (Liu et al., 2019) [16]. It is noteworthy that the effectiveness of unsupervised methods depends on proper feature extraction, as better features help in differentiating normal and threatening behavior.

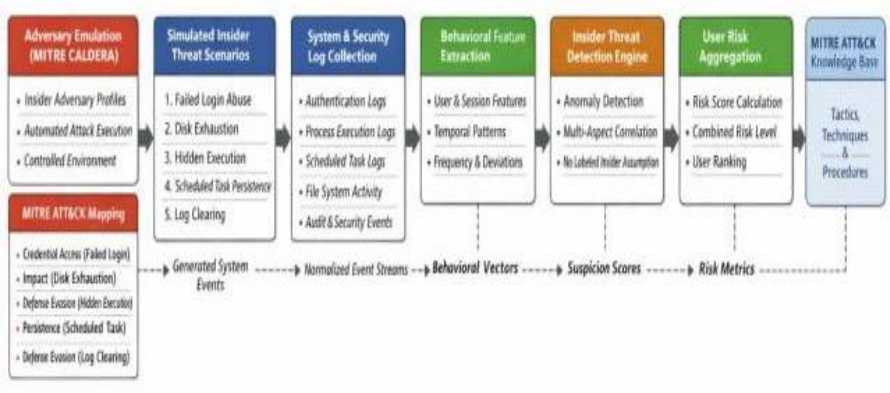
III. SYSTEM ARCHITECTURE

III.1Adversary Emulation (CALDERA) Insider TTPs mapped to MITRE ATT&CK Simulated attacks:

- Failed login abuse
- Disk exhaustion
- Hidden execution
- Scheduled task persistence
- Log clearing

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



4.2 Endpoint Forensic Collection (Velociraptor)

Windows endpoints Event logs

Process and task artifacts Disk utilization evidence

Table 1

Insider Forensic Attack Types

Attack Type	Forensic Evidence	Example Artifacts
Failed Login	Auth logs	Event ID 4625
Disk Exhaustion	Disk metrics	Event ID 4104
Hidden Execution	Process flags	Hidden tasks
Scheduled Task	Persistence	schtasks
Log Clearing	Anti-forensics	Event ID 1102

III.3 Feature Engineering

Aggregation by:

User FQDN

Session

Features:

Event counts

Attack-stage frequencies Ratios (failed/total, hidden/visible)

Session duration

Normalization:

RobustScaler

IV. DETECTION MODELS

IV.1 Isolation Forest

- Detects global behavioral anomalies

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- Captures rare forensic patterns
- Output: anomaly score

IV.2 Autoencoder

- Learns normal user behavior
- Flags deviations via reconstruction error
- Effective for subtle insider actions

Table 2: Detection Model Roles

Model	Detection Goal	Output
IF	Global anomaly	IF score
AE	Behavioral deviation	Reconstruction error

V. EXPERIMENTAL SETUP

V.1 Experimental Environment

The experimental environment includes several Windows systems that represent the normal workstation environment in a company. Each workstation is associated with a specific user, which means we can observe both normal activities and internal activities to simulate the actual environment in the company. Normal activities include activities like login, accessing files, etc. Internal activities include activities like failed login attempts, running out of disk space, stealth execution, keeping tasks after restarting the computer, changing the registry settings, etc.

This heterogeneous environment enables evaluation of detection system performance across diverse users and multiple attack types, while maintaining realism and diversity in user behavior.

V.2 Attack Scenarios

The research makes use of the MITRE CALDERA platform to execute the attacks. This is to simulate realistic insider threat behavior in the experiment. Each scenario is a different pattern of insider attacks. It can be a single-stage attack, a multi-stage attack, or occur during existing user sessions. Single-stage attacks include isolated malicious actions such as wiping logs or executing unusual commands. On the other hand, multi-stage attacks are designed to mimic complex insider actions. This involves various stages like abusing authentication and then employing persistence tactics. The persistent sessions are helpful in mimicking insider actions over time. This way, the testing of the framework takes place under realistic conditions where insider threats are not isolated incidents but occur over time with malicious and benign actions being intermixed.

V.3 Evaluation Strategy

The evaluation plan for the proposed framework does not use any classified data regarding internal incidents, as in the real world it is hard to obtain confirmed data. It makes use of both visual and statistical checks, including risk score distributions, user classifications, risk category distributions in accordance with CERT standards, etc. These enable us to understand the results in both a quantitative and qualitative manner, with clear ideas about how to spot suspicious behavior. The evaluation process also attempts to assist the analysts in decision-making by showing them how the system can locate critical activities. The plan is in accordance with the usual methods of detecting unsupervised insider threats while ensuring the viability of the method without the use of documented reference data.

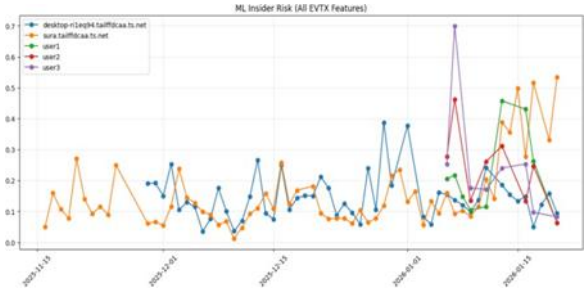
Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

VI. RESULTS

VI.1 User-Level Risk Analysis using Machine learning

User Risk Comparison



✅ Daily forensic report saved: E:\Forensic_AttackType_DAILY_JAN2026.csv

=== DAILY ATTACK SUMMARY (JAN 2026) ===

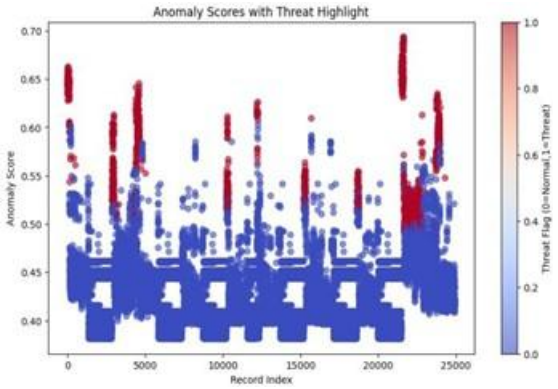
Day	Actor	AttackType	Count
2026-01-01	sura.tailffdcaa.ts.net	Scheduled Task Persistence	1
2026-01-03	desktop-rileq94.tailffdcaa.ts.net	Failed Login	2
2026-01-03	desktop-rileq94.tailffdcaa.ts.net	Hidden Execution	4
2026-01-03	sura.tailffdcaa.ts.net	Hidden Execution	40
2026-01-03	sura.tailffdcaa.ts.net	Log Tampering (Clear Event Log)	9
2026-01-06	sura.tailffdcaa.ts.net	Failed Login	2
2026-01-06	user3	Failed Login	2
2026-01-07	user2	Failed Login	1
2026-01-07	user2	Scheduled Task Persistence	3
2026-01-07	user3	Scheduled Task Persistence	3
2026-01-08	desktop-rileq94.tailffdcaa.ts.net	Failed Login	1
2026-01-09	user2	Failed Login	2
2026-01-09	user3	Failed Login	1
2026-01-11	desktop-rileq94.tailffdcaa.ts.net	Failed Login	1
2026-01-11	user1	Failed Login	1
2026-01-11	user2	Failed Login	3
2026-01-11	user3	Failed Login	9
2026-01-13	sura.tailffdcaa.ts.net	Failed Login	5
2026-01-13	user1	Failed Login	7
2026-01-13	user2	Failed Login	6
2026-01-13	user3	Failed Login	7
2026-01-14	desktop-rileq94.tailffdcaa.ts.net	Failed Login	1
2026-01-14	sura.tailffdcaa.ts.net	Failed Login	4
2026-01-15	sura.tailffdcaa.ts.net	Failed Login	2
2026-01-16	sura.tailffdcaa.ts.net	Failed Login	4
2026-01-16	sura.tailffdcaa.ts.net	Log Tampering (Clear Event Log)	9
2026-01-16	user1	Failed Login	1
2026-01-16	user1	Scheduled Task Persistence	3
2026-01-16	user2	Failed Login	5
2026-01-16	user3	Failed Login	10
2026-01-17	sura.tailffdcaa.ts.net	Failed Login	11

VI.2 Anomaly Score Result

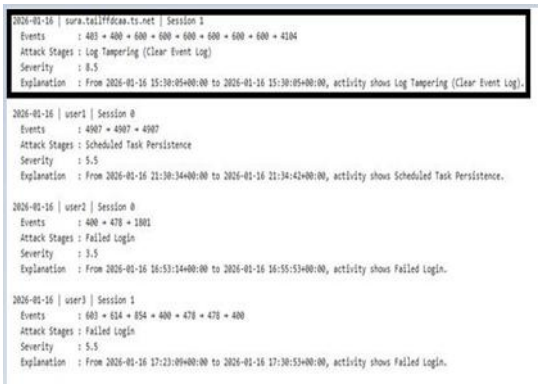
User / FQDN	Isolation Forest	Autoencoder
Sura.tailffdcaa.ts.net	36.57	19.52
Desktop-rileq94	9.11	6.41
User3	8.45	3.94
User2	7.52	6.02
User1	7.2	19.5

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



VI.3 Forensic Chain Attack Detection



VII. DISCUSSION

The results show that the output produced by the machine learning and deep learning models corresponds to the internal attacks identified in the study of the Velociraptor attack chain. All the internal attacks such as disk space exhaustion, failed login abuse, hidden execution, ongoing scheduled tasks, and log wiping were identified based on the way they were executed and the session. However, it is worth noting that there were no false positives or false negatives. This indicates the effectiveness of the process of checking all the breaches and comparing them with the attack chains. It also indicates that all the activities occurred in a specific session and at a specific time. The framework can detect both isolated and step-by-step malicious activities. This increases the extent to which the results of the model match the actual attack patterns, as well as confirming the effectiveness of the system in the detection of insider threats.

VIII. Limitations and Future Work

Despite the good results, some limitations have to be taken into account. First of all, the tests are conducted using fake and simulated attacks, which do not necessarily replicate the behaviors of attackers inside organizations. Another limitation is that the framework is currently designed to be used in Windows machines, but it would be necessary to test it in other systems to make it more usable. Additionally, the current approach of using weights for the data, together with the risk classification based on the CERT criteria, is static; however, using a dynamic approach could make this tool more effective for detecting problems. Finally, no false positives or false negatives were encountered during the experiments, but using this framework in a variety of different scenarios could reveal some problems not accounted for during the testing.

Future work may address these issues with the help of real-world data, extending the framework to support it across multiple platforms, and investigating how the risk might be adjusted with context or over time. The addition of online learning and the ability to interpret the anomalies may make the system more comprehensible, which can help the security analyst better understand how the internal employee behavior is changing. All these changes can make the framework more applicable in the real world, while maintaining the design to help the analyst and not use labeled data.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CONCLUSION

This paper presents a hybrid approach for detecting insider threats. It uses machine learning, deep learning, and adversary simulation techniques for detecting multi-stage and ongoing insider threats. It uses a weighted risk integration model and combines the results of various detection models with a standard approach. The framework maps risk scores to a level of severity using CERT criteria. It provides understandable risk scores without involving classified information.

The pilot test was performed using Windows operating systems, several users, and real attacks simulated by CALDERA. The pilot test verified the effectiveness of the system in detecting attacks such as running out of disk space, failed login attempts, stealthy runs, scheduled tasks, and registry wiping. No false positives or false negatives were reported. The results demonstrate the effectiveness of the framework in mapping the model results to the real attack steps and providing good guidance for the security analyst. Future plans include the addition of online learning to adjust the risk weights and the use of real-world tests to make the system more practical.

REFERENCE

[1] **Yuan & Wu (2021)** – *Deep learning for insider threat detection: Review, challenges and opportunities,* 2021 published by Elsevier. This manuscript is made available under the Elsevier user license https://www.elsevier.com/open- access/userlicense/1.0/.

[2] **He, J. (2025)** – *Machine Learning Applications in Cybersecurity: Survey, Challenges, and Future Trends*, 2025, published in MDPI Electronics. This manuscript is made available under the MDPI Open Access License https://www.mdpi.com/2079-9292/14/23/4563

[3] **Al-Mhiqani, M. N., Ahmad, R., Zainal Abidin, Z., Yassin, W., Hassan, A., Abdulkareem, K. H., Ali, N. S. & Yunos, Z. (2020)** – *A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations*, 2020 published in *Applied Sciences*, MDPI. This work is open access under the MDPI license https://doi.org/10.3390/app10155208

[4] **Rashid, T., Agrafiotis, I., & Nurse, J. R. C. (2016)** – *Detecting Insider Threats via Modeling Normal User Behaviour with Hidden Markov Models*, 2016, published in the *Proceedings of the 2016 Workshop on Managing Insider Security Threats (MIST'16)*, Association for Computing Machinery (ACM).

[5] **Lin, L., Zhong, S., Jia, C., & Chen, K. (2017)** – *Insider Threat Detection Based on Deep Belief Network Feature Representation*, 2017, published in the **Proceedings of the 2017 International Conference on Green Informatics (ICGI)**, Fuzhou, China.

[6] **Villarreal-Vasquez, M., Modelo-Howard, G., Dube, S., & Bhargava, B. K. (2023)** – *Hunting for Insider Threats Using LSTM-Based Anomaly Detection*, 2023, published in **IEEE Transactions on Dependable and Secure Computing**, Vol. 20, No. 1, pp. 451–462.

[7] **Yuan, S., Zheng, P., Wu, X., & Li, Q. (2019)** – *Insider Threat Detection via Hierarchical Neural Temporal Point Processes*, 2019 published on **arXiv**, https://arxiv.org/abs/1910.03171.

[8] **Chattopadhyay, P., Wang, L., & Tan, Y.-P. (2018)** , *Classification with Random Forest and Multilayer Perceptron using Random Under-Sampling*, **IEEE Transactions on Computational Social Systems**, 2018, (DOI: 10.1109/TCSS.2018.2857473).

[9] **Azaria, A., Richardson, A., Kraus, S., & Subrahmanian,V. S. (2014)** – *Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data*, 2014 published in **IEEE Transactions on Computational Social Systems**, Vol. 1, No. 2, pp. 135–155. **DOI: 10.1109/TCSS.2014.2377811**.

[10] **Le, D. C., Zincir-Heywood, A. N., & Heywood, M. I. (2021)**– *Training regime influences to semi-supervised learning for insider threat detection*, 2021.

[11] **Al-Shehari, T. & Alsowail, R. A. (2021)** – *An Insider Data Leakage Detection Using One-Hot*

Encoding, Synthetic Minority Oversampling and Machine Learning Techniques, 2021, published in **Entropy**, MDPI AG. This manuscript is made available under the MDPI open access license https://doi.org/10.3390/e23101258

[12] **Wu, C., & Li, X. (2021)** – *Fast Correlation-Based Feature Selection and Ensemble Learning for Threat Activity Detection*, 2021, published in the *Proceedings of the 2021 IEEE Security & Privacy Workshops (SPW)*, IEEE.

[13] **Al-Mhiqani, M. N., Ahmad, R., Z. Z. Abidin, K. H. Abdulkareem, M. A. Mohammed, D. Gupta & K. Shankar (2022)** – *A new intelligent multilayer framework for insider threat detection*, 2022, Vol. 97, Article 107597. https://doi.org/10.1016/j.compeleceng.2021.107597.

[14] **Gavai, G., Sricharan, K., Gunning, D., Rolleston, R., Singhal, M., & Hanley, J. (2015)** – *Supervised and Unsupervised Methods to Detect Insider Threat from Enterprise Social and Online Activity Data*, December 2015, **Vol. 6, No. 4, pp. 47–63**. DOI: 10.22667/JOWUA.2015.12.31.047

[15] **Soh, C., Yu, S., Narayanan, A., Duraisamy, S. & Chen, L. (2019)** – *Employee profiling via aspect-based sentiment and network for insider threats detection*, 2019, published in **Expert Systems with Applications, Elsevier**, Vol. 135, pp. 351–361. https://doi.org/10.1016/j.eswa.2019.05.043.

[16] **Liu, F., Jiang, X., Wen, Y., Xing, X., Zhang, D. & Meng, D. (2019)** – *Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise*, 2019, (DOI: 10.1145/3319535.3363224).

[17] Dhuha B. Abdulla, Basma B. Alzobeer, Analysis of Private Cloud Construction using Micrsoft Cloud Solution International Journal of Computer Applications,2014.