

Article

Not peer-reviewed version

Structure-Aware Unified Modeling for Root Cause Localization in Microservice Systems Using Multi-Source Observability Data

[Zixiao Huang](#), [Sijia Li](#), [Chengda Xu](#), [Bolin Chen](#), [Yihan Xue](#), Jixiao Yang *

Posted Date: 2 March 2026

doi: 10.20944/preprints202603.0081.v1

Keywords: microservice systems; multi-source observation fusion; dependency graph modeling; root cause localization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Structure-Aware Unified Modeling for Root Cause Localization in Microservice Systems Using Multi-Source Observability Data

Zixiao Huang ¹, Sijia Li ², Chengda Xu ¹, Bolin Chen ³, Yihan Xue ⁴ and Jixiao Yang ^{5,*}

¹ University of Washington, Seattle, USA

² University of Michigan, Ann Arbor, USA

³ Duke University, Durham, USA

⁴ University of Southern California, Los Angeles, USA

⁵ Westcliff University, Irvine, USA

* Correspondence: kin19960523@gmail.com

Abstract

By decoupling services and enabling elastic deployment, microservice architecture improves system scalability and evolutionary capability. At the same time, it substantially increases operational complexity. Failures often exhibit cross service propagation and a mismatch between observed symptoms and underlying root causes. To address the heterogeneity and fragmentation of multi source observability data such as logs, metrics, and distributed traces, this study proposes a unified modeling and intelligent root cause localization method for microservice systems. The approach treats each service as a basic modeling unit and maps heterogeneous observations into a shared representation space. Service dependency structure is explicitly incorporated to characterize system state at a global level. Through structure aware modeling on the dependency graph, anomaly information is propagated and constrained along real invocation relations. This design enables more accurate separation of local disturbances from structural anomalies. In addition, a consistency based measure derived from state deviation is constructed to score service anomalies. Dependency relations are then used for attribution and ranking, which unifies root cause localization and impact analysis within a single framework. Comparative results show that the proposed method achieves more stable and consistent advantages across multiple evaluation metrics. It captures anomaly propagation patterns in microservice systems more effectively and provides a unified and structure aware solution for intelligent diagnosis of complex distributed systems.

CCS CONCEPTS: Computing methodologies~Machine learning~Machine learning approaches

Keywords: microservice systems; multi-source observation fusion; dependency graph modeling; root cause localization

I. Introduction

With the maturation of cloud computing and container technologies, microservice architecture has become the dominant paradigm for large scale Internet systems and enterprise applications. By decomposing complex systems into loosely coupled service units that can be deployed and evolved independently, microservices provide clear advantages in flexibility, scalability, and continuous delivery[1]. At the same time, this high degree of decoupling and dynamic orchestration significantly increases operational complexity. Service instances scale with workload. Invocation chains span multiple components. Failures often appear as cascading effects, propagation, or hidden performance degradation. Under these conditions, operation and diagnosis methods designed for monolithic or

static architectures are no longer sufficient. This gap poses new challenges to system stability and business continuity[2].

To enhance observability, modern microservice platforms commonly adopt multiple monitoring sources, including logs, metrics, and distributed tracing. Each source captures system behavior from a different perspective. Logs record discrete events and abnormal semantics. Metrics describe continuous changes in resource usage and performance. Traces represent the temporal paths and dependency relations of cross service requests. These data sources are highly complementary in theory. However, they differ greatly in data type, time scale, structural form, and noise characteristics. In practice, they are often collected, stored, and analyzed in isolation. A unified modeling perspective is missing. As a result, abnormal signals are difficult to correlate across sources. Operators must switch between tools and viewpoints. Diagnosis efficiency and accuracy are therefore limited[3].

More importantly, failures in microservice systems are rarely isolated events. They are often systemic problems that propagate along service dependency paths[4]. Performance degradation in a low level component can be amplified through invocation chains. It may finally appear in upstream services as increased latency, higher error rates, or abnormal resource consumption. When only local observations or a single data modality are used, diagnosis usually captures outcome level anomalies rather than the true root causes. This mismatch between observed symptoms and actual causes increases troubleshooting time[5]. It can also lead to inappropriate remediation decisions and secondary risks. Therefore, explicitly modeling service dependencies at the system level and mapping multi source observations into a unified representation space is a key requirement for accurate root cause localization.

In real world deployments, the operational state of microservice systems is also highly dynamic and non stationary. Service versions evolve continuously. Deployment topologies change frequently. Business workloads exhibit strong temporal variations. As a result, the notion of normal behavior is inherently time dependent. Diagnostic approaches based on manual rules or static thresholds struggle to remain effective over time. They often produce frequent false alarms or severe missed detections. A unified modeling approach that captures system state evolution and distinguishes transient fluctuations from structural anomalies is essential. Such capability is critical for improving adaptive diagnostic performance. It also supports higher levels of automation and intelligent operations.

Against this background, research on unified modeling of multi source observability data and intelligent root cause localization for microservice architectures is of significant theoretical and practical value. From a theoretical perspective, this direction advances the systematic understanding of complex software system behavior. It supports the study of collaborative modeling mechanisms for heterogeneous data in a shared representation space. It also sheds light on the relationship between structural dependencies and state evolution. From an application perspective, these methods can provide more reliable and efficient fault localization for large scale distributed systems. They can reduce service downtime, lower manual intervention costs, and improve system stability and maintainability. As microservices are increasingly deployed in critical industries and core business systems, research on multi source observability fusion and root cause localization will play an increasingly important role in safeguarding digital infrastructure.

II. Methodological Foundations

A unified diagnosis framework for complex service-oriented systems benefits from representing heterogeneous observability streams as comparable service-level states, while enforcing system-level constraints through explicit dependency structure. Graph neural modeling with temporal dynamics provides a direct methodological basis for integrating node-wise state representations into a topology-consistent global embedding, enabling propagation of abnormal evidence along directed relations rather than relying on purely local detectors [6]. Structure-aware and semantically enhanced graph construction further motivates treating the dependency graph as an active inductive bias that regulates message passing, suppresses spurious correlations, and improves separability between

localized disturbances and structurally induced deviations [7]. To robustly quantify state deviation under evolving baselines, transformer-based change-point modeling illustrates how long-range temporal context can be leveraged to distinguish regime shifts from short-lived fluctuations, supporting deviation-driven scoring that remains stable under non-stationarity [8]. Hierarchical attention mechanisms complement this by emphasizing salient segments and multi-scale temporal dependencies, which aligns with fusing multi-source observations into a shared representation space through adaptive weighting rather than fixed feature aggregation [9]. When labels are scarce and anomaly distributions are highly imbalanced, self-supervised learning provides a principled route to learn invariant representations and consistency constraints from unlabeled streams, strengthening the reliability of unified embeddings across heterogeneous modalities [10]. Residual-regulated forecasting with differencing then offers a practical regularization perspective: by suppressing trend and drift components while preserving diagnostically meaningful residuals, the learned state deviation becomes less sensitive to workload-driven baseline changes [11]. Multi-head self-attention for dynamic anomaly identification further supports constructing deviation measures that adaptively focus on the most informative dimensions, improving the discriminative power of anomaly scores derived from latent state perturbations [12].

Operational deployment additionally demands continual learning under distributed data ownership and streaming updates. Asynchronous real-time federated learning establishes a training organization pattern that tolerates stragglers and heterogeneity while enabling near-real-time model updates, which is aligned with continuously refreshing service-state encoders without centralizing all raw observability signals [13]. Privacy-preserving and communication-efficient federated learning complements this by introducing mechanisms to reduce communication overhead and protect sensitive information, making unified multi-source modeling feasible at cloud scale under realistic governance and bandwidth constraints [14]. For dependency-aware attribution, transformer-based modeling with graph integration demonstrates how temporal representations can be coupled with relational structure for joint reasoning, reinforcing the design choice of constraining anomaly propagation and attribution through explicit invocation relations rather than post-hoc correlation [15]. Multi-hop relational modeling provides an additional foundation for controlled diffusion across multiple dependency steps, enabling the method to capture delayed and indirect propagation while retaining path-sensitive attribution capability [16]. Dynamic spatiotemporal causal graph neural networks further motivate embedding temporal evolution and causal structure into graph state transitions, supporting a principled separation between upstream driving factors and downstream manifestations under topology-constrained propagation [17]. Causal reasoning over knowledge graphs then offers a complementary inference layer for intervention-oriented attribution and ranking, providing methodological support for unifying root cause localization with impact analysis through structured causal-path reasoning [18].

To enhance robustness and interpretability of localization decisions, causal representation learning motivates learning latent factors that isolate mechanism-relevant variation from context-dependent confounders, thereby aligning anomaly scores with causal drivers rather than coincidental co-movements [19]. Causal-invariant retrieval-augmented modeling under distribution shift further suggests enforcing invariance across changing environments while grounding decisions in relevant historical evidence, which directly supports stable scoring and explanation of service anomalies as operating conditions evolve [20]. Knowledge-graph-driven generative modeling provides a methodological route for converting structured relational evidence into coherent, faithful explanations, enabling ranked candidates to be accompanied by topology-consistent rationales derived from paths and neighborhoods in the dependency structure [21]. Trustworthy summarization with uncertainty quantification extends this by encouraging calibrated confidence and risk-aware reporting, which improves operational reliability when observability is incomplete, noisy, or partially inconsistent with the assumed structure [22].

At the automation layer, trust-aware orchestration for multi-agent collaboration motivates robust coordination and policy execution under unreliable or adversarial conditions, informing how

diagnostic components can be composed into resilient end-to-end workflows that integrate detection, localization, and validation actions [23]. Multi-agent assistants for translating intent into deployable artifacts further inspire an operational pipeline in which diagnosis outputs can be systematically transformed into actionable changes, supporting closed-loop workflows that reduce manual intervention while preserving traceability [24]. Resource-aware inference considerations are addressed by proactive, fragmentation-aware serving systems, which motivate designing the representation and graph-reasoning modules to be deployable with elastic scheduling and lightweight updates in dynamic runtime environments [25]. Multi-scale LoRA fine-tuning strategies provide methodological support for parameter-efficient adaptation, enabling rapid and modular refinement of encoders or scoring components without expensive full-model retraining [26]. Transformer-based interaction sequence modeling offers additional evidence that long-range sequential behaviors can be compressed into stable latent states, reinforcing the feasibility of representing complex, multi-step system behaviors as service-level embeddings suitable for deviation scoring and attribution [27]. Finally, diffusion-based generative modeling with conditional control motivates controlled synthesis or completion of representations under constraints, which can be leveraged to improve robustness to missing modalities, enhance consistency of diagnostic views, or support explanation generation while respecting structural and contextual conditions [28].

III. Model Design

This method addresses the heterogeneous, complex, and dynamically evolving nature of multi-source observation data in microservice architectures, constructing a unified system-level modeling and root cause reasoning framework. The overall approach maps different observation sources, such as logs, metrics, and traces, to a shared representation space. Using the microservice dependency structure as a carrier, it organizes scattered local observations into a computable system state representation. First, using services or components as basic analysis units, multi-source observations for each service within a given time window are aligned and aggregated to form a service-level state description. Then, by explicitly modeling the calls and dependencies between services, the local state is embedded into the global structure, enabling anomaly information to propagate and be constrained along dependency paths. Throughout this process, the method does not rely on manual rules or static thresholds, but instead uses consistent mathematical representations to characterize "normal operating state" and "deviation patterns," providing a unified foundation for subsequent anomaly measurement and root cause localization, thereby avoiding information loss caused by fragmented multimodal data analysis. This paper also presents the overall model architecture, as shown in Figure 1.

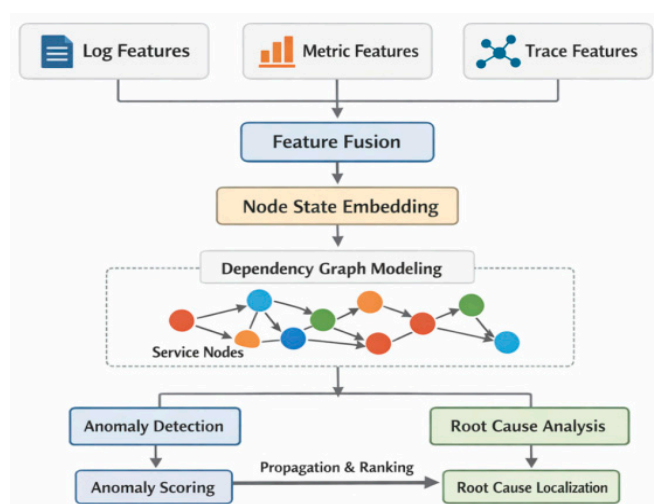


Figure 1. Overall model architecture.

At the representation level, suppose the system contains N microservice nodes at time t . For the i -th service, the feature vectors extracted from different observation sources are denoted as follows:

$$x_i^{(l)}(t), x_i^{(m)}(t), x_i^{(r)}(t) \quad (1)$$

The superscripts correspond to logs, metrics, and tracking observations, respectively. Through linear mapping and nonlinear transformation, these are uniformly mapped to the same latent space and fused to obtain the service-level representation:

$$h_i(t) = \phi(W_l x_i^{(l)}(t) + W_m x_i^{(m)}(t) + W_r x_i^{(r)}(t)) \quad (2)$$

where W_l, W_m, W_r is a learnable mapping matrix and $\phi(\cdot)$ is a nonlinear function. This representation achieves scale alignment and semantic unification of cross-modal features while maintaining the complementarity of information from different observation sources, providing a consistent input for system-level modeling.

To characterize the structural dependencies between microservices, a directed graph $G = (V, \varepsilon)$ is introduced, where the set of nodes V corresponds to service instances, and the set of edges ε represents calls or dependencies. Based on this structure, dependency-aware updates are performed on the service representation, ensuring that the state of each node is simultaneously influenced by its own observations and the states of its neighboring nodes. This is formally represented as:

$$z_i(t) = h_i(t) + \sum_{j \in N(i)} \alpha_{ij} h_j(t) \quad (3)$$

Here, $N(i)$ represents the set of neighbors that are dependent on node i , and α_{ij} is the dependency strength weight, used to characterize the degree of influence of different dependency paths on the current service state. This process enables anomalous signals to propagate and be modeled along the real system structure, thereby avoiding misjudging structural anomalies as isolated fluctuations.

During the root cause analysis phase, the source of anomalies is identified by measuring the deviation between the service's current state and its historical consistency pattern. An anomaly score is defined for each service:

$$s_i(t) = \|z_i(t) - \mu_i\|_2 \quad (4)$$

Here, μ_i represents the reference state of service i under stable operating conditions. Further, by combining the dependency structure, structural constraint propagation is performed on the anomaly scores to obtain the final root cause confidence score:

$$c_i(t) = s_i(t) \cdot (1 - \sum_{k \in P(i)} \beta_{ki}) \quad (5)$$

where $P(i)$ represents the upstream dependency set of node i , and β_{ki} is the proportion of anomalies propagating from upstream to the current node. This design prioritizes attributing anomalies to structurally more likely source nodes, rather than simply ranking them by anomaly magnitude, thereby achieving more reasonable and consistent root cause localization at the system level.

IV. Experimental Evaluation

A. Dataset

This study adopts the open source dataset Anomalies in Microservice Architecture based on version configurations of the train ticket system as the unified input benchmark for both data analysis and method validation. The dataset is collected from a representative microservice benchmark system. It is designed for scenarios that combine unified modeling of multi source observability data with intelligent fault root cause localization. The dataset captures common behaviors in microservice operations. These behaviors include cross service dependency propagation and global symptoms triggered by local anomalies.

In terms of data modalities, the dataset provides three core observation sources. They include logs, distributed tracing based on Jaeger, and metrics collected by Prometheus. These sources correspond to the three pillars of microservice observability. The data are organized into multiple subsets. In total, the dataset contains ten subsets. Each subset represents monitoring records under a specific system version or configuration change. The dataset documentation provides anomaly background information and identification hints. This design supports the full workflow within a single closed data loop. The workflow includes multimodal feature fusion, dependency graph modeling, anomaly scoring, and root cause localization. Logs provide strong semantic information. Metrics provide strong temporal signals. Traces provide explicit structural information. As a result, the dataset naturally supports the modeling objective of heterogeneous information complementarity combined with structure constrained reasoning.

From the perspective of data organization and usage, the dataset is structured according to change targets, dependent libraries or versions, and data collection time. Each subdirectory contains raw and structured log files, metric monitoring directories, and tracing data directories. This structure aligns directly with unified modeling at both service node level and invocation chain level. The data support service level state construction through aggregation of logs and metrics. They also support dependency recovery and propagation inference through traces or call based graph construction. Together, these properties provide the required observation, structure, and reasoning inputs for root cause localization. The dataset has a moderate scale, complete modalities, and open reproducibility. It is therefore well suited as the dataset description for the method section and as the basis for subsequent experimental settings.

B. Performance Results

This article first presents the results of the comparative experiments, as shown in Table 1.

Table 1. Comparative experimental results.

Method	Acc	Precision	Recall	AUC
Deeptralog[29]	0.842	0.831	0.815	0.876
TraceGra[30]	0.861	0.854	0.839	0.892
MADMM[31]	0.873	0.869	0.851	0.904
SRdetector[32]	0.889	0.883	0.872	0.918
Ours	0.912	0.905	0.896	0.941

An overall comparison indicates that the baseline methods show a gradual improvement in accuracy and discriminative performance. The proposed method exhibits more consistent behavior across all four metrics. This pattern suggests that unified modeling of multi source observability data combined with dependency aware reasoning captures cross service anomaly propagation more effectively. Compared with approaches that rely on a single observation source or weak fusion strategies, the results highlight the benefits of multimodal complementarity. The model maintains stable recognition performance even when complex invocation chains and multiple symptoms appear simultaneously. It also demonstrates stronger overall generalization.

The simultaneous improvement in precision and recall shows that the proposed method reduces false alarms while also lowering missed detections. This balance aligns well with the practical requirements of root cause localization in microservice systems. In such environments, fluctuations in isolated metrics or local log anomalies are often insufficient to indicate real faults. True root causes are usually associated with structural impact and propagation along service paths. A unified representation of multi source observations establishes consistent links among semantic signals, performance trends, and invocation paths. This capability helps distinguish transient noise from persistent anomalies. Alerts therefore become more reliable. The input for subsequent root cause

candidate ranking is also cleaner. The advantage observed in comprehensive metrics such as AUC implies that the proposed method preserves strong ranking capability and robustness under different threshold settings. This property is critical for root cause localization tasks. Such tasks rarely involve simple binary decisions. They require interpretable priority rankings across multiple services and anomaly signals. These rankings guide operators to inspect the most likely source components first. With unified modeling and explicit incorporation of dependency structure, anomaly scores form more reasonable distributions along propagation paths. The distinction between source nodes and affected nodes becomes clearer. This clarity improves the practical usability of localization results.

Further comparison among baseline methods reflects differences in fusion depth, use of structural information, and modeling of temporal consistency. These differences lead to uneven metric improvements. The proposed method outperforms others across all four metrics. This outcome indicates that the gains do not arise from isolated local optimizations. They result from the joint effect of unified representation, structural constraints, and anomaly propagation reasoning. Such coordination is closer to real microservice operation behavior. It enables the model to capture root cause signals more reliably under complex dependencies and dynamic conditions. This property provides stronger system level support for intelligent fault diagnosis.

The latent space dimension determines the expressive capacity and information compression strength of multi-source observation features in the unified representation space, thus directly affecting the ability of state representation to characterize anomalous semantics and structural dependencies. To evaluate the impact of this dimension selection on model behavior, it is necessary to examine the performance changes of the proposed method under different dimension configurations while keeping other settings consistent. The experimental results are shown in Figure 2.

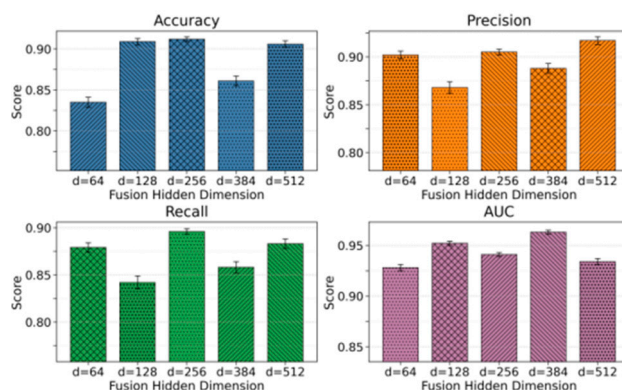


Figure 2. The impact of fusion of latent space dimensions on experimental results.

The overall trend shows that the dimension of the latent space has a clear impact on model behavior. Different dimensional settings lead to noticeable differences in system state representation capability. This observation indicates that, in unified modeling, the latent space serves both as a mechanism for information compression and as a factor that determines how fully multi source observations are expressed in a shared representation. When the dimension is too small, multimodal information is overly compressed. The complex relationships among log semantics, performance fluctuations, and invocation structure cannot be fully captured. The fluctuations observed across dimensions also indicate strong sensitivity to representation capacity. This behavior is consistent with the high diversity of anomaly patterns and the complex dependency structure of microservice systems. An examination of the trends across different subplots shows that evaluation metrics respond differently to changes in latent dimension. This suggests that each metric emphasizes different properties of system state. Some metrics rely more on stable global representation quality. They exhibit relatively smooth adjustments as the dimension changes. Other metrics are more sensitive to local representation capacity. They show stronger fluctuations when the dimension varies. This inconsistency reflects differing emphases of multi source observations within the unified

representation. It also suggests that a single dimensional configuration cannot optimally serve all discriminative objectives at the same time. When combined with the characteristics of dependency structure modeling in microservices, these variations can be interpreted as a balance between latent space capacity and structural information expression rather than random noise. Higher dimensions help preserve fine grained differences along cross service invocation paths. Anomalies become easier to distinguish during dependency propagation. At the same time, higher capacity may introduce redundancy or amplify local disturbances, which can affect overall stability. Lower dimensions suppress noise to some extent. However, they may weaken the ability to represent complex anomaly propagation patterns. Some structural information then becomes difficult to express explicitly.

The results confirm that latent space dimension is an important structural factor in unified modeling of multi source observability data. Its choice directly influences information fusion quality and the reliability of root cause localization. The differentiated fluctuations across metrics indicate that model performance does not simply improve with larger representation capacity. A reasonable balance is required between expressive power and structural constraints. This finding is highly consistent with the mechanisms of anomaly generation and propagation in microservice systems. It also provides clear guidance for parameter selection under different system scales and operational conditions.

The number of message passing layers in the graph determines the scope and depth of information aggregation on the dependency graph, thus affecting the model's ability to represent cross-service anomaly propagation paths and local perturbations. To evaluate the impact of this structural hyperparameter on the unified modeling process, it is necessary to observe the accuracy trend of the proposed method under different layer configurations while keeping other settings unchanged. The experimental results are shown in Figure 3.

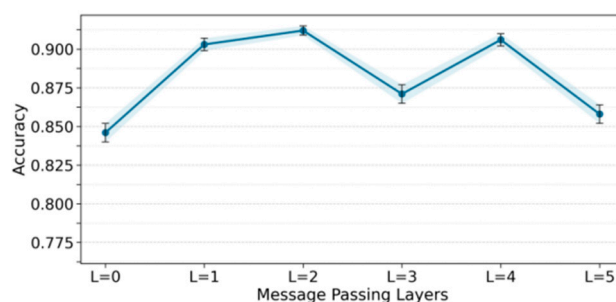


Figure 3. Experiment on the sensitivity of the number of message passing layers to accuracy.

The number of graph message passing layers has a clear influence on model performance, and the effect is non monotonic. This observation indicates that, when aggregating information over a microservice dependency graph, neither very shallow propagation nor unlimited depth leads to optimal results. An appropriate number of layers enables the model to establish effective connections between local service states and cross service dependencies. Anomaly signals can then be integrated gradually along real invocation paths. This process improves the representation of system level behavior.

When the number of message passing layers continues to increase, performance degradation and instability appear. This pattern suggests that excessive structural propagation introduces redundant information or noise. Over aggregation can reduce the discriminability of node specific observations. State representations of different services may become overly similar. In addition, not all long range paths in a microservice dependency graph are relevant to a given anomaly. Overall, the results confirm that the depth of graph message passing is a key structural hyperparameter that directly affects state representation quality in the unified modeling framework. The clear performance fluctuations across depths show that model effectiveness depends on a reasonable propagation setting rather than increased structural complexity. This behavior is consistent with the limited hierarchical depth of microservice dependencies and the locality of anomaly propagation

paths. It also provides practical guidance for parameter configuration under different system scales and topology complexities.

V. Conclusion

This work addresses key challenges in microservice architectures, including complex failure patterns, fragmented multi source observability, and difficulty in root cause localization. It proposes a unified modeling and intelligent root cause localization approach based on logs, metrics, and tracing data. By integrating heterogeneous observations at the system level and explicitly incorporating service dependency structure into state modeling and anomaly reasoning, the method provides a more complete characterization of anomaly generation and propagation in microservice systems. Comparative results demonstrate clear advantages in overall discriminative capability and result consistency. These findings confirm the effectiveness of combining unified representation with structural constraints for diagnosing complex distributed systems.

From a methodological perspective, the study treats a microservice system as a dynamically evolving whole rather than a collection of isolated components. Multi source observations are jointly modeled within a shared representation space. Dependency relations are used to constrain anomaly propagation paths. This design allows the model to preserve sensitivity to local signals while capturing system level anomaly patterns. Such a paradigm mitigates common issues in traditional approaches, including concentrated false alarms and dispersed root cause candidates. The resulting localization outputs better align with real operational logic and provide more reliable inputs for subsequent automated operations decisions. At the application level, the proposed approach is highly relevant to the stable operation of large scale cloud native systems. As microservice architectures are widely adopted in critical domains such as finance, industrial Internet platforms, and online services, system scale continues to grow and dependencies become deeper. Manual driven troubleshooting can no longer meet requirements for timeliness and accuracy. The unified modeling strategy presented in this study offers a feasible path toward intelligent and structure aware operational diagnosis. It helps reduce localization time, lower human intervention costs, and improve reliability and maintainability under complex operating conditions. Looking ahead, the framework provides substantial opportunities for further expansion of intelligent operations capabilities. One direction is to examine how unified modeling supports system evolution and long term stability under larger scale and more complex topologies. Another direction is to integrate this approach with tasks such as resource scheduling, capacity planning, and risk early warning. Such integration can promote a shift from reactive fault response to proactive operational optimization. With continued advances in cloud native technologies and observability platforms, root cause localization based on unified multi source modeling is expected to play a foundational role across broader application domains. It may become a key component in next generation intelligent operations for distributed systems.

References

1. Wang, Y., Yan, R., Xiao, Y., Li, J., Zhang, Z., and Wang, F., "Memory-driven agent planning for long-horizon tasks via hierarchical encoding and dynamic retrieval," 2025.
2. Xie, Z., Zhang, S., Geng, Y., et al., "Microservice root cause analysis with limited observability through intervention recognition in the latent space," Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 6049-6060, 2024.
3. Yao, Z., Pei, C., Chen, W., et al., "Chain-of-event: Interpretable root cause analysis for microservices through automatically learning weighted event causal graph," Companion Proceedings of the 32nd ACM International Conference on the Foundations of Software Engineering, pp. 50-61, 2024.
4. Zheng, L., Chen, Z., He, J., et al., "MULAN: Multi-modal causal structure learning and root cause analysis for microservice systems," Proceedings of the ACM Web Conference 2024, pp. 4107-4116, 2024.

5. Chen, R., Ren, J., Wang, L., et al., "Microegrc1: An edge-attention-based graph neural network approach for root cause localization in microservice systems," *Proceedings of the International Conference on Service-Oriented Computing*, Cham: Springer Nature Switzerland, pp. 264-272, 2022.
6. Q. Zhang, N. Lyu, L. Liu, Y. Wang, Z. Cheng, and C. Hua, "Graph neural AI with temporal dynamics for comprehensive anomaly detection in microservices," arXiv:2511.03285, 2025.
7. N. Lyu, J. Jiang, L. Chang, C. Shao, F. Chen, and C. Zhang, "Improving pattern recognition of scheduling anomalies through structure-aware and semantically-enhanced graphs," arXiv:2512.18673, 2025.
8. C. Hua, N. Lyu, C. Wang, and T. Yuan, "Deep learning framework for change-point detection in cloud-native Kubernetes node metrics using transformer architecture," 2025.
9. Z. Cheng, "Hierarchical attention-based modeling for intelligent scheduling delay prediction in complex backend systems," *Transactions on Computational and Scientific Methods*, vol. 4, no. 4, 2024.
10. Y. Shu, K. Zhou, Y. Ou, R. Yan, and S. Huang, "A self-supervised learning framework for robust anomaly detection in imbalanced and heterogeneous time-series data," 2025.
11. Y. Ou, S. Huang, R. Yan, K. Zhou, Y. Shu, and Y. Huang, "A residual-regulated machine learning method for non-stationary time series forecasting using second-order differencing," 2025.
12. Y. Wang, R. Fang, A. Xie, H. Feng, and J. Lai, "Dynamic anomaly identification in accounting transactions via multi-head self-attention networks," arXiv:2511.12122, 2025.
13. M. Raeiszadeh, A. Ebrahimzadeh, R. H. Glitho, et al., "Asynchronous real-time federated learning for anomaly detection in microservice cloud applications," *IEEE Transactions on Machine Learning in Communications and Networking*, 2025.
14. H. Liu, Y. Kang, and Y. Liu, "Privacy-preserving and communication-efficient federated learning for cloud-scale distributed intelligence," 2025.
15. Y. Wu, Y. Qin, X. Su, and Y. Lin, "Transformer-based risk monitoring for anti-money laundering with transaction graph integration," in *Proc. 2025 2nd Int. Conf. Digital Economy, Blockchain and Artificial Intelligence*, pp. 388-393, 2025.
16. K. Cao, Y. Zhao, H. Chen, X. Liang, Y. Zheng, and S. Huang, "Multi-hop relational modeling for credit fraud detection via graph neural networks," 2025.
17. Q. Gan, R. Ying, D. Li, Y. Wang, Q. Liu, and J. Li, "Dynamic spatiotemporal causal graph neural networks for corporate revenue forecasting," 2025.
18. R. Ying, Q. Liu, Y. Wang, and Y. Xiao, "AI-based causal reasoning over knowledge graphs for data-driven and intervention-oriented enterprise performance analysis," 2025.
19. J. Li, Q. Gan, R. Wu, C. Chen, R. Fang, and J. Lai, "Causal representation learning for robust and interpretable audit risk identification in financial systems," 2025.
20. S. Sun, "CIRR: Causal-invariant retrieval-augmented recommendation with faithful explanations under distribution shift," arXiv:2512.18683, 2025.
21. S. Long, K. Cao, X. Liang, Y. Zheng, Y. Yi, and R. Zhou, "Knowledge graph-driven generative framework for interpretable financial fraud detection," 2025.
22. S. Pan and D. Wu, "Trustworthy summarization via uncertainty quantification and risk awareness in large language models," arXiv:2510.01231, 2025.
23. Y. Hu, J. Li, K. Gao, Z. Zhang, H. Zhu, and X. Yan, "TrustOrch: A dynamic trust-aware orchestration framework for adversarially robust multi-agent collaboration," 2025.
24. T. Guan, "A multi-agent coding assistant for cloud-native development: from requirements to deployable microservices," 2025.
25. Y. Ni, X. Yang, Y. Tang, Z. Qiu, C. Wang, and T. Yuan, "Predictive-LoRA: A proactive and fragmentation-aware serverless inference system for LLMs," arXiv:2512.20210, 2025.
26. H. Zhang, L. Zhu, C. Peng, J. Zheng, J. Lin, and R. Bao, "Intelligent recommendation systems using multi-scale LoRA fine-tuning and large language models," 2025.
27. R. Liu, R. Zhang, and S. Wang, "Transformer-based modeling of user interaction sequences for dwell time prediction in human-computer interfaces," arXiv:2512.17149, 2025.
28. R. Liu, L. Yang, R. Zhang, and S. Wang, "Generative modeling of human-computer interfaces with diffusion processes and conditional control," arXiv:2601.06823, 2026.

29. Zhang, C., Peng, X., Sha, C., et al., "Deeptralog: Trace-log combined microservice anomaly detection through graph-based deep learning," Proceedings of the 44th International Conference on Software Engineering, pp. 623-634, 2022.
30. Chen, J., Liu, F., Jiang, J., et al., "TraceGra: A trace-based anomaly detection for microservice using graph deep learning," Computer Communications, vol. 204, pp. 109-117, 2023.
31. Wang, P., Zhang, X., Cao, Z., et al., "MADMM: Microservice system anomaly detection via multi-modal data and multi-feature extraction," Neural Computing and Applications, vol. 36, no. 25, pp. 15739-15757, 2024.
32. Ge, H., Ji, X., Peng, F., et al., "SRdetector: Sequence reconstruction method for microservice anomaly detection," Electronics, vol. 14, no. 1, p. 65, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.