Knowledge Graph Enhanced AIOps Framework for Root Cause Analysis in Network Slice Service Degradation

Jiank Leawis
School of Computing, Clemson University, Clemson, SC, USA.
lewis370@clemson.edu

Pavel May
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
maypavel@ucf.edu

Ole Woods
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
olemail@unh.edu

Abstract

The provisioning of network slices in fifth-generation and beyond communication systems introduces unprecedented flexibility in delivering tailored services with distinct performance guarantees. However, the inherent complexity of virtualized infrastructures and the tight coupling between logical slices and shared physical resources significantly complicate the detection and diagnosis of service degradation. Artificial intelligence for IT operations has emerged as a promising paradigm to automate operational tasks, yet conventional data-driven root cause analysis often suffers from a lack of semantic context, leading to ambiguous fault localization and reactive remediation. This paper proposes a knowledge graph enhanced AIOps framework specifically designed for root cause analysis in network slice service degradation. By integrating heterogeneous telemetry data, topological models, service level agreement policies, and historical incident records into a unified knowledge graph, the framework enables context-aware reasoning over complex dependencies. We examine architectural considerations, including the layered design of the ingestion, knowledge graph persistence, inference engine, and decision support subsystems. The discussion extends to structural trade-offs between centralized and distributed knowledge graph deployments, the balance between model interpretability and diagnostic accuracy, and the integration of graph neural networks for anomaly propagation tracing. Furthermore, we address operational governance aspects such as federated knowledge sharing across domains, zero-trust security principles, and the policy implications of AI-driven decision-making in regulated telecommunications environments. The framework emphasizes sustainability through resource-efficient knowledge maintenance and fairness in cross-tenant diagnostics. A comprehensive cross-domain comparison situates the proposed approach within the broader landscape of intelligent network management. The paper concludes with a forward-looking perspective on the role of knowledge graphs in achieving autonomous and trustworthy network operations.

Keywords

AIOps; knowledge graph; root cause analysis; network slicing; 5G; service degradation.

1. Introduction

The emergence of fifth-generation wireless networks has established network slicing as a cornerstone architectural paradigm, enabling the concurrent operation of multiple logical networks over a common physical infrastructure, each tailored to specific service requirements spanning enhanced mobile broadband, ultra-reliable low-latency communications, and massive machine-type communications [1]. The operational reality of managing these slices, however, reveals a profound increase in system intricacy, where service degradation can originate from a multitude of intertwined factors: virtual network function failures, resource contention at the radio access or core network, configuration drift, or unexpected interactions between slices sharing underlying hardware. As operators strive to meet stringent service level agreements, the ability to rapidly and accurately identify the root cause of performance anomalies becomes a critical competitive and regulatory necessity.

Artificial intelligence for IT operations has been progressively adopted to move beyond threshold-based alerting towards intelligent anomaly detection and diagnostic reasoning [2]. Early approaches in network fault diagnosis often relied on manually crafted dependency graphs or statistical correlation of metrics. Methods employing hierarchical causality graphs and log pattern analysis provided initial steps toward automated causal inference in distributed systems [3]. More recently, the application of graph neural networks has demonstrated considerable potential in learning latent representations of topological dependencies, proving effective in modeling the propagation of faults along service chains [4]. Simultaneously, machine learning techniques broadly applicable to fifth-generation networks have been extensively surveyed, highlighting both the opportunities of data-driven management and the challenges of data heterogeneity and scale [5]. In the specific domain of quality of service assurance, deep reinforcement learning has been employed to dynamically adjust slice configurations to maintain performance targets [6].

Despite these advances, purely data-driven AIOps pipelines frequently encounter a semantic gap. They operate on numerical telemetry streams or textual logs without a deeper understanding of the functional roles of network components, the contractual context of affected slices, or the operational policies governing resource arbitration. When a latency spike is detected, a model may point to a congested link or an overloaded container, but it struggles to articulate whether the degradation violates a specific customer’s latency budget or whether a pre-existing maintenance window exempts the event from triggering an emergency rollback. This lack of contextualization leads to high false-positive rates in alerting cycles and hampers the automation of trustworthy remediation actions.

To bridge this gap, we propose a knowledge graph enhanced AIOps framework that serves as a foundational semantic layer for root cause analysis in network slice service degradation. The framework ingests diverse operational data, unifies them into an evolving knowledge graph that captures entities, their interrelationships, and domain constraints, and then leverages both symbolicreasoning and graph-based learning to identify causal pathways. By maintaining a dynamic model of the operational domain, the framework provides a structured substrate for both machine learning algorithms and human operators to navigate the intricate web of dependencies that define network slice performance. The proposed system is not merely a monitoring tool but a decision support platform that aligns technical diagnostics with business and policy objectives, thereby advancing the goal of self-healing and autonomous networks.

2. System Architecture and Knowledge Graph Construction

The architectural design of the proposed framework follows a layered approach that separates concerns across data ingestion, semantic unification, analytical reasoning, and actionable insight generation. At the foundational layer, a distributed stream processing fabric ingests heterogeneous telemetry signals from the radio access network, transport stratum, core network functions, and the orchestration layer responsible for network slice lifecycle management. These signals encompass structured performance metrics such as throughput and latency, semi-structured event logs from containerized virtual network functions, and unstructured data including configuration snapshots and trouble ticket annotations. The ingestion subsystem applies schema-on-read principles and lightweight normalization to align temporal granularities and suppress transient noise, enabling the framework to accommodate the velocity and variety inherent in operational carrier-grade environments [7]. Importantly, the ingestion layer is designed with extensibility in mind, allowing new data sources to be incorporated without disrupting the continuous operation of the knowledge-centric reasoning pipeline.

Above the ingestion foundation lies the knowledge graph construction layer, which transforms raw telemetry into an interconnected and semantically rich representation of the managed domain. The graph schema is grounded in a pre-defined ontology that captures multiple entity classes, including network slices, slice subnet instances, virtualized network functions, physical nodes, logical links, service level agreement policies, incident records, and organizational responsibility domains. Relationships encode physical hosting, logical connectivity, protocol dependencies, operational ownership, and causal precedence. This ontological scaffolding draws from established work on network resource modeling and graph-based fault localization that demonstrated the efficacy of topological relationship capture in reducing diagnostic ambiguity [8, 9]. An essential characteristic of the construction process is the fusion of static infrastructure models with dynamically updated state information. When a network function scaling event occurs or a slice is reconfigured to adjust its resource profile, the graph is updated in near real-time through a change data capture mechanism that preserves historical state snapshots for retrospective analysis.

A significant architectural consideration is the trade-off between centralized and distributed knowledge graph persistence. A centralized graph database provides straightforward query semantics and transactional consistency, which are invaluable for global reasoning across slices that span multiple data centers. However, strict centralization can introduce latency and bandwidth overhead, particularly in large-scale deployments where radio access nodes may be geographically dispersed. A federated approach, wherein domain-specific subgraphs are maintained at the edge or in regional management clusters and periodically synchronized or queried through a semantic query broker, offers superior scalability and resilience against wide-area connectivity disturbances [10]. The framework supports both configurations, adopting a federated model with an eventual consistency protocol that preserves correctness for causal reasoning without imposing prohibitive synchronization delays. Special attention is given to data provenance tracking, as each fact in the knowledge graph is decorated with its source, timestamp, and confidence level, which later inform the reliability weighting during root cause inference.

The knowledge graph is further enriched through the integration of external domain knowledge and historical operational experience. Operator-defined policies for incident prioritization, standard operating procedures encoded as causal rules, and regulatory obligations such as data sovereignty constraints are expressed as graph annotations and rule

sets that operate in tandem with statistical models. Historical incident records are mined to extract recurring fault propagation patterns, which are then embedded as weighted causal edges linking classes of symptoms to underlying root cause categories. This symbiosis between human-authored knowledge and data-driven pattern extraction mitigates the cold-start problem that often plagues purely machine learning-based diagnostic systems while allowing the graph to evolve with the network. The maintenance of knowledge freshness becomes a sustainability concern; thus, the framework implements lifecycle management policies that archive stale entities and prune obsolete relationships, preventing the graph from accumulating unmanageable clutter while preserving the traceability required for compliance audits.

3. Root Cause Analysis Mechanism and Inference

The core diagnostic capability of the framework rests on a hybrid inference engine that combines the structured reasoning power of the knowledge graph with the pattern recognition strength of graph neural networks. When a service degradation event is detected by the anomaly monitoring subsystem, which itself relies on unsupervised behavioral baselining techniques attuned to the periodicity and trend characteristics of slice-level metrics, an incident context graph is dynamically extracted from the full knowledge graph. This subgraph encompasses all entities within a configurable topological radius of the affected slice, including co-located virtualized functions, shared physical hosts, upstream and downstream service chain components, and active maintenance or lifecycle operations that may coincide with the incident timeframe. By constraining the analysis to a relevant subgraph, the framework balances computational efficiency with diagnostic completeness, a critical consideration in live production environments where mean time to resolution directly impacts service availability [11].

Root cause analysis proceeds along two complementary pathways. The first is a symbolic reasoning path that applies a rule engine over the extracted subgraph to evaluate predefined causal hypotheses. These rules capture domain-specific failure modes such as memory exhaustion in a virtualized firewall cascading to throughput degradation for all slices that traverse it, or misconfigured quality-of-service marking at a particular gateway leading to dropped packets for slices belonging to a specific service class. The rule engine operates over typed relationships and attributes, enabling it to test conditions such as the co-occurrence of a known fault signature in a linked physical node and an alarm in a dependent network function. When a rule matches, the engine annotates the graph with a causal hypothesis node that links the suspected root cause to the observed symptoms, along with a confidence score derived from the historical accuracy of the rule and the reliability of the underlying telemetry.

The second pathway employs graph neural network models trained to identify fault propagation paths in complex topologies. These models treat the incident context graph as input and learn to assign a root cause probability to each node, considering both node-level features derived from alert severity and metric deviation vectors, and graph structural properties such as centrality, clustering coefficients, and shortest-path distances from the symptom node. The architecture draws from graph attention networks that weigh neighboring nodes according to their diagnostic relevance, a technique proven effective in microservice fault localization [12]. The graph neural network output is calibrated through conformal prediction to produce confidence intervals that quantify the uncertainty level, enabling the decision support layer to distinguish between high-confidence root cause candidates suitable for automated remediation and those requiring human validation. The hybrid design ensures

that the system benefits from the transparency and auditability of rule-based reasoning for well-understood failure scenarios while leveraging the generalizability of learned models for novel or ambiguous faults.

A pivotal challenge in this hybrid scheme is the alignment and fusion of evidence from the two pathways. The framework employs a Dempster-Shafer evidence combination approach that treats each hypothesis as a piece of evidence with an associated mass function, merging symbolic and neural inferences into a unified belief structure. This method naturally handles conflicting outputs, for instance when the rule engine confidently implicates a node that the graph neural network assigns low probability, by preserving the conflict as a separate mass of uncertainty rather than discarding one viewpoint prematurely. The fused belief values are mapped back onto the knowledge graph, creating a layered root cause landscape that visual analytics tools can explore from multiple diagnostic perspectives. This approach respects a fundamental design principle of the framework: all analytical output must remain traceable to its source data and reasoning steps, thereby upholding the explainability requirements that are increasingly mandated by telecommunications operational governance frameworks [13].

4. Operational Governance and Policy Integration

Deploying AIOps frameworks in production telecommunications environments inevitably raises critical governance concerns that extend beyond technical diagnostics. The framework incorporates a comprehensive governance model that addresses data sovereignty, security, fairness in multi-tenant diagnostics, and the sustainability of operational knowledge. In a network slicing context, a single physical infrastructure hosts slices belonging to different enterprises, each with distinct regulatory obligations. A slice might serve an industrial automation client with data localization requirements while coexisting with a public safety network demanding hardened availability. The knowledge graph must therefore enforce access controls that ensure diagnostic queries initiated by an operator do not inadvertently reveal sensitive topology or performance data belonging to other tenants [14]. To this end, the framework implements an attribute-based access control mechanism over the graph, where relationships and entities are tagged with security classifications, and inference processes run within bounded scopes defined by the requesting user's clearance and the incident's tenant context. This zero-trust approach extends to the model training pipelines, which employ federated learning techniques to refine the diagnostic graph neural network across different administrative domains without centralizing raw telemetry, thereby minimizing exposure risk while preserving collective intelligence.

Fairness in root cause analysis is an underexplored yet vital consideration in multi-tenant environments. If the underlying diagnostic model exhibits systematic bias toward attributing degradation to slices operated by smaller customers or to certain network functions that have historically been implicated more frequently, the resulting allocation of repair resources and potential financial penalties from service level agreement violations may become inequitable. The framework addresses this through fairness-aware model calibration, which monitors the distribution of root cause scores across different tenant categories and adjusts the inference output to correct for statistically significant disparities [15]. This is implemented as a post-processing fairness layer that receives the raw causal hypothesis probabilities and applies a constrained optimization to minimize deviation from the original ranking while ensuring that the root cause selection does not disproportionately disadvantage any tenant group. Additionally, policy modules within the knowledge graph capture fairness constraints as explicit rules, preventing, for example, the automated suspension of a slice without

considering the context of maintenance windows that may have been communicated to the customer.

The long-term sustainability of the knowledge graph itself demands rigorous governance of its growth and curation. As networks evolve through expansion and technology refreshes, the graph accumulates an ever-growing number of entities and relationships, many of which gradually lose relevance. Continuous integration of new vendor equipment models and software releases introduces schema drift that, if unmanaged, erodes the semantic consistency on which root cause analysis depends [16]. The framework addresses this through an automated graph quality monitoring module that tracks metrics such as orphaned nodes, inconsistent relationship cardinalities, and deprecated class usage. When quality indicators breach thresholds, curatorial workflows are triggered, which may involve automated pruning based on inactivity periods or flagging for operator review. These mechanisms align with broader sustainability principles by reducing the computational resources unnecessarily consumed by stale knowledge and ensuring that the knowledge graph remains a lightweight and agile digital twin rather than an unmanageable data lake.

The policy implications of AI-driven decision-making in regulated industries also extend to accountability and auditability. The framework preserves a tamper-resistant log of all root cause analysis conclusions, the evidence pathways that led to them, whether the diagnostic triggered an automated remediation action, and the ultimate outcome. This audit trail is linked to the same knowledge graph structure, effectively making operational decisions themselves first-class entities that can be queried for compliance reviews and post-incident analyses. By foregrounding traceability, the framework supports the evolving regulatory expectation that artificial intelligence in critical infrastructure must be explainable not only in a technical sense but also in a legal and procedural sense [17]. The integration of policy constraints as formal graph rules thus closes the gap between operational performance optimization and responsible stewardship of telecommunications infrastructure.

5. Cross-Domain Perspectives and Discussion

The knowledge graph enhanced AIOps framework for network slicing root cause analysis can be productively examined through cross-domain comparisons with other complex systems management paradigms. In cloud computing environments, similar challenges arise in diagnosing performance degradation across microservice architectures where distributed tracing, service mesh telemetry, and orchestration metadata collectively form a diagnostic surface [18]. Early cloud AIOps platforms relied heavily on unsupervised clustering of log data and time-series anomaly detection, yet they encountered the same semantic fragmentation that our framework addresses through graph-based unification. A notable parallel exists in the domain of industrial control systems, where digital twin architectures maintain synchronized representations of physical assets and their relationships, enabling what-if fault simulations and impact analysis [19]. The knowledge graph in our proposal serves a role analogous to the digital twin core, but extends it with richer policy semantics and the hybrid reasoning dualism that bridges operational technology and information technology management styles.

The integration of knowledge graphs with machine learning is gaining traction in several safety-critical sectors, offering insights into the robustness enhancements possible through structured knowledge infusion. In healthcare, similar architectures have been proposed to combine clinical knowledge graphs with patient monitoring data for diagnosis support, where the graph constrains the hypothesis space of a machine learning classifier to physiologically

plausible conditions [20]. The parallel to our network slice scenario is instructive: just as clinical graphs prevent absurd diagnostic associations, our network knowledge graph prevents root cause candidates that violate physical or logical constraints from being entertained, irrespective of how strongly a purely statistical model might favor them. This design principle of using knowledge as a guardrail rather than a sole oracle represents a mature approach to human-machine collaboration in high-stakes environments.

Another critical cross-domain lesson relates to the challenge of knowledge graph evolution under conditions of rapid change, which is common in telecommunications due to continuous feature rollouts and dynamic slicing. In the financial sector, where knowledge graphs are used for fraud detection, similar challenges are managed through incremental ontology evolution frameworks that allow new transaction patterns and regulatory rules to be assimilated without full graph reconstruction [21]. The framework's change data capture and versioned schema management draw from these practices, ensuring that root cause analysis remains accurate even as the underlying network service model diverges from its original design. Moreover, the use of graph neural networks trained on historical fault propagation patterns necessitates careful consideration of distributional shifts; as networks evolve, previously learned causal paths may become misleading. The framework's continuous learning loop re-evaluates model performance against fresh incident data and triggers retraining when concept drift is detected, a practice widely studied in the context of time-evolving graphs in social network analysis and bioinformatics.

The interdisciplinary nature of the proposed system invites reflection on the sociotechnical dimensions of automated root cause analysis. Extensive automation of incident diagnosis will inevitably reshape the operational workforce, shifting the role of network engineers from reactive troubleshooting to proactive governance of the AIOps platform itself [22]. This demands the establishment of algorithmic trust, which the framework fosters by exposing not merely the final root cause label but the entire evidence graph and reasoning chain. As telecommunications networks assume an increasingly vital role in societal functioning, the transparency of automated decisions extends beyond technical merit into public accountability. The policy layers embedded in our knowledge graph acknowledge this by enabling regulatory bodies or independent auditors to replay diagnostic scenarios and verify compliance with fairness and non-discrimination norms. Looking ahead, the convergence of network slicing with edge computing and artificial intelligence inference at the far edge will further expand the semantic scope of operational knowledge graphs, encompassing environmental data, user mobility traces, and application-level intents, thereby enabling an era of truly context-aware and autonomous service assurance [23].

6. Conclusion

The complexity of network slice operations in modern telecommunications systems demands a departure from purely reactive and statistically driven fault management toward semantically enriched, context-aware root cause analysis. In this paper, we have presented a knowledge graph enhanced AIOps framework that bridges the gap between rich operational semantics and scalable machine learning diagnostics. By grounding root cause analysis in a continuously updated knowledge graph that models infrastructure, policies, incidents, and tenant relationships, the framework enables a hybrid inference mechanism that combines the transparency of symbolic rules with the adaptive pattern learning of graph neural networks. The architectural design grapples with real-world constraints including distributed deployment feasibility, evidence fusion under uncertainty, and the governance imperatives of

fairness, security, and sustainability. We have argued that such a framework not only improves diagnostic speed and accuracy but also provides the auditable decision foundation required for trustworthy automation in multi-tenant, regulated environments.

The contributions of this work extend beyond a singular architectural blueprint. By embedding policy and fairness considerations directly into the root cause analysis workflow, we reposition the AIOps paradigm from a technical optimization tool to a sociotechnical governance instrument. The framework acknowledges that in a world of network slicing, where logical isolation intersects with physical sharing, the diagnosis of degradation is never purely a technical act; it carries economic, legal, and ethical weight. The integration of knowledge graphs thus serves as a manifestation of the principle that artificial intelligence in critical infrastructures must be designed not for autonomy in isolation, but for accountable and equitable operation. Future research directions include the development of cross-operator knowledge graph federation protocols that can diagnose faults spanning multiple service provider domains, the integration of natural language processing to mine equipment documentation and maintenance logs directly into the graph, and the study of adversarial robustness of graph neural network models against intentionally crafted telemetry deceptions. As networks evolve towards sixth-generation systems, the role of structured operational knowledge in maintaining the delicate balance between flexibility and stability will only intensify, underscoring the enduring relevance of knowledge-centric AIOps architectures.

References

1. Yang, L., Zhang, H., Li, X., & Ji, H. (2022). A comprehensive survey on network slicing in 5G and beyond: Principles, challenges, and opportunities. IEEE Communications Surveys & Tutorials, 24(2), 1028–1081.

2. Ding, R., Wang, Q., Dang, Y., Fu, Q., Zhang, H., & Zhang, D. (2023). AIOps: from reactive to proactive operations in cloud systems. ACM Computing Surveys, 55(9), 1–38.

3. Kobayashi, S., Otomo, K., & Fukuda, K. (2018). Causal analysis of network system faults using hierarchical causality graph and log data. IEEE/IFIP Network Operations and Management Symposium (NOMS), 1–7.

4. Ma, M., Yin, J., Wang, Y., Chen, X., Li, Z., & Zhang, Q. (2020). Diagnosing root causes of anomalies in microservice systems with graph neural networks. IEEE INFOCOM, 1941–1950.

5. Khan, R., Kumar, P., Jayakody, D. N. K., & Liyanage, M. (2021). A survey on machine learning for 5G and beyond: Use cases and future directions. IEEE Open Journal of the Communications Society, 2, 2174–2207.

6. Li, Q. (2026). QoS Assurance Mechanism for 5G Network Slicing Based on the Deep Reinforcement Learning PPO Algorithm. arXiv preprint arXiv:2605.03345.

7. Kreps, J., Narkhede, N., & Rao, J. (2011). Kafka: A distributed messaging system for log processing. Proceedings of the NetDB Workshop.

8. Chowdhury, S. R., Ahmed, R., Khan, M. A., Shahriar, N., Boutaba, R., & Mathieu, B. (2020). A scalable knowledge graph for network resource management. IEEE Transactions on Network and Service Management, 17(3), 1758–1773.

9. Tang, Y., Li, J., & Zhu, Y. (2021). Graph-based fault localization in cloud and telecom networks: A systematic review. Journal of Network and Systems Management, 29(4), 1–29.

10. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific American, 284(5), 34–43.

11. Gao, J., Xiao, X., & Li, Z. (2022). Real-time root cause localization in 5G network slices using attention-based graph autoencoders. IEEE Transactions on Network and Service Management, 19(4), 3994–4008.

12. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations (ICLR).

13. European Telecommunications Standards Institute. (2020). Zero-touch network and service management (ZSM): Requirements and architecture. ETSI GS ZSM 002.

14. Aviram, A., Sen, S., & Tse, D. (2022). Privacy-preserving AIOps for multi-tenant edge networks. ACM SIGCOMM Computer Communication Review, 52(4), 12–24.

15. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

16. Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43(5-6), 907–928.

17. Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems. Proceedings of the IEEE, 107(3), 509–517.

18. Soldani, J., & Brogi, A. (2022). An AIOps-driven approach for anomaly detection and root cause analysis in cloud-native applications. Journal of Systems and Software, 190, 111358.

19. Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4), 2405–2415.

20. Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., & Sontag, D. (2017). Learning a health knowledge graph from electronic medical records. Scientific Reports, 7(1), 5994.

21. Soylu, A., Modritscher, F., Wild, F., De Causmaecker, P., & Desmet, P. (2017). Mashup ecosystems by design: Enabling web-scale reuse of digital resources through ontologies. IEEE Internet Computing, 21(5), 34–43.

22. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019). Guidelines for human-AI interaction. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–13.

23. Porambage, P., & Liyanage, M. (2022). Orchestrating intelligent edge AI for 6G: A distributed architecture perspective. IEEE Communications Magazine, 60(8), 78–84.