Aim5B: AI Integrated Semantic Framework for 5G and Beyond Network Management

Thanveer Sulthana*, Ava Sharif Jourabchi*, Venkat Rao Manavarthi*, Jayadithya Nalajala*,
Ankitha Srirama Reddy*, Baek-Young Choi†, Sejun Song†
*Department of Computer Science, University of Missouri - Kansas City, MO, USA
†Department of Computer Science, Missouri University of Science and Technology, MO, USA
Email: *{tsbfg, asddx, vm7k4, jnd9g, asdc3}@umsystem.edu, †{choiby, songsej}@mst.edu

***Abstract*—Scalable, interpretable, and intelligent network monitoring and management are critical for 5G and future networks. This paper introduces Aim5B, an AI-integrated semantic framework for 5G and beyond network management to address these challenges. Aim5B processes unstructured logs from key 5G core network functions, and transforms them into a knowledge graph aligned with the semantic structure of control-plane events. Leveraging a large language model (LLM), Aim5B enables natural language queries to be translated into Cypher graph queries, facilitating precise log retrieval, event analysis, temporal correlation, and statistical summarization—without relying on static parsing rules or predefined dashboards. Integrated on a real private 5G testbed, Aim5B achieved the full performances in precision, recall, F1, and Jaccard metrics, across various detailed, domain-specific queries per network function component. Furthermore, it significantly reduces management traffic—by approximately 99.9%—through targeted, event-specific querying. These results demonstrate the effectiveness and efficiency of Aim5B as an intelligence-driven solution for real-time, scalable network observability in future mobile systems.**

***Index Terms*—5G/6G Network Management, Semantic Log Analysis, Knowledge Graphs, Syslog, Large Language Models, AI-Assisted Querying, Graph Databases, Network Observability, Event-Driven Monitoring.**

I. Introduction

5G and future 6G networks introduce unprecedented architectural complexity, service granularity, and dynamic behavior across core network functions. As a result, ensuring operational reliability and timely fault detection has become increasingly reliant on scalable, interpretable, and intelligent observability mechanisms [1]. Traditional monitoring approaches—such as metric-based dashboards, rule-based alert systems, and periodic polling—are ill-suited for modern 5G and beyond environments, where data is highly contextual, event-driven, and component-specific.

This paper presents *Aim5B*, *an Artificial Intelligence (AI) integrated semantic framework for 5G and beyond*, designed to address emerging scalability, interpretability, and observability challenges. As shown in Figure 1, Aim5B processes unstructured, verbose, and vendor-specific logs [2] from key 5G core components, including Access and Mobility Management Function (AMF), Network Repository Function (NRF), and User Plane Function (UPF). These logs are transformed into a knowledge graph (e.g., Figure 2) aligned with a semantic structure that reflects control-plane event dynamics. Extracting actionable insights from such logs in real-time remains a significant challenge, particularly given the Service-Based Architecture (SBA) of 5G, where multiple network functions interact and produce dense, interdependent event sequences via interfaces like the Service-Based Interface (SBI) [3].

To address this complexity, Aim5B integrates a large language model (LLM) agent that interprets natural language queries and translates them into Cypher queries for precise log retrieval, semantic analysis, temporal correlation, and statistical summarization. Unlike conventional solutions, Aim5B does not rely on static parsing rules or pre-configured dashboards. Instead, it enables dynamic, context-aware, query-driven observability that supports pattern matching, event sequence tracking, and component-specific analytics across distributed logs.

In Aim5B, network queries correspond to operator-driven diagnostic questions composed of a target network function, a control-plane procedure, and optional constraints such as time window or failure condition. Accurately answering such queries requires correlating multiple log events generated across different components and execution stages of a procedure. The knowledge graph captures these relationships by organizing log events into an event-centric semantic structure, enabling queries to be evaluated at the level of control-plane workflows rather than individual log lines. We developed 40 detailed, domain-specific queries per network function component by analyzing and reverse-mapping the log structures of AMF, NRF, and UPF. These queries were aligned with control-plane event semantics (e.g., registration failures, network function service status, SBI transaction rejections). To improve semantic alignment and LLM performance, we expanded the instruction prompt to incorporate 5G-specific log features such as message patterns, log-level distinctions, and logger/thread mappings. Additionally, semantic grounding was achieved by embedding consistent patterns and anchor keywords (e.g., "IMSI," "nrf-nfm," "PDUSession"), enabling reliable event extraction. By broadening the query set, we enhanced the LLM's ability to generalize, reducing ambiguity, false positives, and under-specified results. These enhancements enabled deterministic translation into correct Cypher queries for all 120 testbed scenarios.

Aim5B architecture also demonstrates performance scalability, supporting high-volume log ingestion and reducing redundant management traffic—a critical requirement in dense

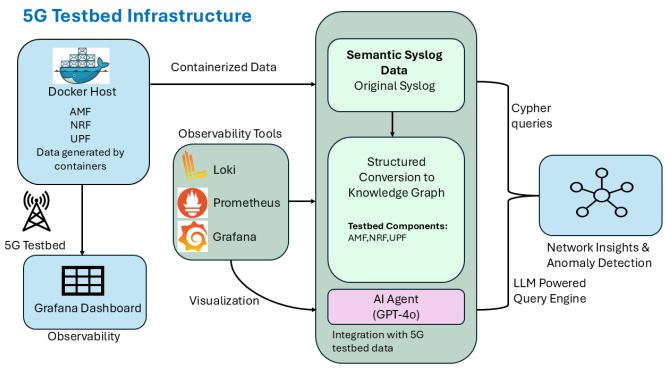


Fig. 1: Aim5B Architecture

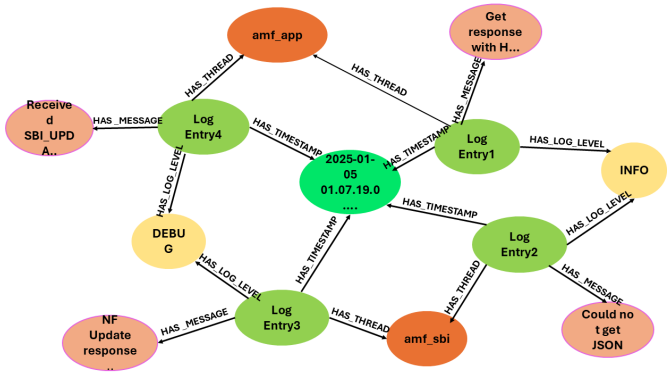


Fig. 2: A Knowledge Graph Example for 5G Testbed Data

5G deployments. We provide a comprehensive overview of the system architecture, log processing pipeline, knowledge graph construction, querying mechanism, and evaluation methodology. Aim5B was implemented and validated on a private 5G testbed. Experimental results show perfect accuracy across precision, recall, F1-score, and Jaccard metrics, for all developed queries, alongside a 99.9% reduction in management traffic through targeted, event-specific querying. These results highlight the effectiveness and scalability of Aim5B as a real-time, intelligence-driven observability solution for next-generation mobile networks. To the best of our knowledge, this is the first work that integrates an AI-agent, network monitoring and management functions, and validates on a real 5G testbed with real 5G network management data.

In this work, we focus on operator-driven, domain-specific diagnostic queries grounded in well-defined 5G control-plane procedures, rather than unrestricted open-domain natural language questioning. This design choice prioritizes deterministic behavior, reproducibility, and operational reliability, which are critical for practical network management. The proposed framework is evaluated using real logs from a private OAI-based 5G testbed, demonstrating its feasibility for high-level semantic diagnosis in realistic deployment settings.

The remainder of this paper is organized as follows: Section II reviews related work. Section III presents the proposed Aim5B system architecture. Section IV discusses our experimental 5G testbed, system setup, and extensive evaluation results. Finally, Section V concludes the paper.

II. Related Work

The use of AI in network management has been explored in many recent studies. Traditional log analysis tools, which depend on static rule sets or metric aggregation, fail to accommodate the semantic depth and context needed for modern 5G networks [4]. These limitations have encouraged the adoption of semantics-driven approaches using knowledge graphs and AI assistance. Graph-based models have shown promising results in log-based fault analysis for distributed systems, offering structured insight beyond simple parsing [5]. Within the 5G context, semantically enriched network telemetry has emerged as a powerful methodology for making observability both intelligent and intent-aware. [4] proposed a semantics-driven telemetry framework tailored for 5G, showing how structured metadata and semantic modeling improve the clarity and actionability of operational data [6].

Recent advancements in large language models (LLMs) have made it feasible to translate natural language queries into structured graph search expressions, such as Cypher or SPARQL [7]. When combined with retrieval-augmented generation techniques [8], these capabilities form a powerful foundation for real-time, intent-driven observability in telecom environments. The integration of knowledge graphs (KGs) with AI models has also proven effective across various domains, including network management. For example, Zhao et al. [9] and Thanveer et al. [10] showcased how an AI agent can interface with knowledge graphs to enable semantic querying over Syslog data through natural language interfaces. Prior research in network management approaches, however, have predominantly emphasized static analysis techniques or isolated applications of AI for specific tasks. These approaches often lack the flexibility and responsiveness required for real-time, adaptive management in dynamic mobile and wireless environments.

On the other hand, testbed-based approaches have emerged to facilitate real-world experimentation and scalability testing of 5G network functionalities. Mamushiane et al. [5] presented a full-stack 5G standalone (SA) testbed deployment using open-source platforms such as srsRAN and Open5GS, with a particular focus on user equipment (UE) integration and the practical challenges of RF tuning, error troubleshooting, and network slice enforcement at both the RAN and core levels. Their work provides valuable empirical insights into the complexities of building carrier-grade experimental setups capable of supporting semantically enriched telemetry and adaptive network behavior. Atalay et al. [6] conducted a resource-centric study using a containerized 5G testbed based on OpenAirInterface, analyzing compute overhead under various slicing configurations. Their findings revealed that resource consumption scales non-linearly with the number and specificity of slices, emphasizing the need for dynamic resource orchestration in semantically aware and AI-managed

5G environments. Existing testbed studies are mostly limited to evaluating specific network functions with limited components of open-source 5G networks.

To address the limitations of task-specific AI agents and the lack of comprehensive testbed implementations and evaluations, we integrate real-time AI-driven querying with a flexible knowledge graph (KG) architecture, and implement it on comprehensive 5G testbeds. This integration enables dynamic data retrieval and semantic analysis, overcoming the constraints of static rule-based methods and delivering enhanced adaptability, scalability, and precision in network observability and control. As far as we are aware, this is the first work to effectively integrate semantically structured knowledge graphs with LLM-based query generation for 5G network observability, and it was experimented on comprehensive open-source-based testbeds.

III. Aim5B System Architecture and Design

Aim5B enables intelligent querying and real-time analytics through a modular, scalable pipeline that ingests syslogs from a private 5G testbed, semantically structures the data, builds a knowledge graph, and employs an LLM agent for query-driven analysis across diverse 5G core logs. The system architecture is depicted in Figure 1.

A. Data Ingestion and Transformation

1) Syslog Data Collection: Logs are continuously collected from 5G testbed containerized deployments of key 3GPP-defined core functions such as AMF, NRF, and UPF components within the 5G testbed, which is deployed using OAI (OpenAirInterface) [11] for both Core and RAN in a containerized environment. These logs represent control-plane events such as UE registration, NF discovery, SBI transactions, and session management, which generates semantically rich syslog data essential for knowledge graph construction and semantic querying.

- **AMF** manages UE registration, authentication, and mobility, interfacing with SMF(Session Management Function) and NRF(Network Repository Function) via SBI(Service-Based Interface). In the testbed, it generates high-volume logs with rich 5GMM(5G Mobility Management) state transitions, IMSI handling, and NGAP/SBI interactions—characterized by temporal and contextual patterns well-suited for semantic graph analysis.
- **NRF** serves as a dynamic registry for 5G Core functions, handling NF discovery, registration, and heartbeats. Its logs capture events like Network Function (NF) registrations, discovery failures, and service status changes—supporting real-time queries on NF health and service availability.
- **UPF** anchors the data plane, handling packet routing, QoS enforcement, and PDU session management under SMF control [12]. Though less frequent, its logs capture key session state changes and data path setups, enabling cross-layer correlation with AMF and SMF events for queries like "Identify failed UPF session activations."

TABLE I: Overview of 5G Data Collection

Data Attribute	Description
Collection	Data collected from a private 5G testbed
Sources	AMF, NRF, UPF Logs
Type	Syslog data from control-plane functions
Duration	December 2024 – February 2025
Total Size	2.2 GB

2) Data Transformation: Given the heterogeneity of log formats across functions, the system applies a standardized transformation pipeline. Each raw log is parsed using a domain-specific regular expression and normalized into a consistent structure.These regular expressions are used solely for lightweight syntactic normalization and field extraction, such as parsing timestamps, severity levels, and message boundaries. They are not employed for semantic classification, event correlation, or query logic, which are handled entirely at the knowledge graph and query layers. This yields five structured fields, including timestamp, severity, thread, module, and message. These are enriched with component metadata (e.g., AMF, NRF, UPF) before entering the knowledge graph construction phase.

If log formats evolve due to software updates (e.g., changes in OAI versions), only the normalization layer requires adjustment, while the semantic schema and query logic remains unchanged.

B. Knowledge Graph Construction

The knowledge graph in Aim5B is generated by transforming raw 5G syslog entries into a structured, event-centric representation. Each log entry is first normalized into basic fields such as timestamp, severity, component, and message content. A corresponding event node is then created in the graph, while additional nodes capture contextual information including time, network function, and execution context. Relationships are used to explicitly connect events to their context and to each other, enabling the graph to represent how control-plane activities evolve across components and over time.

TABLE II: Summary of Accuracy Enhancement Stages

Methods	Description
FGQ	General-purpose log queries were decomposed into event-specific sub-queries. This step enabled precise alignment with distinct control-plane activities, such as session setup failures or NF service registration anomalies.
PSA	The AI agent's instructions were adapted to reflect the internal graph schema and log structures. This alignment minimized syntactic mismatches by guiding the agent to reference nodes and edges accurately in Cypher queries.
SPG	Regular expression patterns and domain-specific keywords (e.g., "IMSI", "nrf-nfm", "PDU Session") were embedded into the query generation framework. These semantic anchors allowed consistent recognition of log semantics across noisy or variable messages.
LTQE	Deterministic Cypher generation was enforced using low-temperature prompting, along with validation steps to ensure query correctness. This included a self-checking loop within the agent, ensuring output stability and eliminating false matches.

In Aim5B, semantics are inferred deterministically rather than learned statistically. The semantic meaning of each event is derived directly from explicit log attributes and identifiers defined by 3GPP control-plane procedures, such as IMSI values, network function instance identifiers, service names, and SBI transaction information. These attributes provide unambiguous cues about the role and context of each event within a control-plane workflow.

Each structured log entry is mapped to a semantic graph representation using a graph database, Neo4j. The graph schema is component-agnostic and event-centric, enabling precise modeling of control-plane behavior across the 5G core [13]. The schema includes the following node types: `Event`, `Time`, `Severity`, `ExecutionContext`, `ServiceModule`, and `EventDetail`. Relationships capture temporal, contextual, and semantic dependencies among nodes [14].

By encoding control-plane events and their semantic relationships as nodes and edges, the knowledge graph enables structured reasoning over 5G workflows. When a query is issued, the system translates the user's intent into graph constraints and retrieves only those event subgraphs that match the specified procedure, components, and conditions. This semantic organization allows Aim5B to answer complex diagnostic queries accurately without scanning raw log streams, thereby improving both precision and efficiency.

C. AI Agent for Querying and Analysis

A domain-adapted LLM agent interfaces with the knowledge graph to support natural language interaction. The agent interprets user queries and generates corresponding Cypher queries. It is fine-tuned with 5G-specific patterns and message structures and maps semantic intent to graph constraints [15].

D. Execution Flow

The user provides a query in natural language. The AI agent parses the intent, generates a Cypher query, retrieves relevant data, and processes the output using the appropriate analysis tools. This flow supports both low-latency responses and deep log mining for scalable network intelligence.

TABLE III: Retrieval Performance Metrics and Results

Metric	**Scores (BL to All)**		
	AMF	**NRF**	**UPF**
Macro-Averaged Precision	0.87 to 1	0.82 to 1	0.84 to 1
Macro-Averaged Recall	0.78 to 1	0.76 to 1	0.77 to 1
Macro-Averaged F1 Score	0.80 to 1	0.78 to 1	0.79 to 1
Macro-Averaged Jaccard Index	0.70 to 1	0.68 to 1	0.69 to 1
Cumulative Accuracy Change	0.0000	0.0000	0.0000
Slope of Accuracy over Time	0.0000	0.0000	0.0000

TABLE IV: Per-Query Log Data Reduction Summary

Query Theme	**Raw Log (Lines)**	**Filtered Log (Lines)**	**Reduction (%)**
All UPF messages from NRF	1000	125	87.5%
NRF Heartbeat Communication	1000	375	62.5%
PDU Session Establishment	1000	125	87.5%
UE Context Release	1000	1	***99.9%***

E. 5G Testbed Environment

The 5G testbed is deployed using the OAI platform, which provides open-source implementations of both the 5G Core and RAN. The testbed includes containerized deployments of key network functions, including AMF, NRF, and UPF. Each function is executed within an isolated container and instrumented to emit syslog messages during control-plane procedures such as UE registration, NF discovery, and session establishment [16]. Log messages capture internal events and state transitions defined by 3GPP-compliant signaling workflows. A centralized observability stack—comprising Prometheus for metric collection and Loki for log aggregation—collects these syslogs in real time. While traditional observability tools such as Grafana support metric visualization, they do not enable semantic interpretation or querying across event relationships. The paper addresses this limitation through structured log modeling and AI-driven query support.

F. AI Assisted Framework

This framework ingests raw syslog entries and transforms them into a labeled property graph stored in Neo4j. Each log is parsed into fields including timestamp, severity level, thread, logger, and message content. These fields are instantiated as nodes, with edges representing semantic and temporal relationships. To enable intuitive log interrogation, a language model–aligned agent is integrated as a query interface. The agent translates natural language questions into Cypher queries conforming to the graph schema. It is instruction-tuned to incorporate domain-specific context (e.g., 5G-specific log patterns) and maintain syntactic and semantic validity. Users can issue queries related to 5G operations—such as "List all NF registration failures" or "Show PDU session setup events"—without any prior knowledge of the graph database or query language. By abstracting syslog data into a semantic graph and enabling natural language interaction, this work provides a scalable and interpretable approach to log analysis that significantly exceeds the capabilities of conventional metric dashboards.

G. Integration with 5G Testbeds

The integration of Aim5B with the 5G testbed marks a significant advancement in its applicability, extending its capabilities beyond static or emulated log environments to real-time, standards-compliant 5G core operations. Unlike traditional monitoring tools, which rely on rule-based correlation or metric-level dashboards, this extension enables semantic-level understanding of syslog data generated during complex 5G procedures. By adapting Aim5B's pipeline to ingest logs from containerized AMF, NRF, and UPF components running within an OAI-based 5G testbed, the framework is validated in a dynamic, service-based architecture that reflects realistic network behavior.

This integration is particularly impactful due to the heterogeneous nature of 5G logs, which include tightly coupled interactions across the SBI interface, service registration events,

and NF-level status transitions. Through semantic preprocessing and transformation, Aim5B converts these verbose and implementation-specific log messages into a structured graph representation, capturing both temporal flow and operational semantics. As a result, the framework enables fine-grained query generation over 5G control-plane events—allowing users to retrieve, for example, only failed NF service updates or registration attempts by specific UEs—without relying on static parsing rules or handcrafted filters. Moreover, Aim5B's language model–driven agent remains fully operable in this extended context. By instruction-tuning the model to handle domain-specific 5G terminology and behavior patterns, the system maintains high precision and recall in query interpretation, even across logs that exhibit variability due to component configurations or SBI transaction flows. Experimental results from over 40 targeted queries on AMF, NRF, and UPF logs confirm that the system maintains its core strengths—semantic robustness, flexible query execution, and high interpretability—even in environments with dense signaling and inter-component interactions. Through this integration, Aim5B demonstrates not only its adaptability but also its relevance to real-world 5G deployments. The framework bridges the gap between low-level observability and high-level semantic reasoning, offering a scalable, intelligent log analysis platform for next-generation network management [17].

IV. Experiments with 5G Testbeds

In this section, we discuss our extensive experimental results using a private 5G testbed with both Core and RAN. The testbed logs used in this evaluation were generated from a private 5G testbed, including procedures such as UE registration, PDU session management, service discovery, and SBI-based NF interactions. Logs from AMF, NRF, and UPF were individually parsed, transformed, and ingested into separate knowledge graphs using the processing pipeline Semantic Network Management framework. The datasets used in our study are summarized in Table I.

A. Testing Environment

The evaluation was conducted within a controlled environment designed to replicate real-world conditions [18]. The testing setup included the following components:

- **Graph Database:** A dedicated server hosted the Neo4j database, ensuring optimal performance and accurate query execution.
- **LLM Model:** An AI agent, built on the GPT-4o model, was deployed within a Streamlit application. The agent processed user queries, generated Cypher queries, and performed subsequent analyses.
- **5G Datasets:** A combination of datasets from three 5G core network functions were used as below.
  1) **AMF:** Registration rejections, authentication steps, context releases, and SBI signaling
  2) **NRF:** NF service registration failures, heartbeat losses, discovery patterns
  3) **UPF:** Session context activation, interface status changes, data-path state transitions

The evaluation relies on domain-specific management queries designed to reflect realistic operator diagnostic tasks grounded in 5G control-plane procedures. Aim5B is not intended to support unrestricted or arbitrary natural language questioning; instead, it prioritizes deterministic, interpretable diagnosis of well-defined control-plane events. This design choice enables reliable semantic querying and avoids ambiguous responses. Handling completely arbitrary queries or unforeseen log patterns is outside the scope of this study and is identified as future work.

B. Semantic Query Response Accuracy and Enhancements

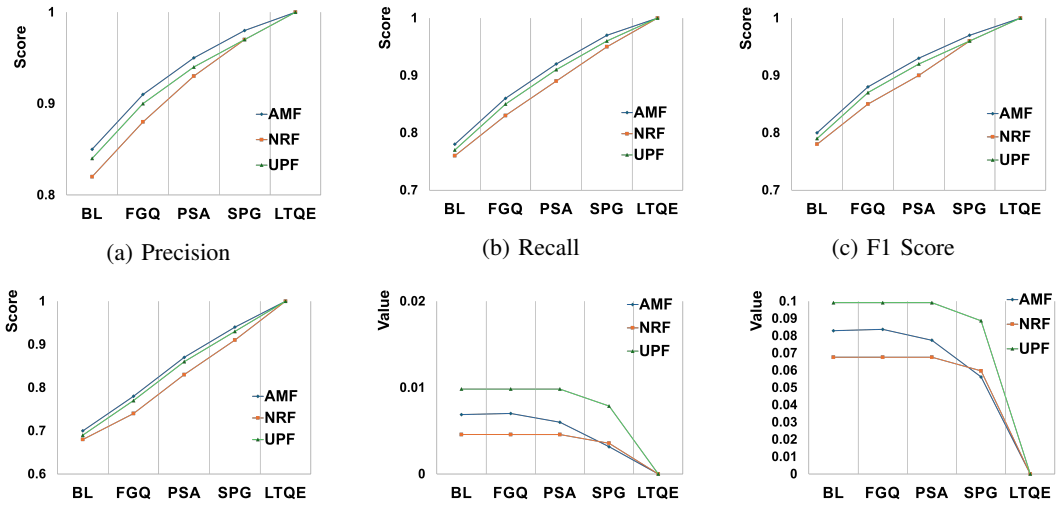
The proposed Aim5B framework was evaluated using 120 crafted, domain-specific network management queries — 40 each for AMF, NRF, and UPF —covering key event types such as registration attempts, SBI failures, PDU session initiations, and NF status transitions.

Aim5B generates deterministic Cypher queries over a fixed knowledge graph schema and executes them on a graph instance; therefore, the reported metrics reflect query formulation correctness rather than statistical generalization, resulting in deterministic, error-free retrieval for the evaluated management query set.

Table III summarizes the accuracy trajectory using Macro-Averaged Precision, Recall, F1 Score, Jaccard Index, Cumulative Accuracy Change, and Accuracy Slope over Time metrics. Beginning with an initial accuracy of approximately $68\% \sim 87\%$, the system ultimately achieved 100% across all metrics. These results confirm that the instruction-tuned AI agent can consistently generate accurate Cypher queries over the constructed KG. To achieve this level of accuracy and eliminate query variability, we employed a staged enhancement strategy comprising Fine-Grained Queries (FGQ), Prompting and Schema Alignment (PSA), Semantic Pattern Grounding (SPG), and LLM-Tuned Query Execution (LTQE), as detailed in Table II.

Ground truth for each query is defined based on explicit control-plane event semantics specified by 3GPP procedures and reflected in OAI logs. Each query targets a bounded event type identified through unambiguous log markers, such as IMSI values, network function instance IDs, service names, and SBI transaction identifiers.

Figures 3a and 3b show consistent improvements in Precision and Recall across all datasets as each enhancement stage is applied. Starting with modest baseline (BL) accuracy, performance steadily improved with each enhancement, ultimately achieving perfect scores (1.0) after the LTQE stage. This progression demonstrates the effectiveness of the enhancement pipeline across diverse log types when guided by structured semantic representations and refined AI-driven query generation. Figure 3c presents the F1 Score, which balances Precision and Recall. At the BL stage, F1 scores were below 0.85 across all datasets, indicating trade-offs between false positives and missed detections. The FGQ and PSA



(a) Precision (b) Recall (c) F1 Score

(d) Jaccard (e) Variation (f) Standard Deviation

Fig. 3: The Impacts of the Accumulative Enhancement Techniques (BL + FGQ + PSA + SPG + LTQE) on Precision, Recall, F1 Score, Jaccard, Variation, and Standard Deviation.

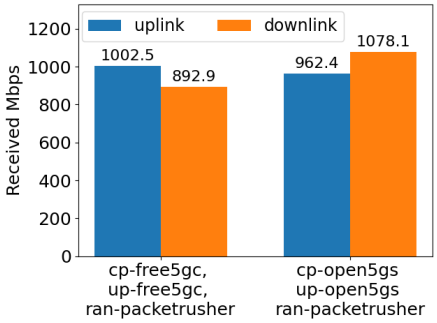


Fig. 4: Received Throughput Comparison: 5G Core

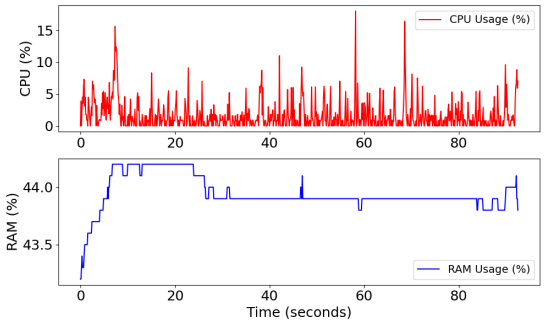


Fig. 5: CPU and Memory Usage over Time during Semantic Query Execution on 5G Testbed Logs

stages significantly improved F1 by jointly boosting both underlying metrics. The final F1 score of 1.0 demonstrates an ideal balance and perfect retrieval performance. As the most sensitive metric to partial mismatches, the Jaccard Index—shown in Figure 3d—revealed the largest initial gap (AMF: 0.70, NRF: 0.68, UPF: 0.69), indicating limited overlap between retrieved and ground truth entries at BL. However, the Jaccard score improved consistently with each enhancement, reaching 1.0 after LTQE for all components. This reflects complete semantic alignment between AI-generated query outputs and the annotated ground truth.

As enhancements were progressively applied, query accuracies increased and variabilities steadily declined across all datasets, ultimately reaching zero at the final LTQE stage—indicating full convergence and stability. Figures 3e and 3f illustrate that early enhancements (e.g., FGQ and PSA) had limited impact on reducing fluctuations, with the exception of PSA for the AMF dataset. In contrast, later stages (SPG and LTQE) significantly improved consistency, demonstrating that the model becomes increasingly deterministic and reliable in generating correct Cypher queries.

C. Impact of Semantic Log Reduction on 5G System Efficiency

In this work, management traffic refers to the volume of log data—measured in log lines and bytes—scanned, transmitted, and processed to answer a network management query. Aim5B reduces this traffic by retrieving only semantically relevant event subgraphs instead of full log streams.

To complement the evaluation of semantic log optimization, we assessed user-plane throughput performance across two 5G core implementations—free5GC [19] and Open5GS [20] —using RAN-PacketRusher [21] in a no-isolation setup. As illustrated in Figure 4, free5GC showed slightly higher uplink performance, while Open5GS exhibited a notable advantage in downlink throughput, reflecting architectural differences in session handling and resource management. Although these measurements originate from the user plane, they are significantly influenced by control-plane efficiency, including signaling latency and log processing overhead. Our semantic

framework mitigates this overhead by converting raw syslog data into structured knowledge graphs and supporting targeted, event-level querying. Rather than scanning thousands of log lines per event, Aim5B leverages Cypher queries over semantically organized logs, reducing irrelevant data processing, enhancing real-time responsiveness, and minimizing control-plane delays.

The throughput measurements provide indirect evidence that reducing control-plane log processing overhead can contribute to improved system efficiency; however, they are not intended as a direct performance benchmark of the semantic querying framework itself.

Table IV demonstrates significant log data reduction achieved through semantic filtering of specific 5G core events. For instance, UE Context Release queries were reduced by 99.9%, while PDU Session Establishment and NRF-UPF interactions saw reductions of 87.5%, highlighting the efficiency of targeted, query-driven retrieval. This reduction in raw data volume also translated into lower computational demands. Figure 5 shows that the resource usage remained modest, with peak CPU utilization at only 16% and RAM usage around 2%, which indicates Aim5B is lightweight and imposes little overhead on an existing system.

Aim5B is designed to support high-level semantic diagnosis of 5G control-plane behavior rather than low-level log retrieval or pattern matching. Traditional approaches such as Loki/Grafana-based keyword querying, regex or template-driven parsing, and manually written Cypher queries are effective for retrieving individual log entries but operate at a lower abstraction level. In contrast, Aim5B reasons over control-plane procedures that span multiple network functions and time intervals. As this paper focuses on validating the feasibility and correctness of procedure-level semantic diagnosis on a real 5G testbed, detailed baseline comparisons are left for future work.

V. Conclusions

This paper presented Aim5B, an AI-integrated semantic framework for 5G and beyond network management. This study focuses on AMF, NRF, and UPF, as these components generate the most semantically rich and interdependent control-plane logs in a 5G core network, covering authentication, service discovery, and session-related signaling. While other components such as SMF or RAN (gNB) logs are not included in the current evaluation, the proposed framework is component-agnostic and can be extended using the same semantic processing pipeline. Aim5B enables natural language-driven insights into control-plane events without relying on static parsing rules or pre-configured dashboards, through a novel integration of semantic structuring and LLM-driven query translation. It has been integrated into and evaluated on a real 5G testbed using real data and with practical management queries. With several enhancements tailored for 5G network management, Aim5B achieves a substantial 99.9% reduction in management traffic through targeted, event-specific querying. These validate the effectiveness and scalability of Aim5B as a practical, intelligence-based solution for real-time network observability and management in future mobile systems. The proposed Aim5B also offers valuable insights for future 6G networks, which are expected to be more complex and require self-optimizing capabilities, where AI must efficiently interpret unstructured network management data in real time. Overall, the experimental results demonstrate both the practical feasibility and correctness of Aim5B for semantic querying and high-level diagnosis of real 5G testbed logs.

References

[1] H. T. Nguyen, V. H. Dang, and N. D. Le, "5g and beyond: A survey on recent advances and future trends in mobile networks," in *IEEE Access*, 2021.

[2] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2019.

[3] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 416–464, 2018.

[4] V. Bajpai, J. Ott, A. M. Kakhki, and A. Dhamdhere, "Semantically enriched network telemetry for 5g observability," *IEEE Communications Magazine*, vol. 59, no. 3, pp. 34–40, 2021.

[5] L. Mamushiane, A. Lysko, H. Kobo, and J. Mwangama, "Deploying a stable 5g sa testbed using srsran and open5gs: Ue integration and troubleshooting towards network slicing," in *2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*. IEEE, 2023, pp. 1–6.

[6] T. O. Atalay, D. Stojadinovic, A. Stavrou, and H. Wang, "Scaling network slices with a 5g testbed: A resource consumption study," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 2649–2654.

[7] L. Jiang and J. Wang, "A survey on large language models in system observability," *IEEE Internet Computing*, 2022.

[8] X. Cheng and et al., "Network telemetry: Towards ai-driven monitoring and control of future networks," *IEEE Network*, 2021.

[9] L. Zhao and Y. Liu, "Integrating knowledge graphs with machine learning for network anomaly detection," *Journal of Network and Computer Applications*, vol. 147, pp. 102–114, 2021.

[10] T. Sulthana, A. S. Jourabchi, S. Song, and B.-Y. Choi, "Sinema: Semantics-driven intelligent network management using ai assistance," in *2024 IEEE 10th International Conference on Collaboration and Internet Computing (CIC)*, 2024, pp. 35–43.

[11] OpenAirInterface Software Alliance, "Openairinterface: Open-source 5g software suite," https://www.openairinterface.org/, accessed: 2025-05.

[12] V. Klyuev and H. Yamashita, "Modeling control plane behavior in 5g networks," in *2020 13th International Conference on Ubi-media Computing and Workshops*, 2020.

[13] Y. Ding and et al., "A survey on integrating semantic knowledge into network analysis," *Computer Networks*, vol. 191, p. 108009, 2021.

[14] M. Sallam and et al., "On the use of semantic knowledge graphs for ai-powered cybersecurity," *Future Generation Computer Systems*, 2021.

[15] J. Zhou and et al., "Graph neural networks: A review of methods and applications," *arXiv preprint arXiv:2004.00387*, 2020. [Online]. Available: https://arxiv.org/abs/2004.00387

[16] K. Zhang and et al., "Graphrag for log intelligence: Combining retrieval-augmented generation with graphs," in *IEEE ICDE*, 2023.

[17] T. Wang and et al., "Benchmarking llm-based agents in operational networks," in *IEEE CNS*, 2020.

[18] M. Tuan and et al., "Real-time observability in 5g using intelligent agents," in *IEEE ICC 2021*, 2021.

[19] "free5GC Project," https://www.free5gc.org/.

[20] "Open5GS: Open Source 5G Core and EPC Implementation," https://open5gs.org/.

[21] V. D'Emmanuele and A. Raguideau, "PacketRusher: High Performance 5G UE/gNB Simulator and CP/UP Load Tester," https://github.com/HewlettPackard/PacketRusher.