

DAGR: Agentic RCA for Internet-Scale Services

Sayan Sinha^{1,2}, Vipul Harsh^{1,3}, B Aditya Prakash^{1,2}, Vyas Sekar^{1,3}, Hui Zhang^{1,3}

¹Conviva

²Georgia Tech

³Carnegie Mellon University

ABSTRACT

Root cause analysis (RCA) is critical for restoring reliability when Internet-scale service KPIs degrade. LLM-based RCA agents can translate natural-language intent into executable investigations, but unconstrained agents are expensive and difficult to verify. We present DAGR, an agentic RCA system organized around four design requirements: expressivity, efficiency, explainability, and reduced practitioners’ effort. DAGR constrains the LLM to emit typed, loop-free RCA programs that compile into a verifiable dataflow DAG of side-effect-free operators, producing explicit evidence traces instead of only post-hoc natural-language rationales. On labeled incidents generated by client and backend-side anomaly injection in DeathStarBench’s Social Network application, DAGR improves end-to-end RCA accuracy by about 9.5× over OpenRCA while reducing estimated LLM cost by about 10×.

1 INTRODUCTION

Internet services serve millions of users across diverse devices, networks, and geographies. Providers track key performance indicators (KPIs) such as error rate and end-to-end latency; when KPIs degrade, incidents demand rapid root cause analysis (RCA). The challenge lies in navigating an enormous hypothesis space under time pressure while separating true causes from correlated symptoms and confounders.

Prior work offers three broad families of solutions. Specialized RCA algorithms can be efficient and explainable but cover a limited incident space. General analytics frameworks execute reliably but require operators to translate intuition into complex scripts during outages. LLM-based agents reduce front-end effort and broaden coverage, but unconstrained generation can hallucinate, waste tokens, and produce results that are difficult to verify. In practice, practitioners want four properties at once: *expressivity*, the ability to map incident descriptions into multi-step investigations; *efficiency*, low token, runtime, and compute cost; *explainability*, explicit evidence from executed operators and intermediate results; and low operator *effort*, meaning less manual scripting during an outage. These goals are not new; however the challenge is achieving them jointly in one RCA system. As shown in §1, existing approaches typically satisfy only subsets of them.

We hypothesize this trade-off is not inevitable. We present DAGR, which occupies the middle ground: it uses LLM world

System Class	Expressive	Efficient	Explainable	Effort
Specialized algorithms (e.g., Sage [1], Murphy [3])	X	✓	✓	X
Frameworks for authoring data analysis workflows (e.g., SQL)	X	✓	✓	X
Agent-based RCA (e.g., OpenRCA [5], RCAgent [4])	✓	X	X	✓
Ours (DAGR)	✓	✓	✓	✓

Table 1: Comparison of related works regarding the four key design requirements for RCA.

knowledge for hypothesis formation and workflow synthesis, but executes only verified workflows grounded in operator intent and local telemetry. Specifically, DAGR defines a typed, loop-free DSL whose programs compile into a dataflow DAG of side-effect-free operators that generate hypotheses (e.g., client cohorts, backend services, URLs, and behavioral event patterns), score and prune candidates, and summarize evidence. Given an incident description and optional analyst hint, the LLM emits a typed DSL program that specifies what will be computed and how results will be combined. A DSL verifier checks that every operator is valid, types compose correctly, and the graph is acyclic before execution. The loop-free restriction keeps execution bounded and auditable during incident response, while side-effect-free operators make evidence extraction reproducible and safe. The execution engine then runs the verified DAG over local telemetry and returns ranked hypotheses together with the intermediate evidence supporting them.

We evaluate our system using labeled incidents generated via anomaly injection on DeathStarBench’s Social Network application [2]. Compared to OpenRCA [5], an LLM baseline that ranks hypotheses directly from serialized telemetry, DAGR improves end-to-end RCA accuracy by about 9.5× on average while reducing estimated LLM cost by about 10×.

2 CONCLUSION

We presented DAGR, an approach to agentic RCA that jointly targets expressivity, efficiency, explainability, and reduced operator effort. DAGR constrains the LLM to a typed, loop-free DSL that compiles to a verifiable dataflow DAG of side-effect-free operators over local telemetry. This design improves auditability through explicit evidence traces while improving accuracy and reducing cost relative to an unconstrained agent baseline.

REFERENCES

- [1] Yu Gan, Mingyu Liang, Sundar Dev, David Lo, and Christina Delimitrou. 2021. Sage: practical and scalable ML-driven performance debugging in microservices. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*. 135–151.
- [2] Yu Gan, Yanqi Zhang, Dailun Cheng, Ankitha Shetty, Priyal Rathi, Nayan Katarki, Ariana Bruno, Justin Hu, Brian Ritchken, Brendon Jackson, et al. 2019. An open-source benchmark suite for microservices and their hardware-software implications for cloud & edge systems. In *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*. 3–18.
- [3] Vipul Harsh, Wenxuan Zhou, Sachin Ashok, Radhika Niranjana Mysore, Brighten Godfrey, and Sujata Banerjee. 2023. Murphy: Performance diagnosis of distributed cloud applications. In *Proceedings of the ACM SIGCOMM 2023 Conference*. 438–451.
- [4] Zefan Wang, Zichuan Liu, Yingying Zhang, Aoxiao Zhong, Jihong Wang, Fengbin Yin, Lunting Fan, Lingfei Wu, and Qingsong Wen. 2024. Rcaagent: Cloud root cause analysis by autonomous agents with tool-augmented large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 4966–4974.
- [5] Junjielong Xu, Qinan Zhang, Zhiqing Zhong, Shilin He, Chaoyun Zhang, Qingwei Lin, Dan Pei, Pinjia He, Dongmei Zhang, and Qi Zhang. 2025. OpenRCA: Can large language models locate the root cause of software failures?. In *The Thirteenth International Conference on Learning Representations*.