A Framework to Evaluate LLM Agents for Network Configuration

Yong Cui, Chang Liu, Xiaohui Xie, Chenguang Du

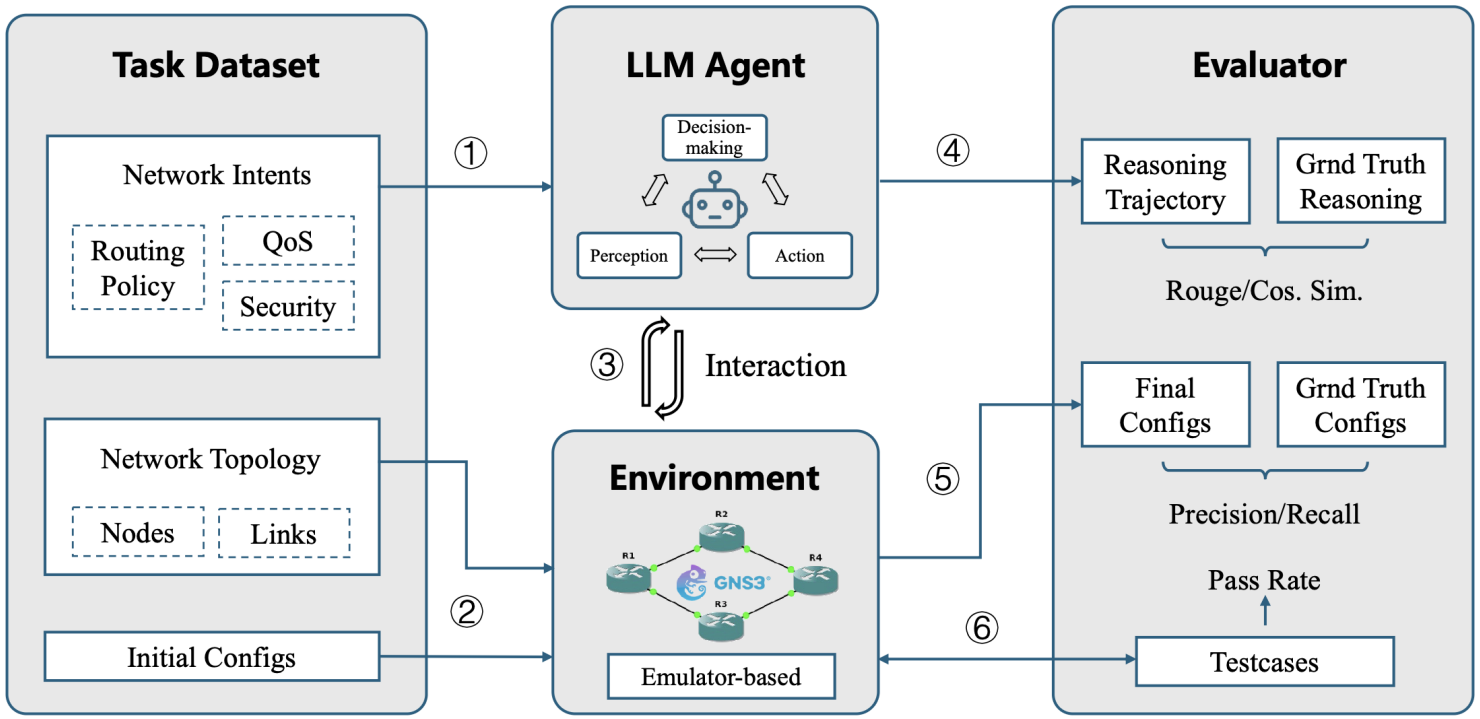
Tsinghua University, Beijing Zhongguancun Laboratory

July, 2025

Motivation

- Challenges
  - Existing evaluation methods are too simplistic to reflect the complexity and diversity of real network environments
  - Lack of standardized evaluation benchmarks for Agent network configuration capabilities
- Propose a evaluation framework for LLM network configuration agents
  - ✓ Define task datasets covering various network configuration tasks
  - ✓ Integrate simulation environment
  - ✓ Offer triple evaluation metrics: reasoning quality, command accuracy, functional correctness

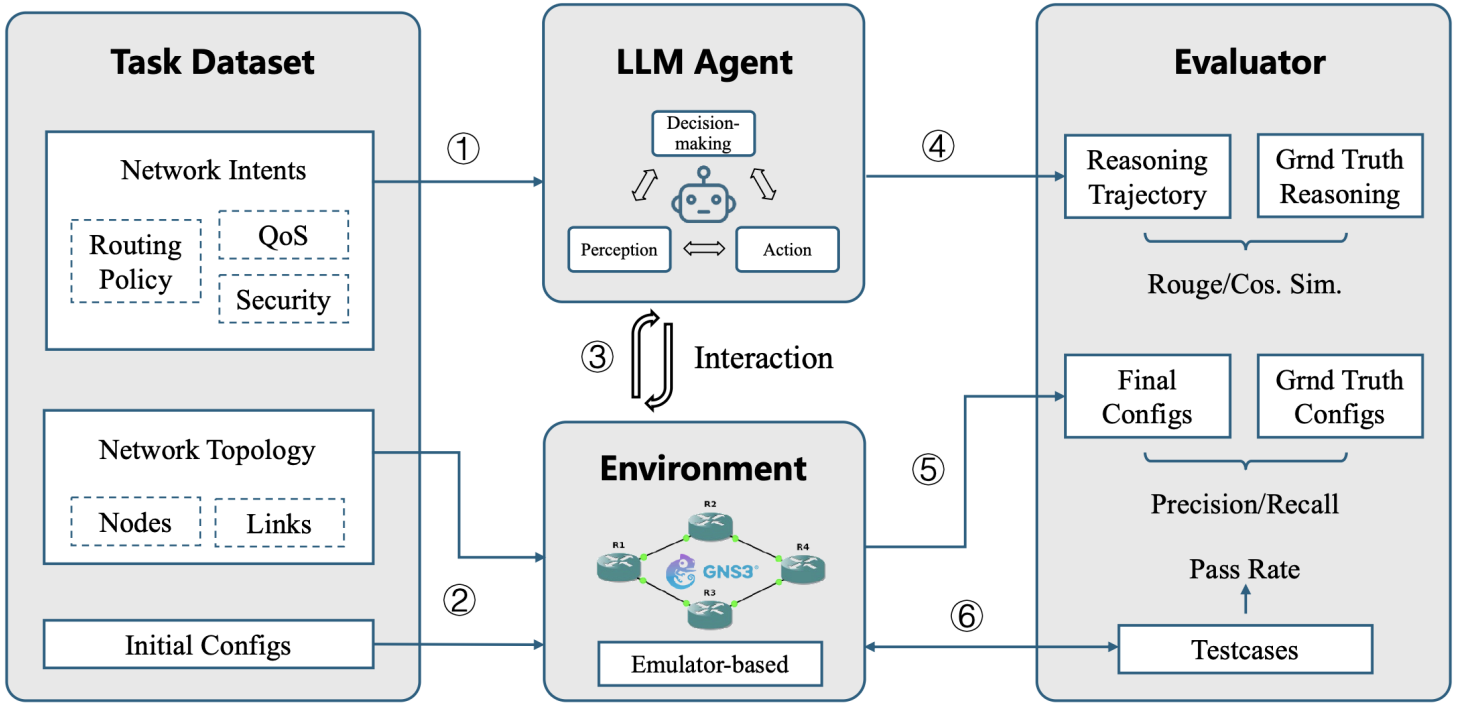
Framework Overview



Key Components:

- **Task Dataset：** Intent, Topology, Initial & GT Configs and Test Cases
- **Emulator Environment：** Built on network simulation software
- **LLM Agent：** The agent to be evaluated
- **Evaluator :** Reasoning, Command and Test Case Scores

Evaluation Workflow



① **Task Assignment:** Select task from Task Dataset
② **Environment Setup:** Create network topology, load initial configurations
③ **Interactive Execution:** Agent issues API calls to interact with the emulated environment
④ **Reasoning Export:** Capture reasoning logs
⑤ **Configuration Export:** Extract final running configurations from all devices
⑥ **Scoring:** Compute metrics

Future Work

- **Future Work**
  - Refine data exchange formats and fields between agents, simulation environment, and evaluator
  - Adapt MCP and A2A protocols to support more tools and multi-agent collaboration
- **Goal: An automated, fair, and comprehensive LLM Agent evaluation framework**

Thanks!
We welcome collaborators!

Q&A

Chenguang Du
ducg@zgclab.edu.cn