

GRAPH-BASED FEW-SHOT THREAT DETECTION USING RAG MODELS

Presented by:

Smith Ardeshan
Devarsh Patel
Dharmitkumar Patel
Lovely Priya
Mansi Tanna

Project Advisor:

Dr. Simon Shim

Introduction

Research Problem

- SOC analysts face high-volume network flow telemetry with limited labeled attacks
- Attack classes are highly imbalanced and rare
- Malicious behavior appears as **relationships between hosts**, not isolated rows
- Existing models lack **explainability and structured reasoning**

Research Gap

- Flat ML models fail to capture relational attack patterns
- LLM-only approaches lack structured classification reliability
- Missing integration of **graph learning + threat intelligence + reasoning**

Proposed Solution

Methodology

End to End Pipeline:
Flow → GNN → RAG → LLM → Adjudicator → Report

Key Idea

- GNN → structural behavior
- RAG → CVE + MITRE context
- LLM → explanation
- Adjudicator → fixes hard cases

DATASETS & KNOWLEDGE BASE

DATASETS & DISTRIBUTION

Combined Distribution

90%

Benign (90%)
Botnet / DoS (5.4%)
Other (2%)
Exploitation (1.7%)
Malware / C2 (0.9%)

RAG KNOWLEDGE BASE

1,151 Total Documents

691 MITRE ATT&CK Techniques

460 NVD CVEs

Analysis and Results

Key Results

Metric	Value
GNN F1 Score	0.989
Pipeline Accuracy	84%
JSON Reliability	100%
GNN vs XGBoost	+85%
Hard Case Accuracy	100% (Phase 7)

3.1 GNN Performance vs Baseline
GraphSAGE significantly outperforms XGBoost baseline across few-shot settings.

3.2 LLM + RAG Comparison

Model	Quality	JSON	Speed
LLaMA 4 Scout	87.5	100.0	0.73
Qwen 32B	84.8	100.0	4.12
Kimi K2	82.6	100.0	1.34
LLaMA 3.1	76.3	100.0	0.53
LLaMA 3.3	100.0	2.5	31.0

RAG Impact:
Without RAG → 17% accuracy
With RAG → 84% accuracy
(+67 percentage points improvement)

3.3 Pipeline Improvement Progression

Accuracy improves from 0% to 84% after rule-based refinements and RAG integration. Phase 7 specifically targets hard cases using adjudication.

3.4 Evaluation Results

Real dataset: 69.2%
10-flow challenge: 80.0%
50-flow challenge: 84.0%
Phase 7: 100% (hard cases)

Contribution & Conclusion

4.1 Phase 7 Adjudicator

Why Adjudicator?
Edge cases cause pipeline errors

Hard Cases Handled
Targeted correction using fine-tuned adjudicator

Impact
Higher accuracy & reliable threat verdicts

Hard Cases Fixed

- Slow C2 (HTTPS 443)
- DNS tunneling
- No-GNN C2 (IP mismatch)
- PPS artifacts (short flows)

Impact (Hard Case Accuracy)

Base Pipeline (Before): 60%
Phase 7 Adjudicator (After): 100%
+40 percentage points

5. CONFIGURATION

- Base: LLaMA 3.1 8B
- Method: LoRA (4-bit NF4)
- Trainable: 0.52%
- Data: 150 samples
- Training: 3 epochs

Evaluation & Conclusion

6. EVALUATION

- Real dataset: 69.2%
- 10-flow: 80%
- 50-flow: 84%
- Phase 7 (hard cases): 100%

7. CONCLUSIONS

- GNN achieves F1 = 0.989
- RAG improves 17% → 84%
- Adjudicator fixes edge cases (60% to 100%)
- Fully integrated explainable system

Limitations & Future Work

8. LIMITATIONS

- High latency (~89s/flow)
- Limited training data
- Not real-time yet

9. FUTURE WORK

- Optimize latency
- Expand dataset
- Real-time deployment
- SOC/SIEM integration

References

- Hamilton et al., GraphSAGE
- Lewis et al., Retrieval-Augmented Generation
- MITRE ATT&CK Framework
- Hu et al., LoRA

Interested in partnering with us on a capstone project or exploring other ways to collaborate with the College of IDS? Contact CIDS_CAP@sjsu.edu to discuss potential projects and partnership opportunities.