ICT9099CS02 - Similarity-Based Alert Clustering and Adaptive Checkpoints in Dynamic Path Expansion

Principal Investigator

Mr. Muhammad Tariq	
Polytechnic ID: 12010823	**Department/School**: ICT

Programme	ICT9099 Master of Science by Research in Computer Science
Research Area and Specialization:	AI and Cybersecurity (Threat Hunting and Intelligence).
Company Endorsement	Beyon Cyber
Industry mentor	Mr. Britto Irudhayasamy

Background of the problem

Recent advances in cybersecurity research highlight the growing adoption of multi-agent LLM systems for threat detection and incident response, offering improved contextual reasoning and cross-domain correlation. However, these systems introduce new challenges, including coordination risks, hallucinations, and lack of standardized evaluation frameworks.
Parallel research in alert clustering demonstrates the effectiveness of similarity-based approaches in reducing alert fatigue, yet fails to leverage similarity for investigation path reuse. Additionally, emerging work on LLM guardrails and human-in-the-loop frameworks emphasizes the importance of safety, trust calibration, and controlled autonomy in high-stakes domains. Despite these advancements, existing approaches remain fragmented, lacking an integrated framework that combines multi-agent orchestration, similarity-based reasoning, adaptive decision checkpoints, and knowledge preservation. This research addresses these gaps by proposing a guardrail-driven multi-agent system with adaptive path expansion and similarity-based investigation reuse.

List of publications related to the field

NA

Research Objectives and Expected Outcomes:

Aim & Objectives

Developing deterministic step execution agents to follow a standard path (for consistency and reliability) for resolving common alert profile.

Developing a decision-gated architecture to implement decision points-based evidence evaluation before expanding the investigation.

Developing a multi-level guardrails system, in which first layer is an LLM-based reviewer to check the completeness of the path and the second layer is deterministic layer to prevent agents from executing unauthorized queries or exceeding time-to-respond (TTR) limits.

Developing a human-in-the-loop authorization framework to ensure manual validation of critical investigation paths.

Expected Outcomes

A framework to guarantee a minimal viable investigation using fixed rules but also allow AI-based discovery and adaptive checkpoints.

Research Design/Methodology

Will be provided later.

Significance and potential impact of research in Bahrain

This research will bridge the gap in AI-assisted cybersecurity by developing a unified framework that will integrate multi-agent orchestration, similarity-based alert clustering, adaptive checkpoints, and multi-level guardrails. It will enable SoCs to preserve and reuse validated investigation knowledge, maintain consistent investigation quality, and reduce duplicated analytical effort. The findings will directly benefit SOC users, improve incident response efficiency, reduce analyst workload, and advance the safe deployment of agentic AI in cybersecurity environments.

Alignment with Bahrain’s national research strategy

This research aligns directly with Bahrain's National Cyber Security Strategy 2025–2028, particularly its pillars of Advanced Cyber Resiliency and Cyber Research, Development, and Innovation. It supports Bahrain's Economic Vision 2030 and the National Digital Economy Strategy by advancing responsible AI adoption, strengthening cybersecurity capabilities, and contributing to the Kingdom's ambition to become a regional leader in digital innovation. The research responds to EDB priorities in the ICT and cyber protection sectors, addresses the growing need to protect Critical National Infrastructure, and supports the National Cybersecurity Center's mandate for scientific research and national capability development. Furthermore, it

contributes to local knowledge production, SOC workforce development, and the safe deployment of agentic AI systems in alignment with Bahrain's national AI governance framework.