**Introduction**

Kairos is a system that analyzes temporal graph networks (TGNs) derived from whole-system provenance graphs. These graphs, built from audit data, represent system events over time, with nodes symbolizing entities like processes, files, and sockets, and edges representing interactions between these entities. Kairos leverages these graphs for anomaly detection and system behavior analysis.

By integrating methodologies like HIPE (Hierarchical Perturbation Explanation) and HELP (Hierarchical Embedding and Learning for Provenance), Kairos can enhance its capabilities for both interpretability and anomaly detection. These methods provide systematic approaches to understanding the critical components and interactions within the graph, ensuring meaningful insights at various levels of granularity.

**Integrating the HIPE Methodology with Kairos. (1)**

The HIPE (Hierarchical Perturbation Explanation) methodology, originally designed for image classification, can be adapted to the Kairos system to introduce hierarchical perturbation strategies for its Temporal Graph Networks (TGNs). Kairos processes whole-system provenance graphs derived from audit data, where nodes represent system entities, (for example processes, files, sockets) and edges represent interactions between entities, with timestamps adding a temporal aspect.

Proposed Approach: -

Hierarchical Perturbation of Provenance Graph: -

Edge Removal: Selectively remove edges to simulate the absence of specific events. This helps determine the importance of these events in anomaly detection or model performance.

Node Removal: Mark or remove nodes to simulate missing entities, highlighting their contribution to the overall system behavior.

Timestamp Alteration: Modify timestamps to evaluate the sensitivity of the model to event timing and reveal whether specific event sequences are critical.

Concept Grouping: Analyze the graph perturbations to group similar events into higher-level concepts such as file access or network communication.

Hierarchical Analysis of Perturbations: -

Systematically perturb different levels of the graph hierarchy (e.g., individual nodes, clusters, subgraphs) to measure their impact on the detection of anomalies or patterns.

Use insights to identify the most critical components and interactions in the system.

### **Integrating the HELP Methodology with Kairos (2)**

The HELP (Hierarchical Embedding and Learning for Provenance) methodology involves dividing the streaming provenance graph analyzed by Kairos into discrete time windows and extracting higher-level concepts from node embeddings and clustering.

Proposed Approach

Temporal Segmentation

Segment the streaming provenance graph into 15-minute time windows (or another suitable duration) as Kairos already supports this windowing mechanism.

Node Embedding with TGN Layers: -

Apply TGN layers to transform node features into embeddings that capture both temporal and structural relationships within each time window.

Node Clustering: -

Use a clustering method, such as DMoN (Deep Modularity Network), to group nodes based on behavioral and structural similarities.

Each cluster may represent entities with similar roles or actions, e.g., a cluster of web server processes.

Concept Node Creation: -

Merge connected components within each cluster to form higher-level concept nodes. For example:

A cluster of processes representing web server activity could be aggregated into a web server instance concept node.

Concept nodes encapsulate the behavior of multiple nodes and their interactions.

Reconstruction Error Evaluation: -

Calculate the reconstruction error of edges associated with concept nodes across time windows.

Use these errors to assess the quality of abstraction and determine whether the higher-level concepts preserve critical system behaviors.

**References: -**

[1] J. Cooper, O. Arandjelovic, and D. J. Harrison, "Believe the HiPe: Hierarchical perturbation for fast, robust, and model-agnostic saliency mapping," Pattern Recognition, vol. 131, p. 108914, Apr. 2022.

[2] Everybody needs a little help: Explaining graphs via hierarchical concepts," unpublished manuscript, under double-blind review, 2024.