LLM-Powered Description Synthesis for Sensitive Behaviors in Software Applications

PI: Dr. Xusheng Xiao, Associate Professor, xusheng.xiao@asu.edu
School of Computing and Augmented Intelligence, Arizona State University

Overview of Proposed Research

This online version of the proposal shows the overview of our proposed research, as indicated in Figure 1. Our proposed research is organized into three phases: Multi-source Data Acquisition and Preparation (Phase I), Privacy-Relevant Node Filter and Summarization (Phase II), and Permission Description Synthesis (Phase III).

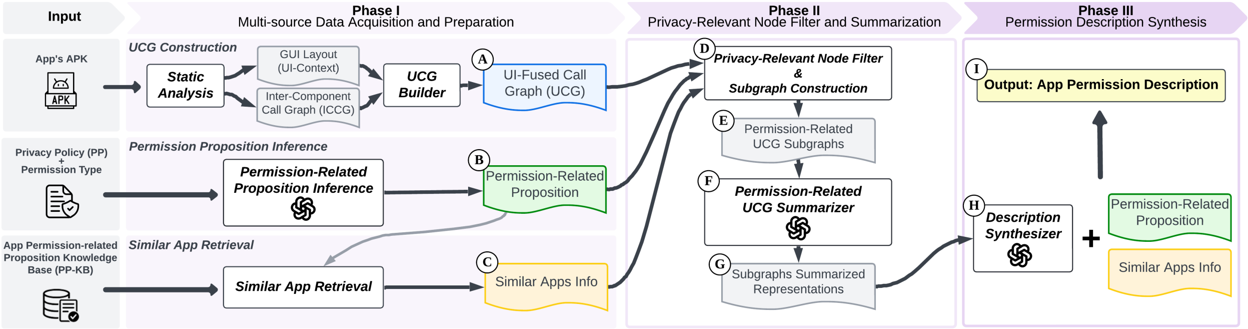
Input
Phase I
Multi-source Data Acquisition and Preparation
Phase II
Privacy-Relevant Node Filter and Summarization
Phase III
Permission Description Synthesis
App's APK
APK
UCG Construction
Static Analysis
GUI Layout (UI-Context)
Inter-Component Call Graph (ICCG)
UCG Builder
A
UI-Fused Call Graph (UCG)
Privacy Policy (PP) + Permission Type
Permission Proposition Inference
Permission-Related Proposition Inference
B
Permission-Related Proposition
App Permission-related Proposition Knowledge Base (PP-KB)
Similar App Retrieval
Similar App Retrieval
C
Similar Apps Info
D
Privacy-Relevant Node Filter & Subgraph Construction
E
Permission-Related UCG Subgraphs
F
Permission-Related UCG Summarizer
G
Subgraphs Summarized Representations
I
Output: App Permission Description
H
Description Synthesizer
+
Permission-Related Proposition
Similar Apps Info

Figure 1: ***Overview of the proposed research***

Phase I will prepare three heterogeneous types of information. First, we will perform static analysis on the input APK to extract GUI layouts containing UI context information and the ICCG, then integrates these elements by mapping UI contextual resources from the GUI layouts to the ICCG nodes to create a "Ⓐ UCG". Second, we will processe the app's privacy policy and a sensitive permission from the app's metadata, employing an LLM agent to generate "Ⓑ Permission-Related Propositions". Third, we will query the Permission-related Proposition Knowledge Base (PP-KB) to find similar apps from the PP-KB based on proposition similarity and retrieves their associated information, forming "Ⓒ Similar App Info".

Phase II will take the output of Phase I (Ⓐ, Ⓑ, Ⓒ) as input, and generates a summary of the relevant information ("Ⓖ Subgraphs Summarized Representations"). Specifically, we will implement a "Ⓓ Privacy-Relevant Node Filter" that leverages multiple predefined critical node patterns to extract nodes containing critical privacy-related information from the UCG. These extracted nodes are then reconstructed into three "Ⓔ Permission-Related UCG Subgraphs". We will then employ an LLM agent ("Ⓕ Permission-Related UCG Summarizer") to summarize each node within these subgraphs to generate the summaries.

Phase III will implement a "Ⓗ Description Synthesizer" that employs LLM agents to process the summaries of the three subgraphs, and performs cross-validation and integration of the heterogeneous analysis results to synthesize the "Ⓘ App Permission Description".

Detailed Example of Proposed Research

Figure 2 provides a step-by-step illustration of LLM4APT 's workflow using a simplified real-world app (`com.gec.MarineApp`), including examples of the input sources, intermediate steps, and the generated description.

In Phase I, Ⓐ depicts a node from the constructed UCG, showing the method signature, a function body that defines three alert dialog types (*poor GPS*, *low battery*, and *outside allowed distance*), and a UI context featuring an *anchor alarm* capability. Ⓑ depicts privacy policy segments related to location data, alongside

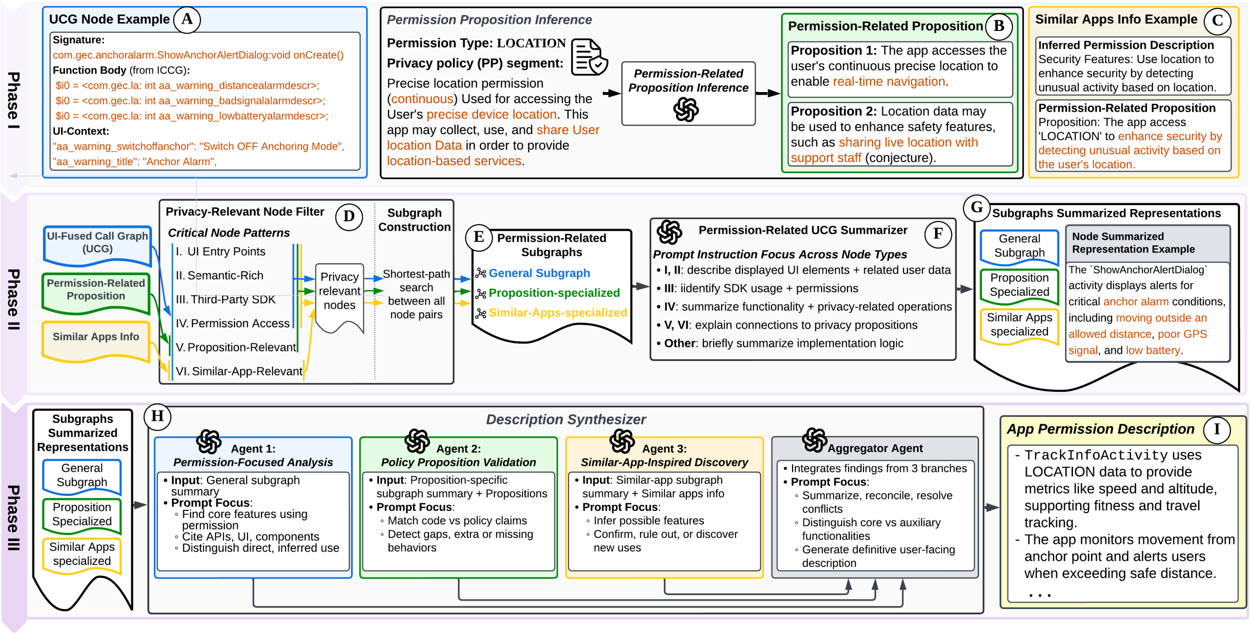
Phase I
UCG Node Example
A
Signature:
com.gec.anchoralarm.ShowAnchorAlertDialog:void onCreate()
Function Body (from ICCG):
$i0 = <com.gec.la: int aa_warning_distancealarmdescr>;
$i0 = <com.gec.la: int aa_warning_badsignalalarmdescr>;
$i0 = <com.gec.la: int aa_warning_lowbatteryalarmdescr>;
UI-Context:
"aa_warning_switchoffanchor": "Switch OFF Anchoring Mode",
"aa_warning_title": "Anchor Alarm",
Permission Proposition Inference
Permission Type: LOCATION
Privacy policy (PP) segment:
Precise location permission (continuous) Used for accessing the User's precise device location. This app may collect, use, and share User location Data in order to provide location-based services.
Permission-Related Proposition Inference
Permission-Related Proposition
B
Proposition 1: The app accesses the user's continuous precise location to enable real-time navigation.
Proposition 2: Location data may be used to enhance safety features, such as sharing live location with support staff (conjecture).
Similar Apps Info Example
C
Inferred Permission Description
Security Features: Use location to enhance security by detecting unusual activity based on location.
Permission-Related Proposition
Proposition: The app access 'LOCATION' to enhance security by detecting unusual activity based on the user's location.
Phase II
UI-Fused Call Graph (UCG)
Permission-Related Proposition
Similar Apps Info
Privacy-Relevant Node Filter
D
Critical Node Patterns
I. UI Entry Points
II. Semantic-Rich
III. Third-Party SDK
IV. Permission Access
V. Proposition-Relevant
VI. Similar-App-Relevant
Privacy relevant nodes
Subgraph Construction
Shortest-path search between all node pairs
E
Permission-Related Subgraphs
General Subgraph
Proposition-specialized
Similar-Apps-specialized
Permission-Related UCG Summarizer
F
Prompt Instruction Focus Across Node Types
• I, II: describe displayed UI elements + related user data
• III: iidentify SDK usage + permissions
• IV: summarize functionality + privacy-related operations
• V, VI: explain connections to privacy propositions
• Other: briefly summarize implementation logic
G
Subgraphs Summarized Representations
General Subgraph
Proposition Specialized
Similar Apps specialized
Node Summarized Representation Example
The `ShowAnchorAlertDialog` activity displays alerts for critical anchor alarm conditions, including moving outside an allowed distance, poor GPS signal, and low battery.
Phase III
Subgraphs Summarized Representations
General Subgraph
Proposition Specialized
Similar Apps specialized
H
Description Synthesizer
Agent 1:
Permission-Focused Analysis
• Input: General subgraph summary
• Prompt Focus:
◦ Find core features using permission
◦ Cite APIs, UI, components
◦ Distinguish direct, inferred use
Agent 2:
Policy Proposition Validation
• Input: Proposition-specific subgraph summary + Propositions
• Prompt Focus:
◦ Match code vs policy claims
◦ Detect gaps, extra or missing behaviors
Agent 3:
Similar-App-Inspired Discovery
• Input: Similar-app subgraph summary + Similar apps info
• Prompt Focus:
◦ Infer possible features
◦ Confirm, rule out, or discover new uses
Aggregator Agent
• Integrates findings from 3 branches
• Prompt Focus:
◦ Summarize, reconcile, resolve conflicts
◦ Distinguish core vs auxiliary functionalities
◦ Generate definitive user-facing description
App Permission Description
I
- TrackInfoActivity uses LOCATION data to provide metrics like speed and altitude, supporting fitness and travel tracking.
- The app monitors movement from anchor point and alerts users when exceeding safe distance.
. . .

Figure 2: ***Detailed example of proposed research***

two inferred permission-related propositions: *Proposition 1* specifies continuous location tracking for real-time navigation, and *Proposition 2* identifies location sharing with support staff for safety purposes. Ⓒ presents retrieved information of a similar app from the PP-KB, including a relevant proposition and a description emphasizing security and safety purposes for location data usage.

In Phase II, Ⓓ shows the Privacy-Relevant Node Filter, which identifies relevant nodes using six patterns. Ⓔ illustrates the construction of three subgraphs from these nodes: a general subgraph, a proposition-specialized subgraph, and a similar-apps-specialized subgraph. Ⓕ shows the Permission-Related UCG Summarizer, where LLM agents use specialized prompts tailored to different node types to generate summarized representations for each subgraph. Ⓖ shows an example of such a summarized representation for the `onCreate()` method of the class `ShowAnchorAlertDialog`.

In Phase III, Ⓗ shows the Description Synthesizer. This module includes three LLM agents: one extracts permission-related app features (Permission-Focused Analysis), another matches code implementation with policy claims (Policy Proposition Validation), and the third infers related features from similar apps (Similar-App-Inspired Discovery). An aggregator agent then integrates the findings from all three agents.

The final output, Ⓘ, shows the generated descriptions for two permission-related features. The *anchor mark and distance alert* functionality is synthesized from the function code and UI evidence in Ⓐ, further informed by the safety-focused usage patterns from the similar app in Ⓒ. The *location tracking* feature is derived from a separate `TrackInfoActivity` node (not shown in this example node Ⓐ), whose UI context includes speed and altitude indicators, consistent with Proposition 1 from Ⓑ. Proposition 2 from Ⓑ (location sharing with support staff) is excluded from the final description due to lack of supporting evidence in the UCG analysis.