TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 02/23 pp1977–1991
DOI: 10.26599/TST.2024.9010220
Volume 31, Number 4, August 2026

Research Article

Bridging the Compliance Gap: Effective and Efficient Detection of Non-Compliant Behaviors in Android Applications

Runqi Fan, Fan Wu, Zifeng Kang, Peng Hu, Weiting Chen, and Song Li*

Abstract: As mobile applications become increasingly complex and privacy regulations continue to evolve, the task of accurately identifying app violations in compliance detection has become a major challenge. Prior works mainly relied on taint analysis and dynamic monitoring to address this issue. However, taint analysis requires specifying sources and sinks for each violation, leading to multiple analysis rounds and inefficiency. Meanwhile, dynamic monitoring suffers from incomplete coverage, resulting in high false negatives. In this paper, we propose a novel graph structure, called Behavior Property Graph (BPG), for detecting non-compliant behaviors in Android applications. BPG integrates the features from various graph representations, including Abstract Syntax Tree (AST), Control Flow Graph (CFG), Call Graph (CG), Program Dependency Graph (PDG), and Pointer Assignment Graph (PAG), enabling comprehensive modeling of complex app behaviors. Violations are identified by querying the BPG using behavioral patterns extracted from real-world apps. We develop a prototype system called BPGen to generate BPGs and evaluated its performance by testing seven types of non-compliant behaviors on a dataset of 200 real-world apps. Notably, BPGen detects 14 violations within 13 previously unreported non-compliant applications. The results show that BPGen can efficiently and effectively detect app compliance violations.

Key words: compliance detection; Android applications; behavior analysis; static analysis, graph query

1 Introduction

With the growing prevalence of mobile applications in daily life and their significant impact on users' privacy and consent rights, ensuring compliance of these apps with regulations has become increasingly essential.

Governments worldwide have introduced numerous compliance regulations to meet these demands. Popular regulations such as the General Data Protection Regulation (GDPR)[1] and the California Consumer Privacy Act (CCPA)[2] have attracted considerable interest from researchers. In China, regulations such as the Data Security Law[3] and the Personal Information Protection Law[4] have established strict standards for how apps manage data and ensure transparency. Additionally, in 2024, the Ministry of Industry and Information Technology (MIIT) increased its focus on issues such as unauthorized launches and unclosable apps[5], further broadening the scope of compliance requirements. Despite the rapid growth of these regulations, research on compliance policies and app behavior in China remains relatively limited.

• Runqi Fan, Fan Wu, and Song Li are with the State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou 310027, China. E-mail: fanrunqi@zju.edu.cn; fanwu01@zju.edu.cn; songl@zju.edu.cn.
• Zifeng Kang is with Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: zkang7@jhu.edu.
• Peng Hu is with Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, Hangzhou 310053, China. E-mail: penghu01@outlook.com.
• Weiting Chen is with Ant Group Co., Ltd., Hangzhou 310023, China. E-mail: weiting.cwt@antgroup.com.
* To whom correspondence should be addressed.
Manuscript received: 2024-08-01; revised: 2024-10-10; accepted: 2024-10-27

Furthermore, the expanding scope of violations and compliance requirements introduces significant challenges to app compliance oversight, particularly in behavior analysis. On the one hand, the growing variety and complexity of violations underscore the need for precise modeling of each violation type, requiring an accurate representation of execution logic and data handling processes. On the other hand, the continuous emergence of new violations and evolving regulatory requirements necessitate greater adaptability and scalability in behavior analysis techniques.

Prior research in Android app behavior analysis primarily employed taint analysis[6–10] and dynamic monitoring[11, 12]. However, these approaches are limited in the current regulatory landscape. Taint analysis necessitates the specification of sources (e.g., user information) and sinks (e.g., external transmissions), requiring multiple rounds of analysis for various types of violations. This not only results in inefficiencies but also restricts the ability to detect non-data leakage behaviors, such as unauthorized cross-app launches, and other emerging violations. Moreover, dynamic analysis relies on real-time execution monitoring—such as network traffic monitoring and system API instrumentation—but is often constrained by its inability to capture all possible violations in large and complex applications, leading to false negatives. Overall, these limitations highlight the urgent need for more comprehensive and efficient methods in app behavior analysis to address the evolving landscape of regulatory compliance.

To overcome these limitations, we introduce a novel graph structure for app compliance detection called Behavior Property Graph (BPG). BPG incorporates various components from multiple graph representations, including the Abstract Syntax Tree (AST), Control Flow Graph (CFG), Call Graph (CG), Program Dependency Graph (PDG), and Pointer Assignment Graph (PAG), to effectively model the full range of app behaviors. This integrated representation demonstrates compatibility in detecting a wide range of violations, extending beyond privacy leakage issues. Once BPG for a specified app is generated, it will be stored in the graph database. By performing graph traversals on BPG, we can check for the existence of violations. In addition, when new compliance requirements arise, they can be accommodated by simply adding query statements, eliminating the need to rebuild BPG. This approach provides a more scalable and flexible solution for compliance detection.

We implement a prototype system called BPGen (BPG Generator) to generate BPGs for Android applications and evaluate the system on a dataset of 200 Android applications, focusing on seven types of non-compliant behaviors. As shown by the experimental results, BPGen completes the analysis for over 90% of the apps within ten minutes. Compared to existing approaches, BPGen can detect a broader range of violations with a lower False Positive Rate (FPR) and False Negative Rate (FNR), demonstrating high efficiency and effectiveness.

In summary, we make the following contributions in this paper:

- We design the BPG for effective and efficient app behavior analysis, which also exhibits high scalability for detecting compliance violations.
- We analyze real-world applications to extract behavioral patterns, guiding the creation of graph queries for detecting non-compliant behaviors.
- We develop a prototype system, BPGen, which identifies zero-day compliance violations in 14 real-world applications during evaluation.

2 Related Work

Behavioral analysis of mobile applications. Static and dynamic analysis methods have garnered extensive research attention in the realms of program optimization, security vulnerabilities, and compliance detection[6–18]. To analyze the behavior of mobile applications, particularly to identify data leakage, numerous static and dynamic analysis methods have been developed. Static analysis methods mostly employ techniques such as taint analysis and property graph[6–9, 13]. Dynamic analysis methods execute the app in a controlled environment to observe its behavior in real-time. These techniques can include dynamic taint tracking, sandboxing, and system call monitoring[10–12].

FlowDroid[6] and VulHunter[13] have also been used in some compliance detection research[19–21], but they have certain limitations. While FlowDroid performs a flow-sensitive, context-sensitive analysis of data flows within an app to detect potential leaks of sensitive information, it does not support checks for the existence and frequency of specific behaviors. For instance, behaviors like Loop Popup After Permission

Denial are not data-related, and FlowDroid cannot detect them. Moreover, FlowDroid requires predefined sources and sinks for taint analysis. Since the call paths for different violations vary greatly, a unified set of sources and sinks cannot be defined. Each time a violation needs to be checked, sources and sinks must be redefined, and the app must be reanalyzed, leading to significant overhead. BPGEN, on the other hand, can analyze an app once and perform multiple queries, reducing time and resource consumption. While VulHunter adopts the concept of property graphs, it does not model resource files and primarily focuses on vulnerability detection, lacking adaptation for compliance violations detection.

Compliance detection of specific behaviors. Recent works also focused on the detection of specific non-compliant behaviors. MOWCHECKER[22] detects inconsistencies in mobile apps' third-party data collection upon user withdrawal. Nguyen et al.[23] proposed a method to detect apps that send out users' personal data without prior consent. OptOutCheck[24] detects inconsistencies between the actual data practices of online trackers and their policy statements regarding user opt-out choices. These studies are based on detecting specific mechanisms of certain behaviors and lack general applicability, making it difficult to conduct broad-scale violation detection.

Privacy policy consistency analysis. In compliance detection research, prior works primarily focused on analyzing the consistency between application behaviors and their privacy policies to determine whether actual behaviors align with the claims made[19, 25–35]. Researchers commonly employ Natural Language Processing (NLP) techniques to model privacy policies, specifying the entities authorized to access user data, the types of data collected, and the purposes of data usage. Additionally, they utilize program analysis methods such as taint analysis, network traffic monitoring, and tracking of sensitive API calls to identify where and how data is accessed within the app. However, these studies primarily focus on whether personal data is misused by applications, limiting their applicability to the detection of other types of violations.

3 Overview

In this section, we first present a motivating example and then delineate the threat model pertinent to app compliance detection.

3.1 A motivating example

We demonstrate the use of BPG in detecting non-compliant behaviors through an example that checks whether an app provides a privacy policy to users in a regulated manner.

The regulations mandate that there must be a hyperlink or button for the privacy policy prominently displayed within the popup, login/registration page, or settings section[36]. Upon clicking this hyperlink or button, the full text of the privacy policy must be readily accessible and displayed to the user. Therefore, developers are required to present the privacy policy text to users in its entirety, adhering to the regulatory requirements.

Figure 1 illustrates a simplified exemplary code snippet from bubei.tingshu, a real-world Android application providing audio digital listening service, to showcase one of the primary approaches adopted by apps to provide privacy policies.

When the user interacts with the app, the predefined privacy policy text from the string resource file is seamlessly integrated into a layout file and displayed to the user via a TextView. To allow the user to access the full policy, a click listener is attached to this TextView in the activity. Upon clicking, the listener executes a predefined action, such as navigating the user to a new screen within the app that displays the full policy or opening a web page in the user's default browser where the policy is hosted, thereby providing the user with easy access to the privacy policy information. The absence of a readily accessible privacy policy text to users signifies a violation of compliance requirements.

The illustrative case motivates the design of BPG in two fundamental ways. Firstly, the extensive manipulation of resource files drives the need to model these files and operations as nodes and edges within BPG. Secondly, the use of resource IDs in findViewById within MainActivity, coupled with user click monitoring via setOnClickListener, underscores the requirement for BPG to model invocation relationships and data dependencies, thereby enabling a comprehensive depiction of various behaviors.

Figure 1 illustrates how the behavior of having a clickable privacy policy within an app is modeled in BPG. The developer locates the resource ID corresponding to the privacy policy using findViewById and binds it to a click event listener setOnClickListener.

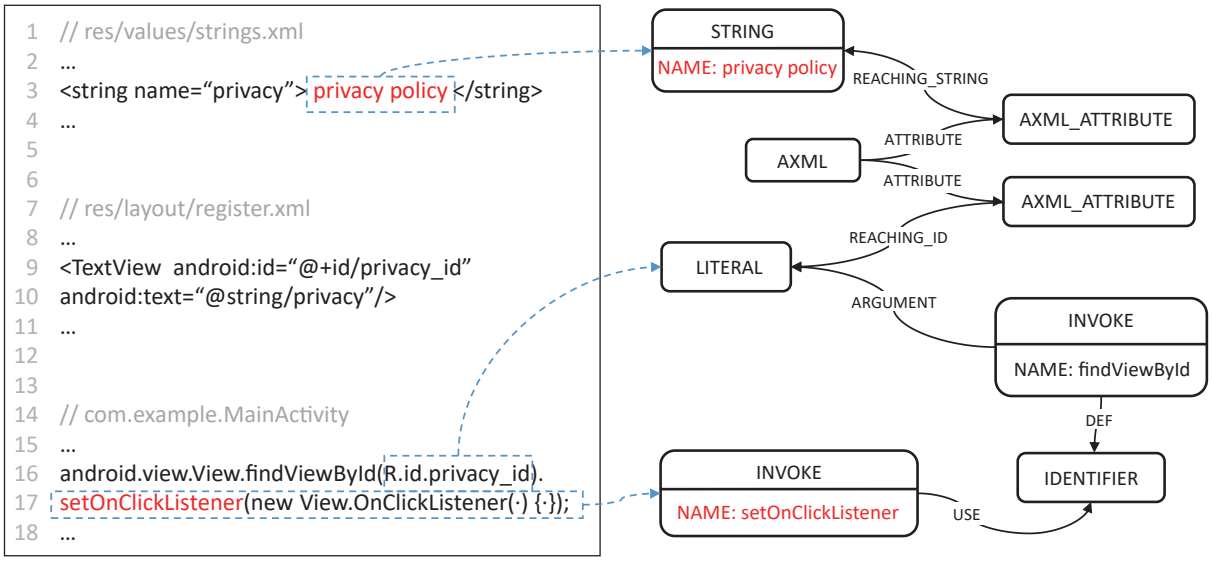


Fig. 1 Example code and the corresponding partial BPG.

In the following sections, we provide a more detailed introduction to the structure of BPG.

3.2 Threat model

The threat model of our study refers to misconduct introduced by app developers, whether deliberate or inadvertent, which leads to violations of regulatory requirements.

We categorize non-compliant behaviors into three types as follows:

- **Existence of behaviors** indicates that the app fails to provide or avoid certain behaviors as required by regulations. Examples of violations in this category include No Privacy Policy, No Account Deletion, Unauthorized Cross-App Launch, App Exits Upon Permission Denial, and Unclosable Targeted Push. These behaviors either violate transparency requirements or undermine user choices, which can be perceived as coercive and in violation of informed consent principles.
- **Content of behaviors** refers to the specific data or information accessed or transmitted during their execution, which may violate regulatory requirements. For instance, violations like Collecting Non-Modifiable Unique Device Identifiers fall under this category, as they often exceed the scope necessary for providing services and violate the principle of data minimization.
- **Frequency of behaviors** indicates that the occurrence of behaviors exceeds the required threshold. A notable example is Looping Permission Pop-ups After Denial, which demonstrates this frequency issue as it repeatedly requests permissions even after users have declined. This behavior disrupts the normal user experience and potentially violates consent norms.

In Section 5.2, we discuss how we utilize graph traversal on BPG to identify these three types of non-compliant behaviors. Subsequently, in Section 6, we analyze the occurrence of the aforementioned violations in real-world apps.

4 BPG

In this section, we describe the definition of BPG and the procedure of constructing BPG.

4.1 Definition

We define BPG as a graph representation, which integrates features from AST, CFG, CG, PDG, and PAG to a joint graph using the concept of property graphs. The app's BPG is constructed based on its resource and code files, and it is stored in a graph database using graph structures comprising nodes, edges, and properties. Next, we introduce these components within BPG.

4.1.1 Nodes

App behavior is shaped by execution logic, data handling, and interactions with resources. To model these aspects effectively, BPG incorporates two types of nodes: resource nodes and code nodes.

Resource nodes represent the configurations and various resources of an app, which can be extracted from files such as AndroidManifest.xml, strings.xml, and register.xml. For instance, when parsing

strings.xml, BPGEN traverses each string resource tag and creates corresponding nodes, with properties that store the resource identifier and value.

Structuring resource nodes enables a comprehensive analysis of how various resources interact with app functionalities. When detecting app violations, examining the resources linked to specific behaviors is often essential. For example, to determine whether an app provides a privacy policy to users, one critical step involves analyzing the string resources to check for the presence of specific privacy policy-related strings.

Code nodes, extracted from.dex files, represent the essential building blocks of an app's logic and structure, encompassing classes, methods, statements, fields, literals, and identifiers. These nodes facilitate a comprehensive analysis of how various code elements influence app behavior. For example, fields are essential for evaluating data manipulation within an app, as they may store sensitive user information. If an app improperly exposes these fields—perhaps through unprotected APIs—it could result in a data breach.

4.1.2 Edges

BPG utilizes the relationships between nodes as edges, which are extracted from various graph representations and then integrated to form a complete graph. Specifically, BPG includes syntax relations from the AST, which capture structural connections between nodes, and calling relations from the CG represent the connections between invocation statements and their corresponding method nodes. Additionally, it outlines control flow from the CFG and data dependencies from the PDG, representing execution paths and def-use chains. Finally, point-to relations from the PAG capture how pointers reference objects in memory, aiding in tracking inter-procedural data flow.

Syntax relations are extracted from AST, which can describe structural relationships between elements within the code of an application. These relations are fundamental in understanding how different syntactic components such as classes, methods, variables, and statements are interconnected.

In BPG, the edges used to describe syntax relations are categorized into three types: CONTAIN edges, ARGUMENT edges, and PARAMETER edges.

CONTAIN edges represent the hierarchical structure, showing how elements like classes, methods, and statements are nested within each other. Firstly, its importance lies in enabling a clear understanding of the scope and context in which operations are performed. For example, in an Android app, permission management is typically implemented within a specific class that includes methods for checking permissions, requesting permissions from the user, and handling the user's response. When analyzing violations like improper permission handling, CONTAIN edges help trace the permission request back to the class and method it belongs to, allowing verification of whether related methods, such as those handling the permission result, follow proper procedures. If not, issues like unauthorized app exit after permission denial or repeated permission requests may arise. Secondly, CONTAIN edges streamline the graph traversal process. When searching for an invoke statement within a method, CONTAIN edges allow traversal of the method's internal structure directly, without needing to check each statement one by one.

ARGUMENT edges illustrate the relationship between a method and its actual arguments, capturing the contextual information of a method invocation. In contrast, PARAMETER edges focus on understanding data flow within a method by linking it to its formal parameters. These edges are fundamental to inter-procedural data flow analysis, enabling precise tracking of data flow across various methods, which will be further elaborated in subsequent sections.

Calling relations, extracted from CG, depict how methods within an application invoke one another. These relations are crucial for understanding inter-procedural information flow and dependencies.

In BPG, calling relations are represented by CALL edges, which connect call statement nodes to the corresponding method nodes. In establishing these calling relations, we primarily address two challenges: asynchronous calls and native calls.

The first challenge is handling asynchronous calls. Android applications often use asynchronous programming techniques, such as listeners and callbacks, to manage events and background tasks. These asynchronous mechanisms can disrupt the direct invocation and execution flow of methods, creating challenges for precise modeling in calling relations.

BPG addresses this challenge through the following approaches.

When app starts a new activity, an intent object is created to specify the current activity and the target activity. This intent is passed as a parameter to the startActivity method. When startActivity is invoked, it automatically triggers the onCreate method of the target

activity. BPG captures this potential relations through graph traversal using chain operations, rather than directly modeling callbacks between methods. Specifically, BPG retains the name of parameters. When an intent is instantiated, it preserves the name of the target activity, and this intent is passed as a parameter to the startActivity method. The ARGUMENT edges and USE edges (to be discussed later) in the graph connect these elements. Additionally, each activity CLASS node retains its NAME property, allowing for matching the callback during the graph traversal phase.

Additionally, developers frequently set click listeners using anonymous inner classes, such as setOnClickListener, which registers a listener for click events on User Interface (UI) components. For these cases, BPG addresses this challenge by accounting for the implicit calls to init (i.e., instance initializer) and clinit (i.e., class constructors). When an anonymous class is instantiated, the init method initializes the listener, and the clinit method handles class-level initialization. Since anonymous classes are instantiated as parameters in the setOnClickListener method, BPG tracks these initialization methods to link the listener setup (i.e., setOnClickListener) with the corresponding callback method (i.e., onClick) using ARGUMENT edges, CONTAIN edges, and CALL edges. Consequently, the asynchronous callbacks are effectively modeled within the graph.

By employing these techniques, BPG enables precise modeling of asynchronous calls within apps.

The second challenge is managing native calls. Developers leverage both Java and Android libraries to invoke their APIs, which are native calls written in languages such as C/C++. In Java-based analysis, these native calls function as black boxes, where corresponding method details are not available, leading to a break in data flow analysis. Therefore, it becomes essential to model these native calls to understand how information propagates through them.

We categorize native methods into the following three types based on how data is propagated through these functions:

- Methods that directly return a value via return statements. For example, android.os.Build.getSerial is used to retrieve the device's serial number in Android.
- Methods that assign values from other parameters to the first parameter without returning anything (e.g., java.lang.StringBuilder.append, used to append the string representation of some argument to the sequence).
- Methods that return the invoking object (e.g., java.lang.String.toString returns the String representation of the object).

To address this challenge, we define a shadow class that mirrors the expected behavior of native methods. Based on data flow patterns, the types of parameters and return values are determined and shadow methods are defined accordingly. Each type of native call is then redirected to its corresponding shadow method. This approach allows us to simulate the behavior of native calls and maintain data flow analysis continuity, ensuring that the flow of information through these native calls can be effectively modeled and analyzed.

Control flow edges represent the flow of execution within a program. In BPG, control flow is modeled with specific edges to capture both intra- and inter-procedural flows.

For intra-procedural control flow, BPG establishes CF (control flow) edges to connect individual statements within a method, preserving their sequential execution. This structure allows for flow-sensitive analysis, where the order of operations is critical for accurately interpreting program behavior. For instance, the app is expected to collect the user's explicit consent before sending any sensitive data. In this case, the order of operations matters: the app must first ask for consent, then collect the data, and only after receiving consent, send it over the network. Flow-sensitive analysis ensures that the correct sequence is followed.

Inter-procedural control flow edges represent transitions between different methods and are vital for elucidating how control transitions between methods, allowing for a comprehensive analysis of program behavior. To enable inter-procedural control flow analysis in BPG, edges are added between the invoking statements and the invoked methods, capturing the control flow initiated by method calls. Additionally, return edges are included to signify the return points in the invoked methods, allowing the analysis to track how control flows back to the calling method after the invoked method has completed execution. This comprehensive representation facilitates a deeper understanding of program behavior, enabling accurate detection of compliance violations and enhancing data flow analysis.

BPG incorporates two types of edges to model data transfer within apps: data dependency edges and point-

to edges.

Data dependency edges, drawing inspiration from the PDG, employ a def-use chain to map the lifecycle of variables and fields within a method. On the other hand, **point-to edges**, originating from the PAG, delineate how pointers—encompassing variables and fields—reference objects in memory. Such insights are essential for mapping inter-procedural data dependencies, offering a clear view of the interactions between different methods and their shared data.

In BPG, we define DEF and USE edges to model intra-procedural data flow, while point-to edges enable the tracking of inter-procedural data transfer. By integrating these relations, BPG provides a comprehensive view of data interactions, making it easier to identify potential compliance violations. For instance, if an app retrieves sensitive information, such as the International Mobile Equipment Identity (IMEI) number, by invoking a system API, it typically stores this information in a variable or field. This sensitive data may then be passed between various methods. BPG's data dependency and point-to relations can effectively track this flow. If the app subsequently transmits this data without obtaining proper user consent, BPG's analysis would highlight this as a compliance violation.

4.2 BPG construction

BPGen first constructs an app's BPG according to its resource files and classes.dex files, and then stores it in a graph database, which uses graph structures (including nodes, edges, and properties) to represent and store data.

Specifically, BPGen first utilizes the libraries provided by FlowDroid[6] to parse all resource files, including AndroidManifest.xml, and establishes corresponding resource nodes. Then, it uses Soot[37] to disassemble the classes.dex into Jimple IR code. The construction of a BPG for code begins by traversing all classes, fields, methods, and statements and creating the corresponding code nodes. Based on the structure of the AST, BPGen identifies the containment relationships among these elements and adds CONTAIN edges between the corresponding nodes. Similarly, BPGen creates edges based on the program's CG, CFG, and PAG. When encountering call statements, variable definition and usage statements, and conditional statements, BPGen creates new nodes for the variables and establishes corresponding edges (such as ARGUMENT, DEF, and USE edges) between the variable nodes and the relevant statements to represent syntactic relations or the propagation of data values.

The constructed BPG is then stored in a Neo4j[38] graph database, which effectively handles the graph structures and facilitates efficient querying and analysis of the app's behavior and potential security issues.

5 BPG Query

In this section, we describe graph queries to BPG for non-compliant behaviors. We first present how to model queries as several types of graph traversals in Section 4.1 and then describe how to represent non-compliant behaviors via those graph traversals in Section 4.2.

5.1 Graph traversals

BPGen models non-compliant behaviors as graph traversals and perform them over BPGs to identify violating apps.

A graph traversal is a function $T: P(V) \to P(V)$ that maps a set of nodes to another set of nodes on top of BPG where V is a set of BPG nodes and P is the power set of V. There are three operations: function composition $\circ$, function intersection $\cap$, and function union $\cup$, which allow chaining, intersecting, and combining the results of two graph traversals on V, respectively.

Through those simple operations, we break a complex graph traversal into multiple basic traversal components. We utilize the following symbols to represent them:

- $\text{MATCH}^{p}_{\text{type}}$ represents matching nodes with type (type) and properties (p),
- $N_1 \xrightarrow{[\text{REL}^{p}]^{\text{len}}} N_2$ denotes a path from node N_1 to node N_2. The path is connected by relationship REL with properties (p) and length (len).
- $a..b$ represents paths of variable length, ranging from a to b relationship hops inclusive.

5.2 Behavior patterns

In this subsection, we describe how to use graph traversals to represent three categories of non-compliant behaviors. As introduced in Section 2.2, we categorize non-compliant behaviors into three types: behavior existence, behavior content, and behavior frequency. We then provide examples to illustrate the

graph query representation for each type, as shown in the Table 1.

For behavior existence, we use Cross-App Launching as an example, as Formula (1) shows. Cross-App Launching refers to the behavior in which one app automatically initiates the startup of another app, potentially without the user's consent. We define the following graph query, which identifies method calls to setComponent within an Android app where the argument type is android.content.ComponentName. The setComponent method is used to specify the target component for an intent, which could be an activity, service, or BroadcastReceiver, either within the same app or across different apps. The android.content.ComponentName class provides the fully qualified name (package and class name) of the target component, allowing precise identification of which app or component the intent should address. If the package name in the ComponentName differs from the app's own package name, the app is targeting a component in a different app. The presence of such invocation statements within the app indicates the existence of cross-app launching behavior.

For behavior content, we take Collect Non-Modifiable Unique Device Identifiers as an example, as Formula (2) shows. We define the following graph query, which searches for method invocations that retrieve device identifiers, such as IMEI, Device ID, Subscriber ID, Mobile Equipment Identifier (MEID), Subscriber Identity Module (SIM) number, and active subscription info list. When an app accesses these identifiers, it indicates the app is attempting to gather unique device information, which could be used for tracking user identification or targeting specific devices.

For behavior frequency, we take Loop Popup After Permission Denial as an example, considering the scenario in which an app repeatedly displays permission request dialogs despite the user having denied authorization requests. We define the following graph query, as Formula (3) shows, which identifies a flow starting from an onClick method containing an android.permission literal, traversing through constructor calls, UI setup (setNegativeButton), static initializers, and additional onClick methods, ultimately leading to a dismiss method invocation. It also finds a separate path from a JIdentityStmt, through the CFG, to the same dismiss method, which then returns void. The goal is to uncover a specific sequence of interactions

Table 1 An example list of graph queries for different types of violations.

Violation type	Graph query
Behavior existence	$\text{MATCH}^{\{\}}_{\text{METHOD}} \xrightarrow{[\text{CONTAINS}^{\{\}}]^{1}}$ $\text{MATCH}^{\{\text{NAME:“setComponent”}\}}_{\text{CALL}} \xrightarrow{[\text{ARGUMENT}^{\{\}}]^{1}}$ $\text{MATCH}^{\{\text{TYPE:“android.content.ComponentName”}\}}_{\text{IDENTIFIER}}$ $\xleftarrow{[\text{REACHING\_DEF}^{\{\}}]^{1..5}} \text{MATCH}^{\{\}}_{\text{LITERAL}}$ (1)
Behavior content	$\text{MATCH}^{\{\text{NAME:“getImei”}\}}_{\text{JInvokeStmt}} \cup$ $\text{MATCH}^{\{\text{NAME:“getDeviceId”}\}}_{\text{JInvokeStmt}} \cup$ $\text{MATCH}^{\{\text{NAME:“getSubscriberId”}\}}_{\text{JInvokeStmt}} \cup$ $\text{MATCH}^{\{\text{NAME:“getMeid”}\}}_{\text{JInvokeStmt}} \cup$ $\text{MATCH}^{\{\text{NAME:“getSimSerialNumber”}\}}_{\text{JInvokeStmt}} \cup$ $\text{MATCH}^{\{\text{NAME:“getSerial”}\}}_{\text{JInvokeStmt}} \cup$ $\text{MATCH}^{\{\text{NAME:“getActiveSubscriptionInfoList”}\}}_{\text{JInvokeStmt}}$ (2)
Behavior frequency	$\Big(\text{MATCH}^{\{\text{NAME:“onClick”}\}}_{\text{METHOD}} \xrightarrow{[\text{CONTAINS}\|\text{CALL}]^{0..10}}$ $\text{MATCH}^{\{\text{NAME:“android.permission.X”}\}}_{\text{LITERAL}}$ $\xrightarrow{[\text{ASSIGNMENT}\|\text{ARGUMENT}\|\text{DEF}\|\text{USE\_OBJ}]^{1..5}}$ $\text{MATCH}^{\{\}}_{\text{InvokeStmt}} \xrightarrow{[\text{CALL}]}$ $\text{MATCH}^{\{\}}_{\text{METHOD}} \xrightarrow{[\text{CONTAINS}]}$ $\text{MATCH}^{\{\text{NAME:“init”}\}}_{\text{InvokeStmt}} \xrightarrow{[\text{CALL}]}$ $\text{MATCH}^{\{\}}_{\text{METHOD}} \xrightarrow{[\text{CONTAINS}]}$ $\text{MATCH}^{\{\}}_{\text{any}} \xrightarrow{[\text{USE\_FIELD}]^{2}}$ $\text{MATCH}^{\{\}}_{\text{any}} \xleftarrow{[\text{CONTAINS}]}$ $\text{MATCH}^{\{\}}_{\text{METHOD}} \xrightarrow{[\text{CONTAINS}]}$ $\text{MATCH}^{\{\text{NAME:“setNegativeButton”}\}}_{\text{InvokeStmt}}$ $\xrightarrow{[\text{ARGUMENT}\|\text{DEF}\|\text{USE\_OBJ}]^{1..5}}$ $\text{MATCH}^{\{\}}_{\text{any}} \xrightarrow{[\text{CALL}]}$ $\text{MATCH}^{\{\text{NAME:“clinit”}\}}_{\text{METHOD}} \xleftarrow{[\text{CONTAINS}]}$ $\text{MATCH}^{\{\}}_{\text{CLASS}} \xrightarrow{[\text{CONTAINS}]}$ $\text{MATCH}^{\{\text{NAME:“onClick”}\}}_{\text{METHOD}} \xrightarrow{[\text{CONTAINS}\|\text{CALL}]^{1..5}}$ $\text{MATCH}^{\{\text{NAME:“dismiss”}\}}_{\text{InvokeStmt}}\Big) \cap$ $\Big(\text{MATCH}^{\{\text{NAME:“dismiss”}\}}_{\text{IdentityStmt}} \xrightarrow{[\text{CF}]^{\text{any}}}$ $\text{MATCH}^{\{\}}_{\text{ReturnVoidStmt}}\Big)$ (3)

related to permission handling and the authorization cancellation UI. The presence of such a call sequence in an app indicates a behavior pattern of repeatedly prompting the user for permissions after authorization has been denied. In contrast, a benign app would typically redirect the user to a new page following a permission denial, rather than remaining on the original page and coercively prompting for authorization to proceed with further actions.

6 Evaluation

In this section, we evaluate BPGen by answering the following research questions.

- **RQ1-Coverage:** What are the primary types of non-compliant behaviours, and is BPG capable of modeling them?
- **RQ2-Effectiveness:** How effective is BPGen in detecting compliance violations?
- **RQ3-Performance:** What is the performance overhead of BPGen?

We perform our experiments on a server with 256 GB = 8 × 32 GB 3200 MHz DDR4, Intel(R) Corporation MontageJintide(R) C6248R 3000 MHz 24 cores 48 threads 1536/24 576/36 608 KB cache, and 2 × 400 GB 6 Gbps SATA HWE62ST3480L003N Hard Drive and 10 TB remote network disk.

The test set comprises 200 real-world apps, with half consisting of officially reported non-compliant apps, and the other half randomly selected from app markets such as TapTap[39], Huawei AppGallery[40], and Tencent MyApp[41]. Figure 2 illustrates the distribution of the size of Android Package Kit (APK) files within the test set.

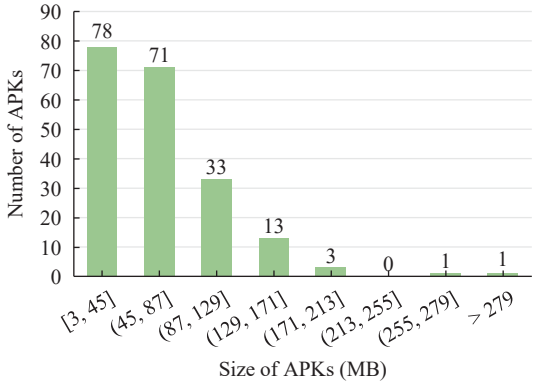


Fig. 2 Distribution of the number of APKs based on size in the dataset of 200 APKs.

6.1 RQ1-Coverage

In this subsection, we answer the research question on the capability of BPG in modeling non-compliant behaviors.

We start by summarizing seven prevalent non-compliant behaviors of mobile apps that frequently appear in official reports, as illustrated in Table 2.

First, we compare the violation detection scopes of various methods, including static taint analysis, dynamic analysis, and BPG.

Taint analysis excels at detecting personal privacy data leaks, such as unauthorized collection of unique device identifiers. Based on the behavior patterns of real-world apps, string resources in resource files—such as "privacy policy", "account deletion", and "targeted push"—can be considered as sources, while interactive components (e.g., buttons) serve as sinks, thereby enabling taint analysis to check for the presence of these behaviors. However, taint analysis is

Table 2 Coverage of methods: BPG, taint-based, and dynamic-based methods for behavior analysis, and coverage of systems: BPGen and three popular compliance detection platforms for seven types of violations (RQ1-Coverage). "Rounds" refers to the number of analysis needed by the method to detect *n* types of violations in a given application. "—" indicates that the metric is not applicable to the system. "√" denotes that the method or system supports detecting the corresponding violation, while "×" indicates that it does not.

ID	Non-compliant behaviour	Method			System			
		Taint analysis	Dynamic monitoring	BPG	OPPO	VIVO	Alibaba	BPGen
1	No Privacy Policy	√	×	√	×	×	×	√
2	No Account Deletion Function	√	×	√	×	×	×	√
3	Unauthorized Cross-App Launch	×	√	√	×	×	×	√
4	Loop Popup After Permission Denial	×	√	√	√	√	×	√
5	App Exits Upon Permission Denial	×	√	√	√	√	×	√
6	Collect Non-Modifiable Unique Device Identifiers	√	√	√	×	×	×	√
7	Unclosable Targeted Push	√	×	√	×	×	×	√
	Rounds	*n*	1	1	—	—	—	—

unable to detect behaviors such as auto-launching of apps or unexpected app exits, which do not directly involve the flow of tainted data from sources to sinks. Dynamic analysis refers to the process of analyzing a program by executing it and observing its behavior in real-time, but it is limited to identifying actions that should not occur, rather than confirming the presence of required behaviors. For example, if an app does not provide a privacy policy or an account deletion function, dynamic analysis may struggle to determine whether these actions were never triggered or were simply not provided by the app.

Additionally, Table 2 presents the number of analysis rounds required for different methods to detect n types of violations for a specific app. Taint analysis necessitates the specification of sources (e.g., sensitive data) and sinks (e.g., network transmission interfaces), often requiring multiple rounds of analysis for various types of violations. In contrast, dynamic analysis and BPG can analyze an app in a single round, resulting in greater detection efficiency.

Second, we introduce several advanced compliance detection platforms for the purpose of comparing with BPGen in Table 2. OPPO and VIVO's compliance detection platforms[42, 43] predominantly rely on dynamic analysis, leveraging their advantage as smartphone manufacturers to conduct real-device debugging. In contrast, Alibaba's detection platform[44] primarily employs static analysis and manual review methods.

As shown in the Table 2, BPG supports a broader range of behavior detection methods compared to the two most mainstream behavior analysis methods, and the corresponding system, BPGen, also exhibits greater detection coverage compared to popular platforms.

6.2 RQ2-Effectiveness

In this subsection, we address the research question of effectiveness evaluation by examining two metrics: FPR, i.e., FP/(TP + FP), and False FNR, i.e., FN/(TP + FN), where TP, FP, and FN denote the numbers of True Positives, False Positives, and False Negatives, respectively. In this context, discovering violations is considered positive data.

We compare BPGen with other platforms on these two metrics based on seven types of non-compliant behaviours. Alibaba platform does not support these detection items so it is excluded in the comparison.

We establish the benchmark for FP and FN through manual analysis of APK files. Table 3 shows the FPR and FNR of BPGen and existing platforms in detecting non-compliant behaviours. It is evident that BPGen has the lowest FPR and FNR, outperforming other compliance detection platforms. This superior performance is due to its precise modeling of the app's internal logic including control flow and data flow.

The analysis below explores the causes of FN and FP. We have defined the occurrence of non-compliant behaviors as positive. Some behaviors are considered compliant when present (e.g., the existence of a privacy policy), while others are compliant when absent (e.g., unclosable targeted push). As a result, both FN and FP can arise from similar causes. Therefore, our primary focus is to figure out why certain non-compliant behaviors were not detected in apps where they actually exist.

The primary cause of detection errors in BPGen arises from limitations in processing certain Chinese characters. Specifically, when assessing whether an app provides privacy policy, account deletion functions, or targeted push notifications, BPGen relies on matching specific strings or resource IDs. However, problems may arise if Chinese characters are not correctly parsed by Soot[37] or if these features are presented through web pages or customer service. In such cases, the absence of these keywords in the analyzed data can lead BPGen to incorrectly conclude that the app lacks these functionalities. This results in a false positive, where the app is wrongly deemed non-compliant with policy requirements.

To summarize our findings, BPGen detectes 14 violations within 13 previously unreported non-

Table 3 FPR (FP/(TP + FP)), and FNR (FN/(TP + FN)), of BPGen compared to two popular detection platforms across seven types of violations in the test set (RQ2-Effectiveness). "N/A" (Not Applicable) signifies that the platform lacks the capability to detect this violation.

(%)

ID	BPGen		OPPO		VIVO	
	FNR	FPR	FNR	FPR	FNR	FPR
1	0	5.03	N/A	N/A	N/A	N/A
2	2.86	44.85	N/A	N/A	N/A	N/A
3	0	0	N/A	N/A	N/A	N/A
4	0	0	100.00	0	100.00	0
5	0	0	100.00	0	100.00	0.50
6	0	0	N/A	N/A	N/A	N/A
7	2.86	20.00	N/A	N/A	N/A	N/A

compliant applications. Additionally, it uncoveres 20 new violations across 20 known apps that have been reported for other issues. In total, BPGen discoveres 34 zero-day violations. A selected summary of these apps and violations is provided in Table 4. Since the findings have not yet been communicated to the company, we have to hide the full package name. We reach out to the company as soon as possible. These apps exhibit high download volumes, indicating that the violations may have caused extensive impact.

6.3 RQ3-Performance

In this subsection, we answer the research question of the performance overhead of BPGen in generating BPG for real-world Android applications.

Our methodology is as follows: We run BPGen on all the APKs in the test set until the analysis was complete or until it timed out. We record the time taken to generate the graph through static analysis and the time required to store the graph in the graph database, as Fig. 3 shows. In general, the duration of static analysis (including both Soot analysis and graph generation) remains below 200 s, while the storage time (to save the graph to the database) remains below 300 s. The fluctuation in analysis time can be primarily attributed to the analysis of resource files within the APKs, like large images, which significantly prolongs the duration.

Additionally, we analyze the distribution of APK analysis time, which highlights the efficiency of BPG construction. Figure 4 shows that 90% of APKs can be analyzed within ten minutes. The results indicate that BPGen is efficient, with the majority of APKs processed within a reasonable time.

Finally, we measure the execution time of different graph queries separately. As illustrated in Fig. 5, the majority of queries are executed within 200 s. Notably, a subset of queries encountered timeouts. That is because the graph queries are unable to detect the specified behavior within BPGs after a long time (over 20 minutes). However, this often indicates the presence of compliance violations. For example, during the execution of the ExistPrivacyPolicy query, if the app does not provide a privacy policy, the relevant string resources information will not be recorded in BPG. This absence prevents the graph query from detecting the expected behavior, leading to a timeout. Nevertheless, it also indicates that the app has violated regulatory requirements for not having a privacy policy available to the user.

7 Discussion

In this section, we discuss the limitations of BPGen and propose directions for future work that will enhance its applicability.

Firstly, this paper focuses on detecting compliance in Chinese apps with local regulations, addressing a gap in existing research. However, BPG has the potential to be extended to other regulatory frameworks, such as the GDPR and the CCPA. This extension requires extracting relevant clauses from these regulations, identifying specific behavioral requirements or prohibitions, and adapting the detection criteria and graph queries by analyzing differences in implementation practices across different countries.

Secondly, the range of detectable violations still needs to be expanded. Since customizing graph query

Table 4 A selective list of non-compliant apps found by BPGen (RQ2 -Effectiveness). "Violation", "# Download", and "Source" columns represent violations detected in the app, the number of downloads at the source (which reflects the popularity of the app), and origins of the tested APK, respectively.

APK name	Version	Violation	# Download	Source
App A	2.1.9	No Account Deletion Function	1.17 million	TapTap
App B	2.19.18	No Account Deletion Function	0.56 million	Tencent MyApp
App C	14.7.2	No Account Deletion Function Loop Popup After Permission Denial	12.48 million	Huawei AppGallery
App D	10.10.56	No Account Deletion Function	20.46 million	Huawei AppGallery
App E	6.4.6.4	No Account Deletion Function	62.04 million	Huawei AppGallery
App F	8.10.19	No Account Deletion Function	1.00 billion	Huawei AppGallery
App G	6.7.3	No Account Deletion Function	23.99 million	Tencent MyApp
App H	6.6.9	No Account Deletion Function	13.87 million	Huawei AppGallery
App I	1.6.9	App Exits Upon Permission Denial	0.23 million	Huawei AppGallery
App J	3.5.2.1014	No Account Deletion Function	100.00 million	Huawei AppGallery

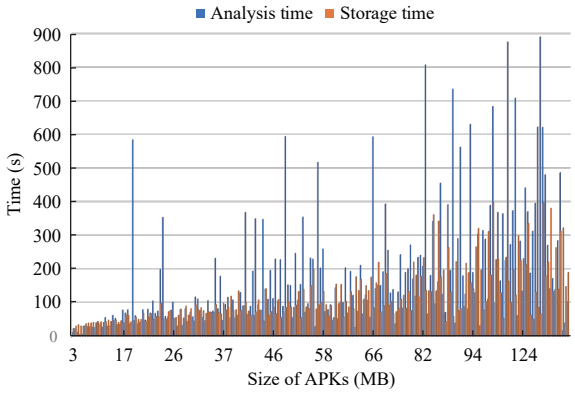


Fig. 3 Analysis time and storage time for APKs of varying sizes (RQ3-Performance).

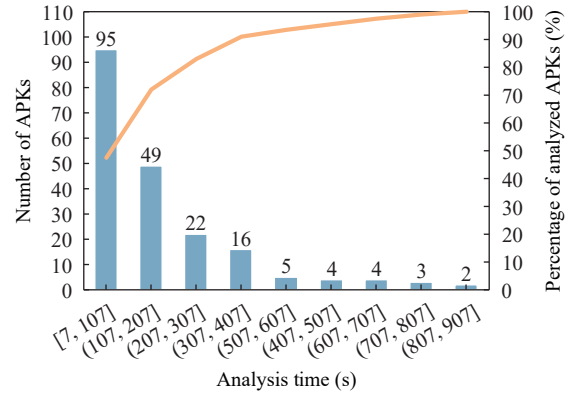


Fig. 4 Cumulative APKs Analyzed Over Time (RQ3-Performance). The bars show the number of APKs completed within each time interval, and the yellow line represents the cumulative percentage of all APKs analyzed.

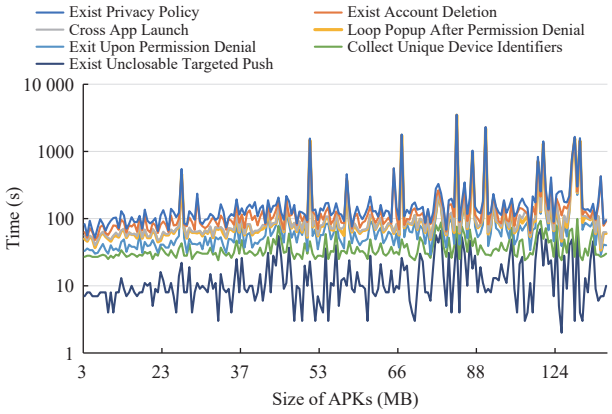


Fig. 5 Times required to execute different graph queries (RQ3-Performance).

statements for different violations requires manual participation, Section 6 only presents detection results for seven common types of non-compliant behaviors. Support for detecting a broader range of violations is still under development.

Additionally, BPGen currently cannot identify third-party libraries, limiting its ability to distinguish between behaviors originating from the app itself and those from Software Development Kits (SDKs). In future work, BPGen can support the detection of non-compliant behaviors from SDKs by integrating third-party library identification tools[45].

Lastly, as BPGen is a static analysis tool, it cannot effectively support the detection of violations that require dynamic analysis methods. For instance, regulations may impose restrictions on aspects such as font and color usage within an app, which necessitates screen analysis techniques such as Optical Character Recognition (OCR). Therefore, incorporating dynamic analysis will be a crucial step toward enhancing the completeness of compliance detection in future work.

8 Conclusion

In conclusion, this paper presents a universal and practical solution for compliance detection in Android apps. We introduce a novel graph structure called BPG, which integrates features from AST, CFG, CG, PDG, and PAG to effectively model an app. By extracting behavioral patterns from real-world applications and applying graph traversal techniques, the model efficiently identifies potential privacy violations. We implement a prototype system called BPGen to generate BPGs for Android apps and evaluate its efficiency and effectiveness on a test set comprising 200 real-world apps. The results indicate that BPGen completes the analysis for over 90% of the apps within ten minutes. Moreover, when compared to existing compliance detection platforms, BPGen significantly outperforms these systems in terms of detection scope, false positive rates, and false negative rates. Notably, BPGEN detectes 14 violations within 13 previously unreported non-compliant applications.

Acknowledgment

This work was supported by the National Key R&D Program of China (No. 2024YFE0203800), the National Natural Science Foundation of China (No. 62302438), and Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security.

References

[1] European Parliament, General Data Protection Regulation, Technical Report Regulation (EU) 2016/679, http://data.europa.eu.eli.reg/2016/679/oj, 2016.

[2] California Attorney General, California Consumer Privacy Act (CCPA) Regulations. Technical Report Cal. Civ. Code §§ 1798. 100–1798. 199, https://oag.ca.gov/privacy/ccpa, 2020.

[3] Standing Committee of the National People's Congress, Data Security Law of the People's Republic of China, Technical Report Order No. 84, (in Chinese), http://www.law-lib.com/law/law_view.asp?id=723460, 2021.

[4] Standing Committee of the National People's Congress, Personal Information Protection Law of the People's Republic of China, Technical Report Order No. 91, (in Chinese), National People's Congress, https://sfj.sx.gov.cn/art/2021/9/18/art_1488809_58923977.html, 2021.

[5] Ministry of Industry and Information Technology, Notice on Apps (SDKs) violating user rights (2024, batch 1, total batch 36), (in Chinese), https://www.miit.gov.cn/jgsj/xgj/fwjd/art/2024/art_615e90b52bd2458bb71b3cc0fb68355b.html, 2024.

[6] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, Y. Le Traon, D. Octeau, and P. McDaniel, FlowDroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps, *ACM SIGPLAN Not.*, vol. 49, no. 6, pp. 259–269, 2014.

[7] X. Pan, Y. Cao, X. Du, B. He, G. Fang, and Y. Chen, FlowCog: Context-aware semantics extraction and analysis of information flow leaks in android apps, in *Proc. 27th USENIX Conf. Security Symp.*, Baltimore, MD, USA, 2018, pp. 1669–1685.

[8] F. Wei, S. Roy, X. Ou, and Robby, Amandroid: A precise and general inter-component data flow analysis framework for security vetting of android apps, *ACM Trans. Priv. Secur.*, vol. 21, no. 3, p. 14, 2018.

[9] T. T. Nguyen, D. C. Nguyen, M. Schilling, G. Wang, and M. Backes, Measuring user perception for detecting unexpected access to sensitive resource in mobile apps, in *Proc. 2021 ACM Asia Conf. Computer and Communications Security*, Hong Kong, China, 2021, pp. 578–592.

[10] W. Enck, P. Gilbert, S. Han, V. Tendulkar, B. G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones, *ACM Trans. Comput. Syst.*, vol. 32, no. 2, p. 5, 2014.

[11] R. Bhoraskar, S. Han, J. Jeon, T. Azim, S. Chen, J. Jung, S. Nath, R. Wang, and D. Wetherall, Brahmastra: Driving apps to test the security of third-party components, in *Proc. 23rd USENIX Security Symp.*, San Diego, CA, USA, 2014, pp. 1021–1036.

[12] Z. Yang, M. Yang, Y. Zhang, G. Gu, P. Ning, and X. S. Wang, Appintent: Analyzing sensitive data transmission in android for privacy leakage detection, in *Proc. 2013 ACM SIGSAC Conf. Computer & Communications Security*, Berlin, Germany, 2013, pp. 1043–1054.

[13] C. Qian, X. Luo, Y. Le, and G. Gu, VulHunter: Toward discovering vulnerabilities in android applications, *IEEE Micro*, vol. 35, no. 1, pp. 44–53, 2015.

[14] X. Liu, Y. Zhang, Q. Yu, J. Min, J. Shen, R. Zhou, and Q. Zhou, SmartEagleEye: A cloud-oriented webshell detection system based on dynamic gray-box and deep learning, *Tsinghua Science and Technology*, vol. 29, no. 3, pp. 766–783, 2024.

[15] F. Yan, R. Wu, L. Zhang, and Y. Cao, SPIDER: Speeding up side-channel vulnerability detection via test suite reduction, *Tsinghua Science and Technology*, vol. 28, no. 1, pp. 47–58, 2023.

[16] L. Xu and J. Zhai, DCVAE-adv: A universal adversarial example generation method for white and black box attacks, *Tsinghua Science and Technology*, vol. 29, no. 2, pp. 430–446, 2024.

[17] S. Khan and M. Nauman, Interpretable detection of malicious behavior in windows portable executables using multi-head 2D transformers, *Big Data Mining and Analytics*, vol. 7, no. 2, pp. 485–499, 2024.

[18] Y. Qu, L. Ma, W. Ye, X. Zhai, S.Yu, Y. Li, and D. Smith. Towards privacy-aware and trustworthy data sharing using blockchain for edge intelligence. *Big Data Mining and Analytics*, vol. 6, no. 4, pp. 443–464, 2023.

[19] R. Slavin, X. Wang, M. B. Hosseini, J. Hester, R. Krishnan, J. Bhatia, T. D. Breaux, and J. Niu, Toward a framework for detecting privacy policy violations in android application code, in *Proc. 38th Int. Conf. Software Engineering*, Austin, TX, USA, 2016, pp. 25–36.

[20] L. Yu, X. Luo, X. Liu, and T. Zhang, Can we trust the privacy policies of android apps? in *Proc. 46th Annu. IEEE/IFIP Int. Conf. Dependable Systems and Networks*, Toulouse, France, 2016, pp. 538–549.

[21] L. Yu, X. Luo, J. Chen, H. Zhou, T. Zhang, H. Chang, and H. K. N. Leung, PPChecker: Towards accessing the trustworthiness of android apps' privacy policies, *IEEE Trans. Softw. Eng.*, vol. 47, no. 2, pp. 221–242, 2021.

[22] X. Du, Z. Yang, J. Lin, Y. Cao, and M. Yang, Withdrawing is believing? Detecting inconsistencies between withdrawal choices and third-party data collections in mobile apps, in *Proc. 2024 IEEE Symp. Security and Privacy*, San Francisco, CA, USA, 2024, pp. 735–751.

[23] T. T. Nguyen, M. Backes, N. Marnau, and B. Stock, Share first, ask later (or never?) studying violations of GDPR's explicit consent in android apps, in *Proc. 30th USENIX Security Symp.*, Virtual Event, 2021, pp. 3667–3684.

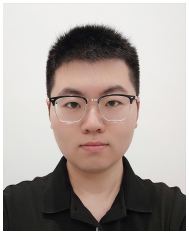
[24] D. Bui, B. Tang, and K. G. Shin, Do opt-outs really opt me out? in *Proc. 2022 ACM SIGSAC Conf. Computer and Communications Security*, Los Angeles, CA, USA, 2022, pp. 425–439.

[25] H. D. Bui, Assessment of privacy risks in mobile and web applications/services, Ph.D. dissertation, University of Michigan, Ann Arbor, MI, USA, 2022.

[26] B. Andow, S. Y. Mahmud, J. Whitaker, W. Enck, B. Reaves, K. Singh, and S. Egelman, Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with POLICHECK, in *Proc. 29th USENIX Conf. Security Symp.*, Virtual Event, 2020, p. 56.

[27] D. Bui, Y. Yao, K. G. Shin, J. M. Choi, and J. Shin, Consistency analysis of data-usage purposes in mobile apps, in *Proc. 2021 ACM SIGSAC Conf. Computer and*

Communications Security, Virtual Event, 2021, pp. 2824–2843.
[28] K. Zhao, X. Zhan, L. Yu, S. Zhou, H. Zhou, X. Luo, H. Wang, and Y. Liu, Demystifying privacy policy of third-party libraries in mobile apps, in *Proc. IEEE/ACM 45th Int. Conf. Software Engineering*, Melbourne, Australia, 2023, pp. 1583–1595.
[29] D. Bui, B. Tang, and K. G. Shin, Detection of inconsistencies in privacy practices of browser extensions, in *Proc. 2023 IEEE Symp. Security and Privacy*, San Francisco, CA, USA, 2023, pp. 2780–2798.
[30] Y. Wang, M. Fan, J. Liu, J. Tao, W. Jin, H. Wang, Q. Xiong, and T. Liu, Do as you say: Consistency detection of data practice in program code and privacy policy in mini-app, *IEEE Trans. Softw. Eng.*, vol. 50, no. 12, pp. 3225–3248, 2024.
[31] M. M. Ali, D. G. Balash, M. Kodwani, C. Kanich, and A. J. Aviv, Honesty is the best policy: On the accuracy of apple privacy labels compared to apps' privacy policies, *Proc. Priv. Enhanc. Technol.*, vol. 2024, no. 4, pp. 142–166, 2024.
[32] S. Wang, Y. Li, K. Wang, Y. Liu, H. Li, Y. Liu, and H. Wang, $M_{INI}S_{COPE}$: Automated UI exploration and privacy inconsistency detection of miniapps *via* two-phase iterative hybrid analysis, *ACM Trans. Softw. Eng. Methodol.*, vol. 34, no. 6, p. 167, 2025.
[33] B. Andow, S. Y. Mahmud, W. Wang, J. Whitaker, W. Enck, B. Reaves, K. Singh, and T. Xie, PolicyLint: Investigating internal privacy policy contradictions on Google play, in *Proc. 28th USENIX Security Symp.*, Santa Clara, CA, USA, 2019, pp. 585–602.
[34] M. Ferreira, T. Brito, J. F. Santos, and N. Santos, RuleKeeper: GDPR-aware personal data compliance for web frameworks, in *Proc. 2023 IEEE Symp. Security and Privacy*, San Francisco, CA, USA, 2023, pp. 2817–2834.
[35] I. Reyes, P. Wijesekera, J. Reardon, A. E. B. On, A. Razaghpanah, N. Vallina-Rodriguez, and S. Egelman, "Won't somebody think of the children?" Examining COPPA compliance at scale, *Proc. Priv. Enhanc. Technol.*, vol. 2018, no. 3, pp. 63–83, 2018.
[36] Cyberspace Administration of China Secretariat, Ministry of Industry and Information Technology Office, Public Security Bureau Office, and State Administration for Market Regulation Office, Methods for Determining Illegal Collection and Use of Personal Information by Apps, (in Chinese), https://wap.miit.gov.cn/jgsj/waj/wjfb/art/2020/art_8663d2afe61b40c3beb7c65bf6ec2a64.html, 2019.
[37] P. Lam, E. Bodden, O. Lhoták, and L. Hendren, The Soot framework for java program analysis: A retrospective, in *Proc. Cetus Users and Compiler Infrastructure Workshop*, 2011. https://www.researchgate.net/publication/236980420_The_Soot_framework_for_Java_program_analysis_a_retrospective.
[38] Neo4j, Neo4j, https://neo4j.com/, 2024.
[39] TapTap, (in Chinese), https://www.taptap.cn/, 2024.
[40] Huawei Technologies Co. Ltd., Huawei AppGallery, https://developer.huawei.com/consumer/en/appgallery/, 2024.
[41] Tencent Technology (Shenzhen) Company, Tencent MyApp, https://sj.qq.com/, 2024.
[42] OPPO mobile open platform, (in Chinese), https://open.oppomobile.com/new/introduction?page_name=audit-open, 2024.
[43] Vivo identity management system, (in Chinese), https://id.vivo.com.cn/, 2024.
[44] Alibaba cloud E-MAS developer tools, (in Chinese), https://emas.console.aliyun.com/service/devTool, 2024.
[45] X. Zhan, T. Liu, L. Fan, L. Li, S. Chen, X. Luo, and Y. Liu, Research on third-party libraries in Android apps: A taxonomy and systematic literature review, *IEEE Trans. Softw. Eng.*, vol. 48, no. 10, pp. 4181–4213, 2022.



Runqi Fan received the bachelor degree from Sichuan University, China, in 2023. He is currently pursuing the PhD degree at Zhejiang University, China. His research interests include program analysis and mobile security.

Zifeng Kang received the PhD degree from Johns Hopkins University, Baltimore, MD, USA, in 2025. He is now an assistant professor at Beijing University of Posts and Telecommunications, China. His research interests include security and privacy issues on the Web, such as leveraging dynamic analysis to systematically detect vulnerabilities in real-world websites.

Fan Wu received the MS degree from National University of Defense Technology, China, in 2013. He is currently pursuing the PhD degree at Zhejiang University, China. His research interests include network security, information security, and artificial intelligence.

Peng Hu received the MS degree from University of Electronic Science and Technology of China, China, in 2023. He is currently a researcher and a developer at Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China. His current research interests include Android static and dynamic analysis.

Song Li received the PhD degree from Johns Hopkins University, Baltimore, MD, USA, in 2022. He is currently a professor at Zhejiang University, China. His research interests include static/dynamic analysis and privacy, trying to solve real-world challenging problems such as increasing the accuracy and efficiency of vulnerability detection.

Weiting Chen received the bachelor degree from Jinan University, China, in 2012. She is currently an application security specialist at Ant Group Co., Ltd., China. She has over a decade of experience in mobile application security. Her research interests focus on mobile security.