Automatically Exploiting Potential Component Leaks in Android Applications

Li Li
Alexandre Bartel
Jacques Klein
Yves Le Traon

University of Luxembourg / SnT / SERVAL, Luxembourg

4 June 2014

978-2-87971-133-1
Automatically Exploiting Potential Component Leaks in Android Applications

Li Li, Alexandre Bartel, Jacques Klein, Yves le Traon
University of Luxembourg - SnT, Luxembourg
{li.li, alexandre.bartel, jacques.klein, yves.letraon}@uni.lu

Abstract—We present PCLeaks, a tool based on inter-component communication (ICC) vulnerabilities to perform data-flow analysis on Android applications to find potential component leaks (e.g., another component can potentially exploit the leak). To evaluate our approach, we run PCLeaks on 2000 apps, randomly selected from Google Play store. PCLeaks reports 986 potential component leaks in 185 apps. For each leak reported by PCLeaks, PCLeaksValidator automatically generates an Android app which tries to exploit the leak. By manually running a subset of the generated apps, we find that 75% of the reported leaks are exploitable leaks.

I. INTRODUCTION

The number of Android apps has increased exponentially in recent years. As of May 2012, Android became the most popular mobile operating system, running on the largest set of activated devices, and being the market leader in most countries [1]. Currently, more than 1.17 millions of apps exist in Google play and more than 80% of the apps are free of charge [2]. Not surprisingly, Android phone users are increasingly relying on the apps to manage their personal data. Because of that, the number of malware is also increasing. Kaspersky [3] has reported in its 2013 security bulletin that there are more than 148,427 mobile malware variants in 777 families and that 98.05% of the found mobile malware target the Android platform. As reported by Securelist [4], nearly half of the found Android malware are Trojan (e.g., SMS-Trojan) that steal personal data stored on the user’s smartphone.

Some families of malware focus on private data leaks. The more private data they want to leak, the more permissions they have to declare. An application asking for numerous permissions or for permissions the application should not require may alert the user [5]. For instance, a note book application asking for permission to send SMS looks suspicious. Instead of drawing the attention of the user by asking for permissions, malware may exploit vulnerabilities existing in other apps to leak sensitive data. Thus, it is essential to detect those vulnerable apps and thereby keep them from entering the app stores.

State-of-the-art approaches are focusing on either exploiting ICC vulnerabilities or detecting full private data leaks. For example, Epicc [6] is designed to detect ICC vulnerabilities (e.g., Activity Hijacking). But it does not perform data-flow analysis based on the detected ICC vulnerabilities. In other words, Epicc only knows where component may leak something, but it does not know if any data is flowing through the leak which yields many false positives. For private data leaks detection, AndroidLeaks [7], for example, uses static analysis technique to automatically find sensitive data leaks in Android apps on a massive scale. Another tool named IccTA [8], which performs inter-component (and also inter-app) communication based taint analysis to detect privacy leaks in Android apps. However, those tools are mainly focusing on private data leaks. To sum up, none of these tools tackle potential component leaks.

The only existing tool analyzing potential component leaks is ContentScope [9]. However, it only focuses on Content Providers, one of the four component types of an Android application. In this paper we present PCLeaks which finds potential component leaks on the other three components: Activity, Service and Broadcast Receiver. Note that when we talk about potential component leaks, each leak is always within a single component. It is not necessary to exploit inter-component potential leaks since such leaks are covered by intra-component potential leaks. Thus, we only perform intra-component data-flow analysis in this paper.

PCLeaks uses a static taint analysis technique to detect potential component leaks. A “traditional” leak starts with a source, a statement retrieving sensitive data from the system, and ends with a sink, a statement sending data outside of the application. In this paper, we focus on two types of “potential” leaks. The first type, Potential Passive Component Leak (PPCL), starts at an Android component entry-point and ends at a sink. For this leak, the component passively leaks data that it receives from other components. The second type, Potential Active Component Leak (PACL), starts at a source and ends at a component exit-point. For this leak, the component actively sends sensitive data to other components, which may leak the sensitive data intentionally or carelessly.

An example of the two types of component leaks is shown in Figure 1. The single component contains two sources, two sinks, two entry-points and two exit-points. There are four data-flow paths, marked as (A), (B), (C) and (D). Path (A) represents a private data leak. Those kind of leaks are well studied by tools such as Flowdroid [10], ScanDal [11] or DroidChecker [12]. Path (D) transfers data from an entry-point to an exit-point, it does not contain any real source or sink. Path (B) is a potential passive component leak (PPCL), starting at an entry-point and ending at a sink. Path (C) represents a potential active component leak (PACL), starting at a source and ending at an exit-point. In this paper we focus on detecting paths similar to (B) and (C), that is, potential component leaks. Roughly speaking, the fact that a leak of a component is exploited or not is dependent on another component.

Our approach relies on the Control Flow Graph (CFG) of the analyzed apps. If a sink node is reached from an entry-point node, a PPCL is detected. If a exit-point node is reached
from a real source node, a PACL is detected. The precision of our approach depends on the precision of the generated CFG. The event-driven nature of the Android system causes discontinuities in the CFG that PCLeaks models by generating a dummy main method for those discontinuities. Then, it performs a data-flow analysis on the precise CFG and finally outputs the detected potential component leaks. We also present a tool called PCLeaksValidator to automatically generate apps to validate leaks reported by PCLeaks. The purpose of the generated apps is to check whether leaks are true positives (e.g., really send sensitive data outside of the app) or not.

The contribution of this paper are as follows:

- PCLeaks, a static taint analysis tool to detect potential component leaks (PACLs and PPCLs).
- PCLeaksValidator, a tool to automatically generate applications to validate leaks reported by PCLeaks.
- An empirical experiment to evaluate PCLeaks and PCLeaksValidator over 2000 real-world Android applications.

The paper continues as follows. Section II explains the necessary background on Android security. Section III gives a motivating example and Section IV introduces the details of potential component leaks. In Section V, the paper discusses the implementation details of our approach. Section VI evaluates our approach. Limitations are discussed in Section VII. Section VIII presents the related work and Section IX concludes the paper.

II. BACKGROUND

A. Android Components

Components are the essential building blocks of an Android apps. As most components can be shared among applications, they act as entry points to the application. Four different types of components exist in Android. The fist one is Activity, which represents a screen with a user interface. The second one is Service, which is used to run long-time jobs in the background of the app. The third one is Content Provider, which provides a standard interface for other components to manage a shared set of data. The last one is Broadcast Receiver, which responds to system-wide broadcast announcements. Of these four types of components, only Activity provides a user interface.

An abstract object called Intent is used to communicate between two components. It describes an action to be performed (e.g., launching an Activity) and the data (extras) transferred by the action. There are two kinds of intents in Android: explicit intents and implicit intents. Explicit Intents, specify the target component. Implicit Intents do not specify the target component, but instead, they hold enough other information (e.g., action, category and data) for the system to determine an available component to run.

An app must declare all its components in a configuration file named AndroidManifest.xml. Implicit Intents can only reach components that declare one or more intent filters. A component not declaring any intent filters can still receive explicit Intents which normally come from the same app. However, it is still possible to receive explicit Intents coming from other apps. The only limitation is that those apps need to be signed by same signature. Intent filters are used to declare the capabilities of components (e.g., what types of broadcasts a receiver can handle). An example about declaring a component with its intent filter is shown in Listing 1. A service called SendSMSService is declared by element service. An intent filter is declared by element intent-filter. Under intent-filter, an action, a category and a data are declared by element action, category and data respectively. SendSMSService can receive all the implicit Intents which hold the same values of action, category and data as the ones declared by the intent-filter (e.g., action equals to "action SEND_SMS", category equals to "category SEND_SMS" and mime type equals to "text/plain").

```
<service android:name="SendSMSService">
  <intent-filter>
    <action android:name="action.SEND_SMS" />
    <category android:name="category.SEND_SMS" />
    <data android:mimeType="text/plain"/>
  </intent-filter>
</service>
```

Listing 1: An example about declaring a component with its intent filter.

B. Android Event-Driven Nature

Android apps are written in Java and thereby share the event-driven nature of Java. The event-driven nature introduces disconnections between parts of the code. In particular, the callback mechanism introduced in Java is used to implement the event-driven nature. For a concrete example, taking into account the java.lang.Thread class. It is used to implement native thread and execute long-time jobs. A developer can extend this class and override the run method, and then call the start method to send a launching thread event to the system. Then, the system will select an appropriate time to launch the thread by executing the run method. There is no code connection between start and run. When performing a static analysis this has to be modeled.

Similarly, the Android system introduces specific callback methods, called lifecycle methods. Each Android component

---

1 Note that for Broadcast Receivers, it is also available to programmatically register them.

2 For more information about intent resolution refer to the official documentation available at http://developer.android.com/guide/components/intents-filters.html

Resolution

Fig. 1: An example of Potential Component leaks, where (B) is a potential passive component leak (PPCL) and (C) is a potential active component leak (PACL).
has its own internal state. The Android system switches between states of a component by calling specific lifecycle methods of the component. Lifecycle methods are executed by the Android system according to user or system events. For example, when a user navigates back to an existing activity, the onResume method is called. The problem is that there is no direct code connection between lifecycle methods. Thus, it is essential to model the Android’s event-driven nature to precisely analyze Android apps.

C. Control-Flow Graph (CFG) and Taint Analysis for Android

To simplify our analysis process and to better describe our approach, we use CFG which are introduced by Reps et al. [14] to intermediarily and visually represent the relationships of the app codes. The CFG is made up of a collection of intra-precedence control-flow graph (IPCFG) and the IPCFGs are connected through the call relations in the CFG. In an IPCFG, sname and ename are used to specify the start node and the end node respectively. For a procedure call, two nodes (call and return-site) and three edges (call-to-return-site, call-to-start and exit-to-return-site) are used to represent them.

An example about CFG and its codes is shown in Figure 2. For procedure output(x), two nodes (call and return-site) are used to represent it. Three edges are also involved for procedure output(x). The first edge is call-to-return-site from node call to node return-site in procedure output(x). The second edge is call-to-start form node call to node soutput and the last edge is exit-to-return-site from node eoutput to node return-site.

In recent years, many tools to generate call-graphs for Android applications and perform taint-analysis have been developed by researchers such as CHEX [15], TrustDroid [16] or LeakMiner [17]. However, few of them are available online. In this paper we use FlowDroid [10] ³, a highly precise tool for data-flow analysis on Android applications. FlowDroid models the lifecycle of Android components and performs a context, flow, field and object-sensitive taint analysis.

III. Motivating Example

We start by giving a motivating example shown in Listing 2 and Listing 3. Two Android apps, namely Application1 and Application2, are introduced. Application1, an example of PPCL, contains a Service named SendSMSService in which a content of short message is obtained from the received Intent when the Service is launched. Then, the content of the message is sent outside of the app by SMS. Application2, an example of PACL, contains an Activity named GetDeviceIdActivity in which a device id is obtained when the Button btn is clicked. Then, the device id is stored in an Intent and is sent to other components by method startService().

³FlowDroid is open-source and can be downloaded at https://github.com/secure-software-engineering/soot-infoflow-android
malicious applications. The next Section describes the different kinds of potential leaks that we detect in this paper.

IV. POTENTIAL COMPONENT LEAKS

In this section, we detail the classification of potential component leaks. As already said, Potential Passive Component Leak (PPCL) starts at an Android component entry-point and ends at a sink. Potential Active Component Leak (PACL), starts at a source and ends at a component exit-point. Moreover, this paper focuses on three types of Android components: Activity, Service, and Broadcast Receiver. As a result, since PPCL and PACL can occur in each of these component types, we define six \((3 \times 2)\) different kinds of potential leaks. Figure 3 illustrates these six kinds of potential leaks.

![Figure 3: Classification of potential component leaks.](image)

A. PPCL

Figure 3a shows the three kinds of PPCL, that can be triggered by Intent spoofing: Activity Launch Leak, Service Launch Leak and Broadcast Injection Leak.

**Activity Launch Leak.** Exported Activities can be launched by other components (or applications) with either explicit or implicit Intents. In some case, an Activity may be launched with an Intent and then leaks the Intent’s data outside the activity or application. This can be used by malicious apps to passively leak sensitive data. We call this specific leak Activity launch leak.

**Service Launch Leak.** As for an exported activity, exported services can also be launched by other components or applications. If the service leaks the received Intent’s data outside the service or application, the leak is called a Service launch leak. For example, Application1 in Listing 2 contains a Service launch leak.

**Broadcast Injection Leak.** A Broadcast Receiver may leak the data it receives from other components or applications. A malicious app can use this to make the Broadcast Receiver passively leak sensitive data. We call this specific leak a Broadcast injection leak.

B. PACL

Figure 3b shows the three kinds of PACL, that may be triggered by Intent hijacking: Activity Hijacking Leak, Service Hijacking Leak and Broadcast Theft Leak.

**Activity Hijacking Leak.** A malicious Activity can be launched through an Intent hijacking. If the original component reads sensitive data and stores them into an Intent (e.g., extras). The malicious Activity may hijacking the Intent and thereby manipulates the sensitive data. Therefore, when sensitive data is obtained and is sent to other Activities through inter-component communication (e.g., startActivity), we call it Activity hijacking leak.

**Service Hijacking Leak.** A malicious Service may hijacking an Intent, which contains sensitive data in its Extras. In this situation, we call the original component contains Service Hijacking leak. For example, a Service hijacking leak exists in Application2 of Listing 3.

**Broadcast Theft Leak.** We call a component that contains Broadcast Theft Leak as it reads sensitive data and sends them through an Intent to a Broadcast Receiver. Because the Intent can be stolen by a malicious Broadcast Receiver.

To summaries, We define six kind of potential leaks (two categories: PPCL and PACL) in this paper. The reason why we distinguish the different component types in each category is that different semantics are performed by Android system when multiple receivable components exist. In detail, for multiple launchable Activities, the system will pop up a selection box to let user decide which Activity is going to be launched. The Android system will randomly select a Service to launch for multiple launchable Services. The Android system launches all available Broadcast Receivers for multiple launchable Broadcast Receivers.\(^4\)

V. IMPLEMENTATION

In this section we discuss the implementation of our approach to find and validate potential component leaks. Our approach strongly relies on the FlowDroid tool and features four steps as illustrated Figure 4. The first steps describe PCLeaks. In Step1 (Section V-A), PCLeaks extracts the list of reachable Android components. In Step2 (Section V-B), PCLeaks builds a precise CFG with the information provided by Step1. In Step3 (Section V-C), PCLeaks uses the source and sink methods collection computed by SuSi [18] to perform taint analysis on the precise CFG provided by Step2 and then reports a list of potential leaks it found. In the last step (Section V-D), we use PCLeaksValidator to automatically generate applications to validate leaks reported by PCLeaks. Finally, for each leak reported by PCLeaks we manually run the app, which is generated by PCLeaksValidator to check whether PCLeaks reported a real leak or not.

A. Step 1: Preprocessing

In the first step (Step 1: apktool in Figure 4), PCLeaks extracts the Android XML file using apktool\(^3\). Based on the generated XML file, PCLeaks extracts the following artifacts:

1. The list of declared components;
2. The permission attribute of components;
3. The exported attribute of components.

Computing (1) gives the list of components of an application. If a component is exported it means it can be reachable from other applications. If it is not exported, it cannot receive Intents from other apps. In other words, PCLeaks does not

\(^3\)This is not always the case for ordered broadcasts, where the available Broadcast Receivers are executed one by one. As each receiver executes in turn, it can abort the broadcast so that it won’t be passed to other receivers.

\(^4\)https://code.google.com/p/android-apktool/
analyze non-exported component. We use the algorithm listed in Algorithm 1 to check whether a component is exported or not. First, we check whether the attribute exported is explicitly set in the manifest or not. If it is set explicitly, we directly return the value of the attribute (true or false). If it is not set explicitly, we need to analyze the default value of the component which is related to the component’s type. If the component’s type is ContentProvider and its app’s version is less than or equal to 16, then the component is exported by default. If the component’s type is not ContentProvider and it contains an intent-filter element, then the component is exported. Otherwise, the component is not exported.

But are all those exported components always accessible from another application? Only components not protected by a permission are reachable. Computing (2)\(^6\) gives the list of permission protected components. If a component is protected by a permission, an app trying to communicate with it must declare the same permission. Note that this only works for permissions with protection level normal or dangerous. Since four protection levels (normal, dangerous, signature and signatureOrSystem) exist for permissions in Android, if a component is protected by permissions at signature or signatureOrSystem level, it is impossible for a malicious app to access the components because the malicious apps would need to be signed with the same signature of they accessed app \(^7\).

We performed a short study of permission protected components on 2000 Android apps which have 11,584 components in total. Among these components, only 13 are protected by a permission and 3 out of the 13 permissions are protected by signature level. The rest 10 permissions are protected by normal level. Since the number of protected components is negligible, we do not take permissions into consideration for the results in this paper and only include exported and non-permission protected components in the list of reachable components.

B. Step 2: Precise CFG Building with FlowDroid

As mentioned in Section II-B, due to the Android event-driven nature, CFG of apps are imprecise. The imprecision is mainly caused by two kinds of methods: lifecycle methods and callback methods. We use FlowDroid [10] to model lifecycle and callback methods of Android components. We modified FlowDroid to only take into account reachable components extracted at step1 to build the precise CFG. Since we experienced that FlowDroid cannot properly analyze some apps because of the memory limitation, the precise CFG does improve the efficiency of FlowDroid.

For lifecycle methods, because the call sequence is well-defined in the Android’s documents, what is needed is to simulate the sequence to call all the lifecycle methods. Since a component has different states in its life time, when leaving a state, it may has different choices and thereby executing different lifecycle methods. Let us take an Activity as a concrete example. When another Activity comes into the foreground (onPause() will be executed) two lifecycle methods may be selected by the Android system depending on user events. If the user returns to the activity, the onResume() is executed. If the Activity is no longer visible, onStop() is executed. FlowDroid generates a graph where both methods are reachable.

For callback methods, FlowDroid has a collection of callback methods extracted by analyzing the Android documentation. Then, for each component, FlowDroid checks whether it contains callback methods or not. If a callback methods exists

---

\(^6\)Referring to the second item presented above.

\(^7\)This would be possible if the private key of the developer of the benign app is leaked.
in a component, FlowDroid makes the component reachable in the application graph.

To model the application graph, FlowDroid generates a dummy main method to call all the involved components and their lifecycle and callback methods. The CFG of generated dummy main method for GetDeviceIdActivity illustrated in Listing 3 is shown in Figure 5. The lifecycle methods are connected from onCreate() to onStart() (shown in A) and the callback methods are also connected from onStart() to onClick() (shown in B). Method onClick() is only called after method onResume() which simulates an Activity going to the running state. Because onResume() is not explicitly overridden in GetDeviceIdActivity, we do not model it in the generated dummy main method.

C. Step 3: Taint Analysis

Taint analysis is a kind of data-flow analysis. In this work, we leverage FlowDroid which is based on Heros [19], an IFDS/IDE problem solver, to perform a inter-procedural data-flow analysis for Android apps. The collection of source and sink methods used in this work is generated by SuSi [18]. SuSi is an open source tool that automatically identifies the source and sink methods in Android applications.

Because in this paper we focus on detecting potential component leaks, we also add all the entry-points and exit-points of components as source methods and sink methods respectively. The original FlowDroid is only sensitive to the sinks exactly defined in its configuration document. The problem is that FlowDroid will ignore all the override methods of the defined sinks. For example, if an sink startActivity() of class Activity is defined and startActivity() (or this.startActivity()) is called in the body of class CustomActivity which extends from class Activity, FlowDroid will not take into account startActivity() in CustomActivity as a sink even it should be. We corrected FlowDroid’s strategy of identifying sink methods so that FlowDroid is sensitive to all the override sink methods as well.

From the leaks detected by FlowDroid, we filter all the non-potential component leaking paths if it does not starts with a component entry-point method or if it does not ends with a component exit-point method. If a potential component path starts with a component entry-point method, it means a PPCL is detected. If a potential component path ends with a component exit-point method, it means a PACL is detected.

D. Step 4: Validation

In order to validate the reported potential leaks, we developed a prototype tool called PCLeaksValidator, which automatically generates Android apps to check the validity of reported potential leaks. For a PPCL, the generated app contains only one component (Activity), which appropriately prepares an Intent and uses it to launch the target component. The extra data of the generated app always uses the default value we defined no matter what key of the extra is. For a PACL, the generated app contains also only one component, where the type depends on the exit-point of the PACL.

Take Listing 3 as a concrete example. The code generated by PCLeaksValidator is shown in Listing 4, in which the Intent-Filter of Service1 is also appropriately configured so that it can be launched by Application 2. Note that for PACLs, PCLeaksValidator uses method Log.i as sink method for all the generated apps.

```
1 class Service1 extends Service{
2   int onStartCommand(Intent i, int f, int id){
3     String str1 = i.getStringExtra("sms-content");
4     Log.i("PCLeaksValidator", str1);
5     return;
6   }
```

Listing 4: The code generated by PCLeaksValidator for Listing 3.
We manually run the two apps (one is the analyzed app and the other is generated by PCLeaksValidator) to validate the exploited component leak. Currently, we are not able to automatically run the two apps to validate the results. Because it needs us to automatically trigger the communication event between the tested two apps. In addition, some communication events rely on the user inputs at run time and thereby make them becomes harder to be automatically triggered. As future work, we would like to enhance PCLeaksValidator to automatically validate the two apps.

VI. Evaluation

In Section VI-A, we present our experimental results of running PCLeaks on a set of 2000 real-world applications. Then, we detail two case studies (one for PPCL and the other for PACL) in Section VI-B and Section VI-C respectively.

A. Experimental Results

We run PCLeaks on 2453 apps randomly selected from the Google play store. The apps are different from the collection we use to perform a short study about the permission protection of components. We run PCLeaks on an a Core i7 CPU and give 8 Gb to the Java VM heap. In average, PCLeaks processes an app in about 40 seconds. Out of 2453 apps, 453 of them could not be processed (e.g., due to errors of insufficient memory, Type mask not found for type or Manifest contains more than one manifest node), so in this paper, we show our experimental results for 2000 apps.

Potential component leaks detected by PCLeaks are shown in Table I. PCLeaks reports 143 PACLs among 43 apps, which is shown in row 1, where only implicit Intents are taken into consideration. There is a significant difference between explicit and implicit Intents for PACL results. PCLeaks reports 15,260 leaks among 1149 apps, where 14,286 leaks are Activity Hijacking Leaks. These results match our expectation that the leaks related to Activity Hijacking Leaks is higher than Service related leaks. This matches the design philosophy of Android components. On the one side, Broadcast is designed to communicate with other apps. But instead of exporting the Service component, those apps export a Broadcast for other apps and in this case, Broadcast is designed to communicate with other apps. On the one side, Broadcast is designed to communicate with other apps.

We randomly select 20 leaks (10 for PACLs and 10 for PPCLs) and run PCLeaksValidator on them. Then, we manually run the generated apps with their related source apps to check the detected leaks. We confirm that 7 PACLs and 8 PPCLs are true positives. The false positives are introduced by the condition-insensitivity of PCLeaks, insufficient string analysis of PCLeaksValidator or bugs in specific case. For example, a false positive is caused by a leak containing an unfeasible condition in its path. Since PCLeaks is currently condition-insensitive, it over-approximates all the possible paths whenever condition statements exist. Another confirmed false positive comes from insufficient string analysis. Since PCLeaksValidator performs a simple string analysis that only traverses the single intra-procedural control-flow graph to determine the value of a string variable. If the value of strings are not determined, we simply ignore them currently. This makes PCLeaksValidator yield false alarms (e.g., the key of an extra data is missing). In the future work, we would like to perform precise string analysis [20] to obtain better results.

Figure 6 classifies the detected PACLs and PPCLs according to the types of leak. The highest number of detected potential component leaks are Activity Launch Leaks. PCLeaks reports 534 leaks on Activity Launch Leaks. Indeed, it is easy to launch an Activity since all the Activity’s information are defined in AndroidManifest.xml. In PACLs, the highest number of detected leaks are Activity Hijacking Leaks, where 110 leaks are reported. Since Activity is well-used in Android, the number of Activity Hijacking Leaks and Activity Launch Leaks confirm our expectation that the leaks related Activity should be the highest detected leaks.

The number of Broadcast related leaks (including system broadcasts) is higher than Service related leaks. This matches the design philosophy of Android components. On the one side, Broadcast is designed to communicate with other apps. It should have more potential leaks. On the other side, some apps (e.g., In-app payments) are designed to provide functional services. But instead of exporting the Service component, those apps export a Broadcast for other apps and in the Broadcast they access the non-exported Service through explicit Intent. Therefore, It is normal to have more Broadcast related leaks.

To summaries, PCLeaks reports 986 leaks among 185 apps, where 5 of them contains both PACLs and PPCLs. We manually check 20 results, where 15 of them are true positives. PCLeaks reaches a precision of 75% on the randomly selected results.

<table>
<thead>
<tr>
<th>Leaking Type</th>
<th>#. of Leaks</th>
<th>#. of Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>PACLs (with explicit Intent)</td>
<td>143</td>
<td>43</td>
</tr>
<tr>
<td>PACLs (with explicit Intent)</td>
<td>15260</td>
<td>1149</td>
</tr>
<tr>
<td>PPCLs (without permission, without non-exported)</td>
<td>843</td>
<td>147</td>
</tr>
<tr>
<td>PPCLs (without permission, without non-exported)</td>
<td>848</td>
<td>150</td>
</tr>
<tr>
<td>PPCLs (with permission, with non-exported)</td>
<td>5540</td>
<td>514</td>
</tr>
</tbody>
</table>

We take these two rows as potential component leaks result, the other rows are used to demonstrate the influence of explicit/implicit Intent, permission and export attributes to the results.

8 Package name is com.beenverified.android.tests.in_app
Listing 5: The log data of com.beenverified.android.tests.in_app when it receives a broadcast.

Fig. 6: Breakdown of detected PACLs and PPCLs by type.

B. Case Study on PPCL

**In-app payments** is an Android application that offers in-app purchases. It contains a Broadcast Receiver named `com.example.dungeons.BillingReceiver`, which logs everything of the received Intent in method `logIntent()`. `logIntent()` is called by the entry point method named `onReceive()`. Both `logIntent()` and `onReceive()` are defined in class `com.example.dungeons.BillingReceiver`.

PCLeaks reports a PPCL for **In-app payments**. To verify the report, PCLeaksValidator automatically generates an app, which sends a broadcast message to the detected app with action named `com.android.vending.billing.IN_APP_NOTIFY` and two extras within the Intent. The two extras are named `inapp_signature` and `response_code`. Then, we manually run the two apps on our test device. No matter what data stored in the two extras, **In-app payments** logs all of them. A piece of the log data is shown in Listing 5.

Note that malicious apps can use this PPCL (without user intervention) to communicate with each other to avoid transferring data directly between malicious apps and thereby bypassing the detection of some specific malware detection tools. For example, **Malicious app 1** first sends the data to **In-app payments** through a broadcast. Then, **In-app payments** logs all the data it received to disk. At last, **Malicious app 2** parses the log data of **In-app payments** to obtain the data transferring from **Malicious app 1**.

C. Case Study on PACL

In this case study, we show an application which contains potential active component leak. More specifically, it contains an Activity hijacking leak. **GetPhoneInfo** is an app to obtain the information of the running Android operating system as well as the running phone itself. The obtained information include SubscriberId, name of the phone (e.g., HTC One), SimOperator Name (e.g., Orange) and many others.

All the sensitive data (phone’s information) are obtained in method `getInfo()` of class `GetPhoneInfo`. Then, an implicit Intent is initialized with an action named `android.intent.action.SEND` and a type named `message/rfc822`. The sensitive data is stored into the Intent with an extra named `android.intent.extra.TEXT`. After that, `startActivity()` is called to communicate with other applications (or components), which also send the sensitive data to other applications.

**MaliciousApp** is an Android application automatically generated by PCLeaksValidator to test whether the app **GetPhoneInfo** will actively leak sensitive data or not. In **MaliciousApp**, we developed a component which can receive action `android.intent.action.SEND` and the data type is set to `*/*`.

We run the two apps on our test device. The results (screenshots) are shown in Figure 7. Note that to better illustrate the results, we refactored the generated malicious apps to explicitly show the received data in a text view. Figure 7a shows the app selection screen when `startActivity()` is executed. Figure 7b shows the received sensitive data when user launch the **Malicious** application. The sensitive data is
transferred from Figure 7a to Figure 7b. That means a PACL can become a real leak, which exposes the user’s private data to other applications.

VII. LIMITATIONS

At the moment, we do not handle URIs, which are well used by Content Provider. Therefore, PCLeaks is not able to exploit potential leaks on Content Provider. PCLeaks is not aware of multiple threads, reflections and condition statements. PCLeaksValidator does not handle URIs to generate incomplete malicious apps, which is not able to exploit the corresponding potential component leaks. Currently, PCLeaksValidator only performs string analysis within a single method which may cause false alarms. PCLeaksValidator is not able to run the apps to automatically validate the reported potential component leaks. Some rarely used ICC methods such as startActivities are not tackled in this work. The native code used by some apps is not analyzed as well.

VIII. RELATED WORK

PCLeaks is developed to detect potential component leaks in Android apps. Information leaks and component vulnerabilities detection in Android apps are two main research topics mostly related to our work.

Information leak detection has been studied for decades and new discoveries continue to be found in recent years [21, 22, 23]. Among them, both static analysis and dynamic analysis are performed. One of the most sophisticated static analysis approach is Flowdroid [10], a context-, flow-, field-, object-sensitive and lifecycle-aware static taint analysis tool for Android apps. It tracks tainted data between pre-defined source and sink methods. If a tainted data is transferred from a source method to a sink method, then a information leak is reported. Several other approaches including SCanDroid [24], LeakMiner [17] and AndroidLeaks [7] also use static analysis to detect privacy leaks. TaintDroid [25] is one of the most sophisticated dynamic approach on detecting privacy leaks in Android apps. It extends from the Android mobile-phone platform and tracks the flow of privacy sensitive data through third-party apps at run time. CopperDroid [26] is another dynamic testing tool which uses stimulating technique to exercise the app to find malicious activities. More recently, DroidTrack [27] tracks and visualizes the sensitive information diffusion on Android to prevent sensitive data leaks. However, these approaches mainly focus on detecting private data leaks, our approach is different from them and focuses on detecting potential component leaks.

Component Vulnerabilities detection is another hot topic related to our work. CHEX [15] is a tool to detect component hijacking vulnerabilities in Android applications by tracking taints between externally accessible interfaces and sensitive sources. The entry-point model of CHEX requires an enumeration of all possible "split orderings" which is not necessary in PCLeaks. ComDroid [28] and Epicc [6] are another two tools that focus on detecting inter-component vulnerabilities. However, they do not perform sensitive data-flow analysis. In other words, ComDroid and Epicc are able to detect component vulnerabilities (e.g., Activity Hijacking, Broadcast Injection). But they do not detect whether a component vulnerability leaks sensitive data or not. PCLeaks is different from these tools that it based on component vulnerabilities to detect potential sensitive leaks.

The related works introduced in this section are either focusing on privacy leaks detection or focusing on component vulnerabilities. Our approach research on the both sides to perform potential component leaks detection. ContentScope [9] is a tool like our approach to detect sensitive data leaks on components in Android applications. But it only focus on Content Provider. ContentScope also detects content pollution in Android applications, which is not handled currently by PCLeaks. But with little modification (e.g., defining pollution methods as sink methods for PCLeaks ), our approach is able to detect component pollution. However, detecting component pollution is out of scope of this paper, we take it as our further work.

Another state-of-the-art works are trying to enhance the user privacy by permission removal [29, 30] or trying to reduce the attack surface of Android applications [31]. Those approaches could be used to complement our approach to prevent components from leaking sensitive data but are out of scope of this paper.

IX. CONCLUSION

In this work, we present a tool named PCLeaks to exploit potential component leaks and a tool named PCLeaksValidator to automatically generate a correspond malicious apps to validate the results of PCLeaks. Concretely, PCLeaks first builds a precise control-flow graph for the analyzed apps. Then, it performs statically taint analysis with a well-defined set of source and sink methods to identify potential active component leaks and also potential passive component leaks. We test PCLeaks on 2000 apps randomly selected from Google Play. Among the 2000 apps, PCLeaks reports PACLs in 43 apps with 143 leaks and also reports PPCLs in 147 apps with 843 leaks. By manually checking 20 results through running the generated malicious app with its source app, we confirm that 15 (or 75%) of them are true positives.

In the future work, we would like to enhance PCLeaksValidator to support automatically validating the exploited leaks. We also like to automatically repair the Android application containing potential component leaks.

REFERENCES


