Vincenzo Riccio

Tutor: Anna Rita Fasolino

XXXI Cycle - III year presentation

ENHANCING AUTOMATED GUI EXPLORATION TECHNIQUES FOR ANDROID MOBILE APPLICATIONS

Personal Background

- **Candidate:** Vincenzo Riccio
- **Cycle:** XXXI
- **Fellowship**: PhD grant

- **Graduation**: MSc with honors in Computer Engineering at the University of Napoli Federico II
- **Research Activity**: Software Testing Automation
- **Research Field**: Software Engineering
- **Collaborations:**

Universidade do Porto
Faculdade de Engenharia
FEUP

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Research Group

- REsEarch gRoup of Software Engineering (REvERSE) at the University of Naples Federico II





- **Mission:** REvERSE@Unina aims at developing novel methods, techniques and tools that advance development and evolution of software systems. We are interested in all the software lifecycle processes, with a special focus on: Software Maintenance, Reverse Engineering, and Testing

Credits Summary

	1st Year	2nd Year	3rd Year	Entire PhD Course	Check
Modules	17	19	0	35,5	**30-70**
Seminars	10,2	5,2	2,7	18,1	**10-30**
Research	34	46	58	138	**80-140**
Total	61,2	70	61	191,6	**>180**

Experience Abroad

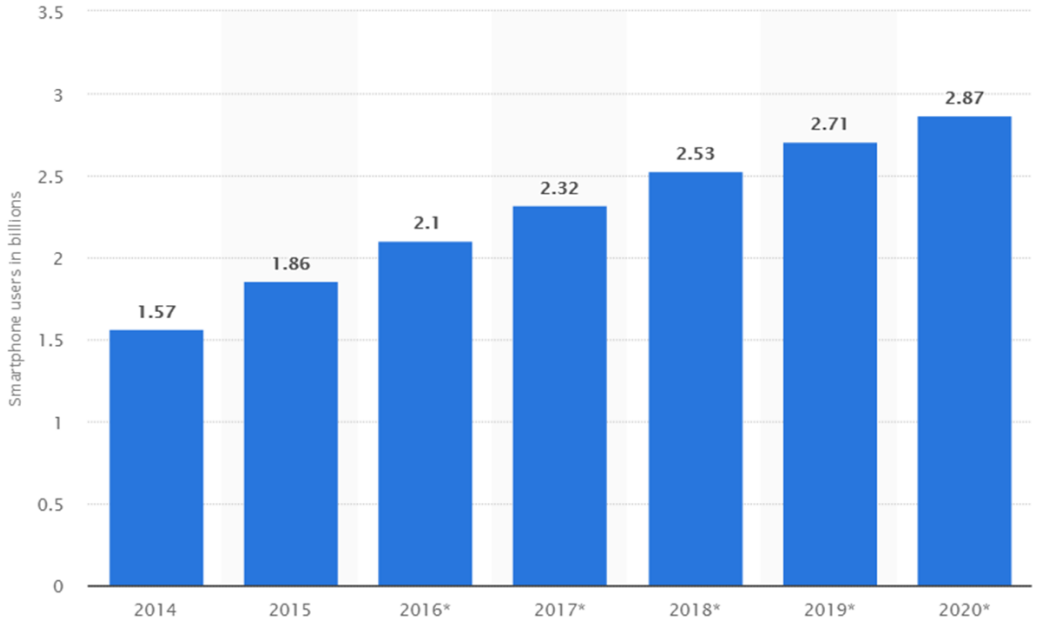
UNIVERSITÄT PASSAU

Fakultät für Informatik und Mathematik

- **Topic:** Novel evolutionary search algorithms for testing mobile applications
- **Start:** 17 April 2018
- **End:** 7 August 2018

Prof G. Fraser, Chair of Software Engineering II

Smartphone users worldwide



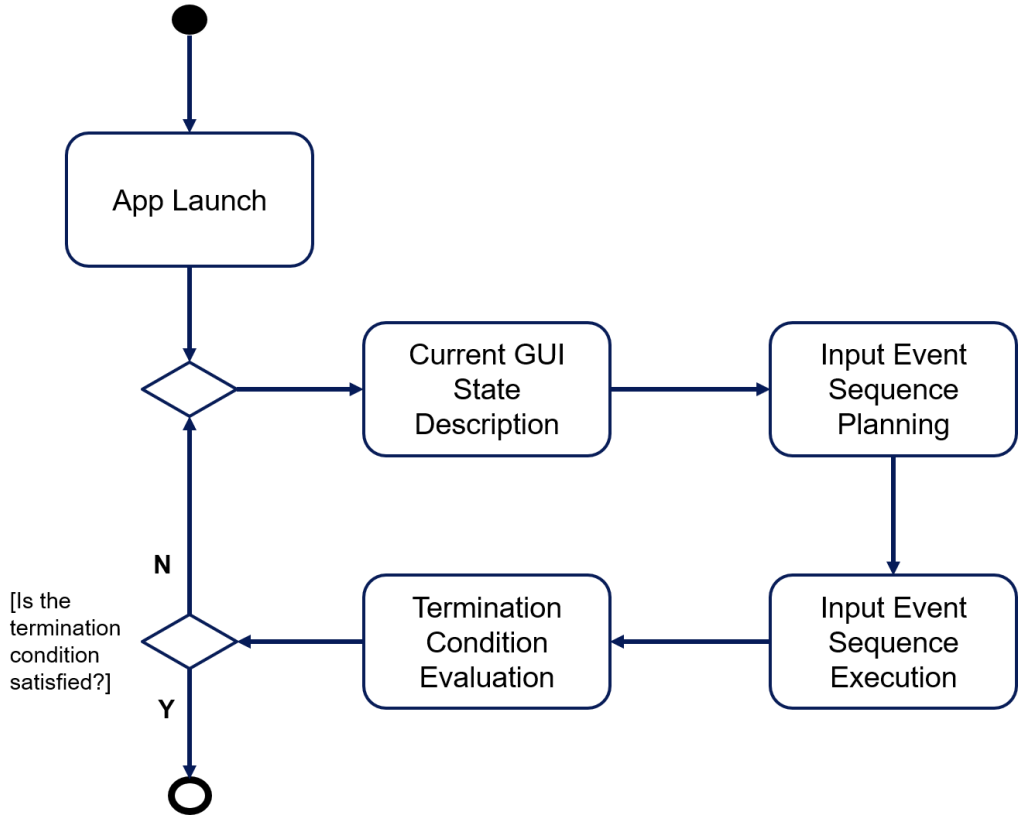
https://www.statista.com/statistics/330695/

- There is a constant demand for new mobile apps
- Android is today the world's most popular mobile operating system

Automation Tools

- The demand for app quality has grown together with their spread
- Automation tools can facilitate software quality engineering activities since they save humans from routine, time-consuming and error-prone manual tasks

Automated GUI Exploration Techniques (AGETs)



D. Amalfitano, N. Amatucci, AM. Memon, P. Tramontana, AR. Fasolino, “A general framework for comparing automatic testing techniques of Android mobile apps”, Journal of Systems and Software, 2017

Challenges

Enanche AGETs by:

1. targeting mobile-specific features
2. exploiting app-specific knowledge that only human users can provide

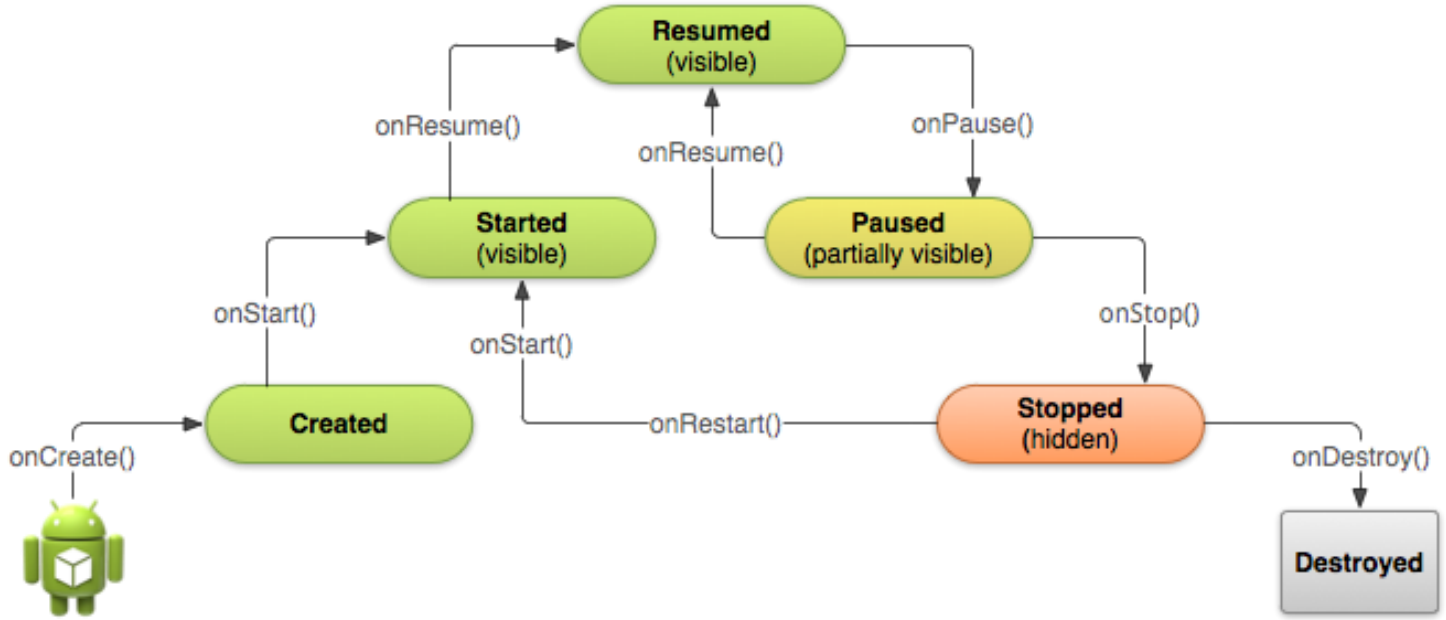
Challenge #1

Targeting mobile-specific features

The Android Activity Lifecycle

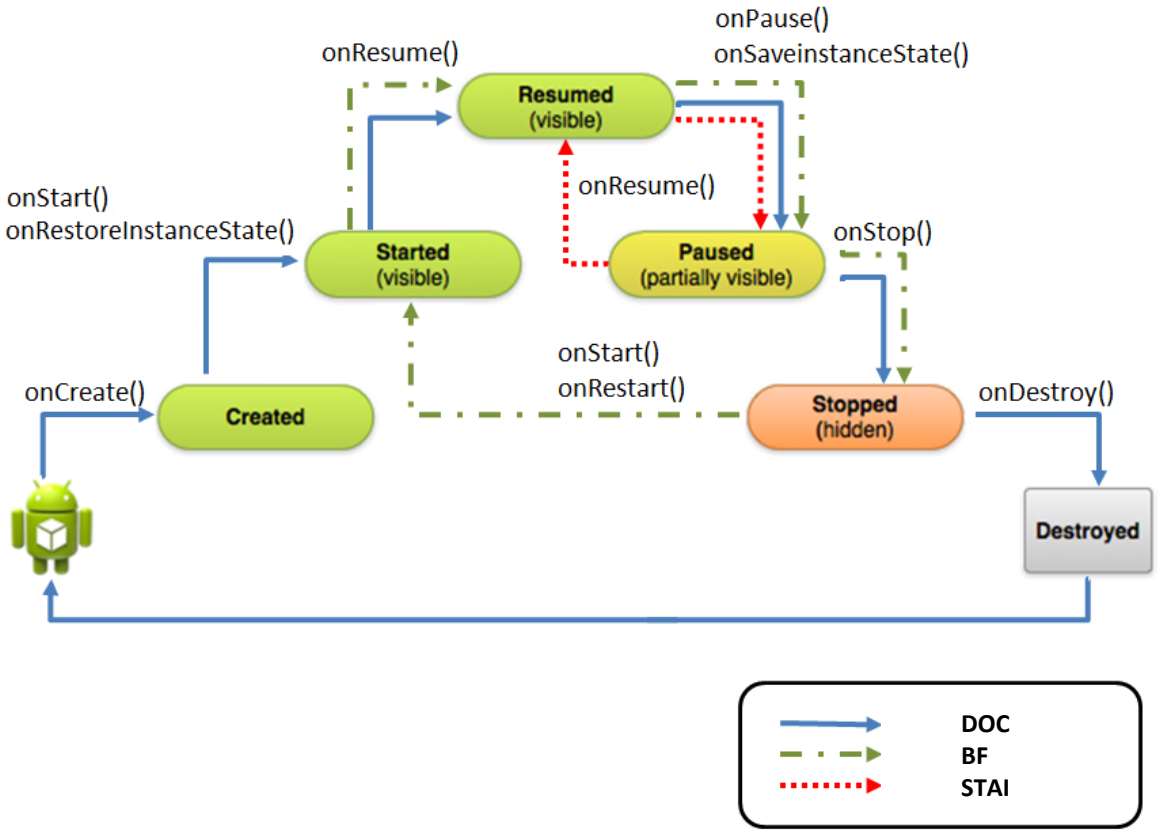
Android Activity Lifecycle

- An Android app is composed by one or more Activities
- Each Activity represents a single screen
- The Android Framework defines a peculiar lifecycle for Activity instances



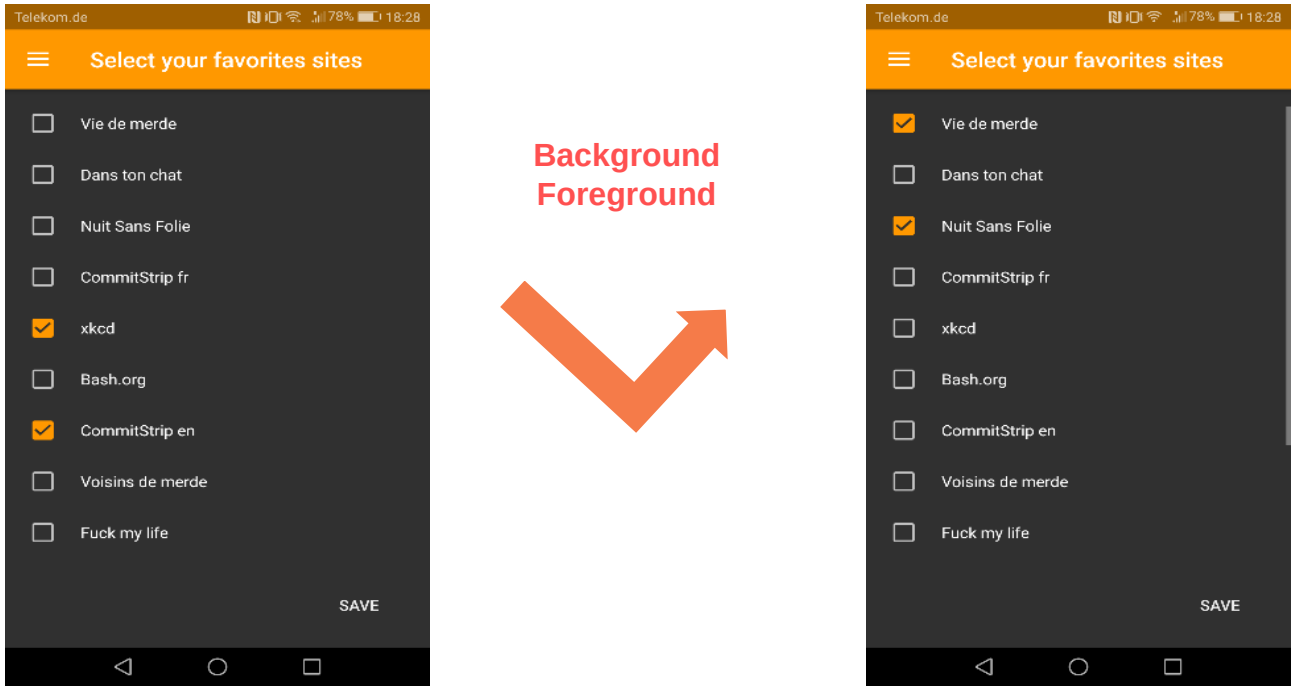
Lifecycle Event Sequences

- Mobile-specific events able to exercise the Activity lifecycle



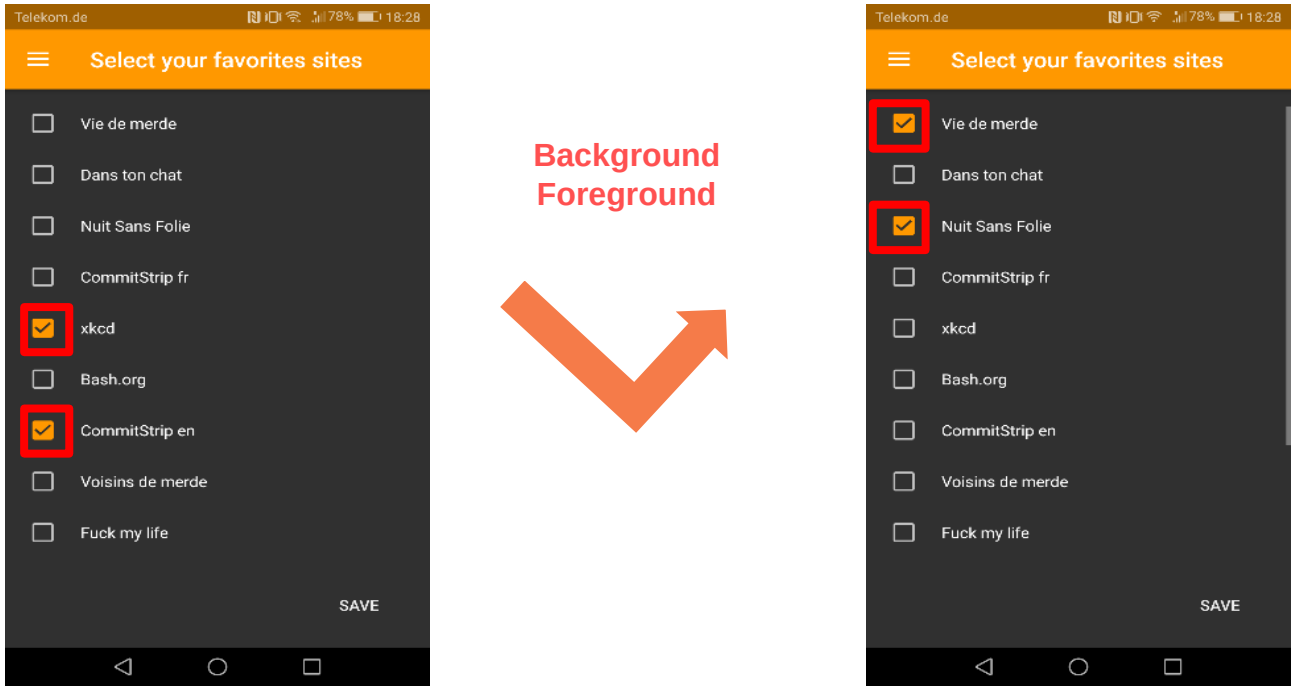
Motivating Example: GUI Failure

- GUI failures consist in the manifestation of an unexpected GUI state

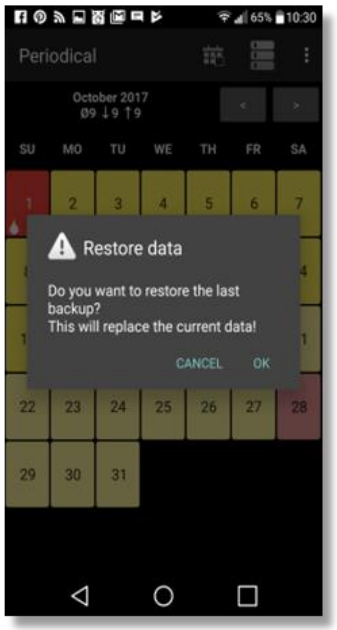


Motivating Example: GUI Failure

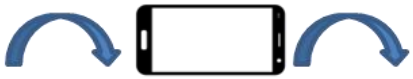
- GUI failures consist in the manifestation of an unexpected GUI state



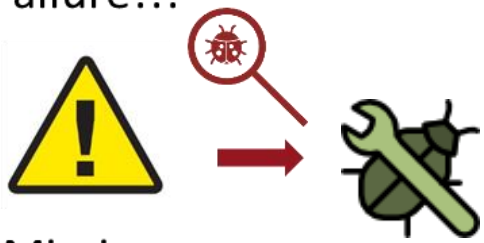
Exploratory Studies



Double
Orientation
Change



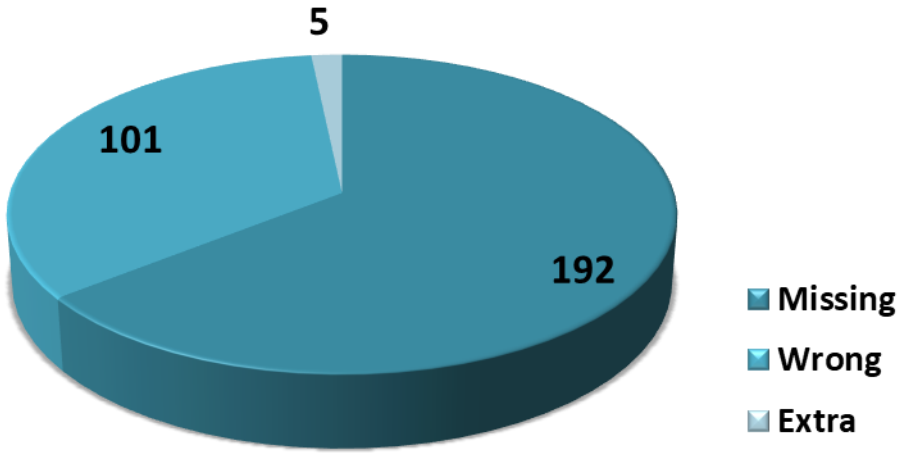
DOC GUI Failure!!!



Missing Dialog

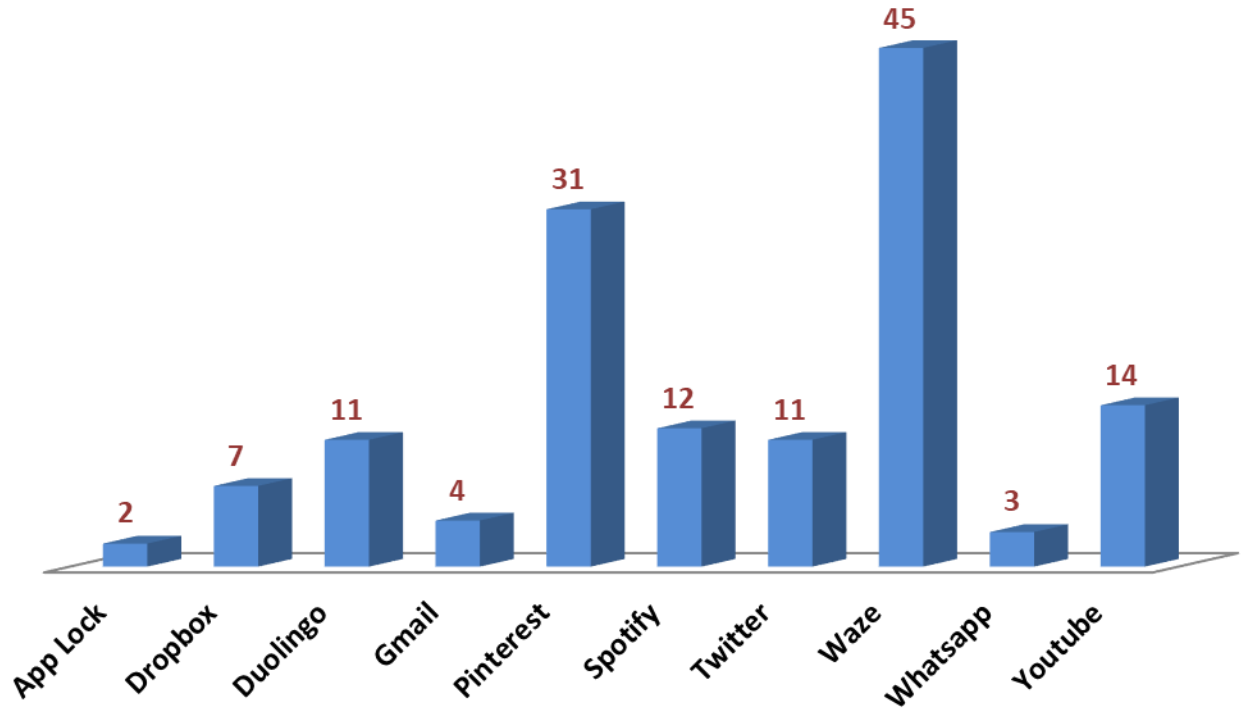
Exploratory Study 1

- 68 open-source apps
- 86% of the considered apps are affected by GUI failures due to orientation changes
- Most of the detected failures involve Dialog objects missing from the GUI after the DOC
- 6 classes of common faults causing GUI failures have been identified

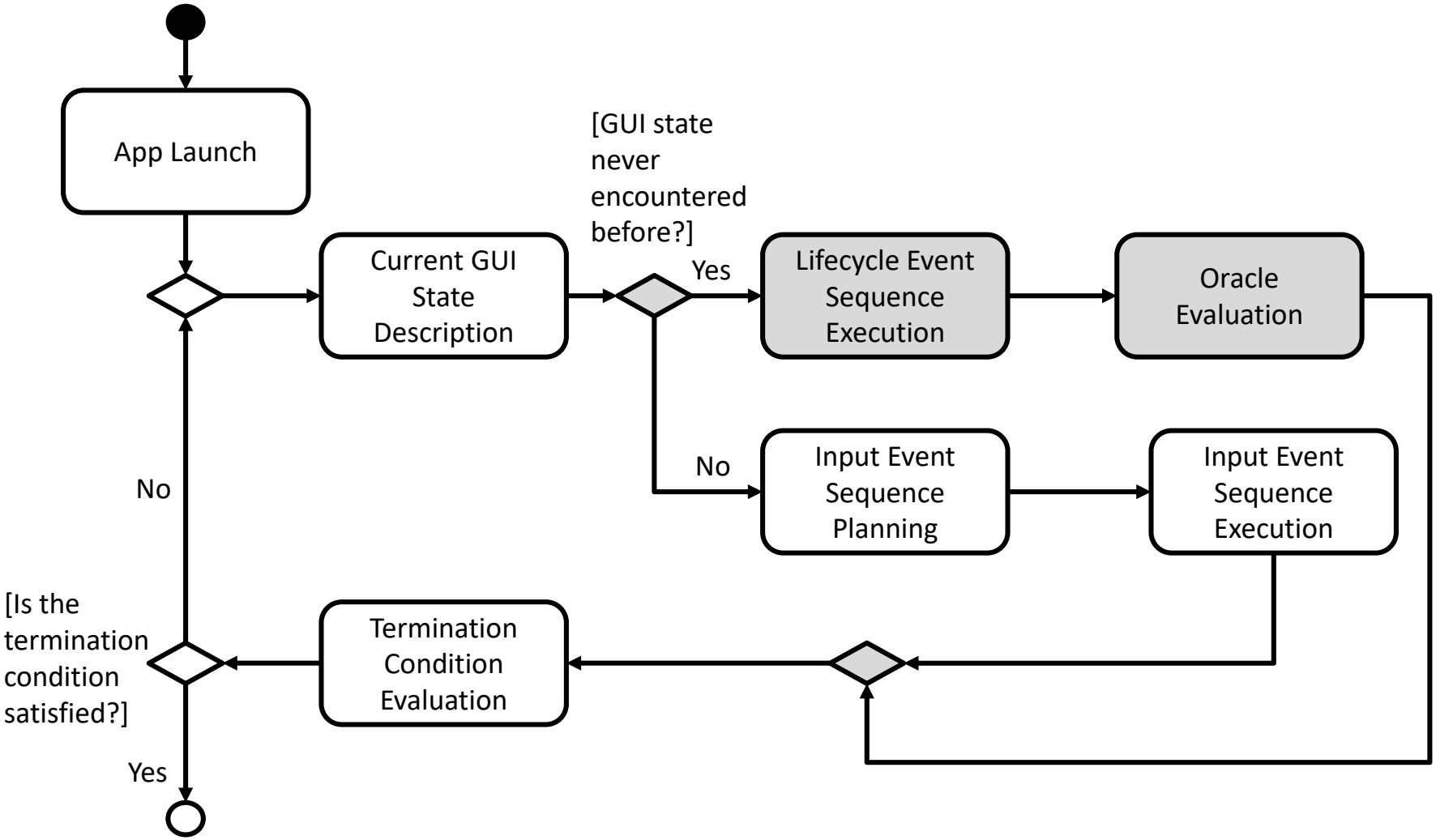


Exploratory Study 2

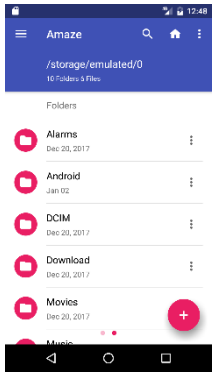
- 15 industrial-strength apps
- All the considered apps are affected by GUI failures due to orientation changes



The ALARic Approach

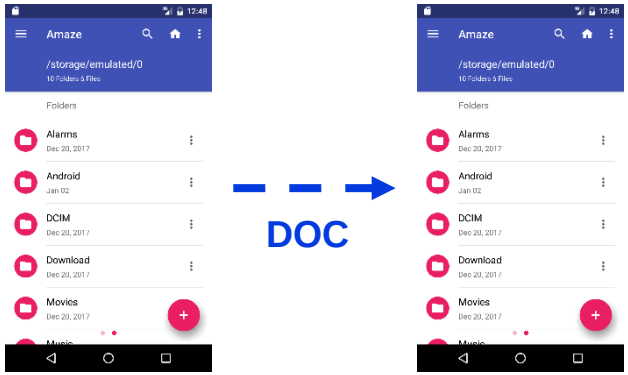


ALARic Workflow Example



A

ALARic Workflow Example

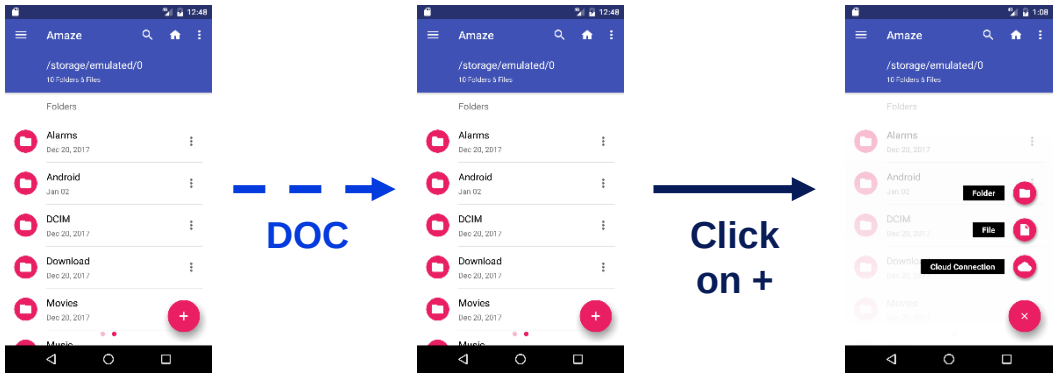


A **B**

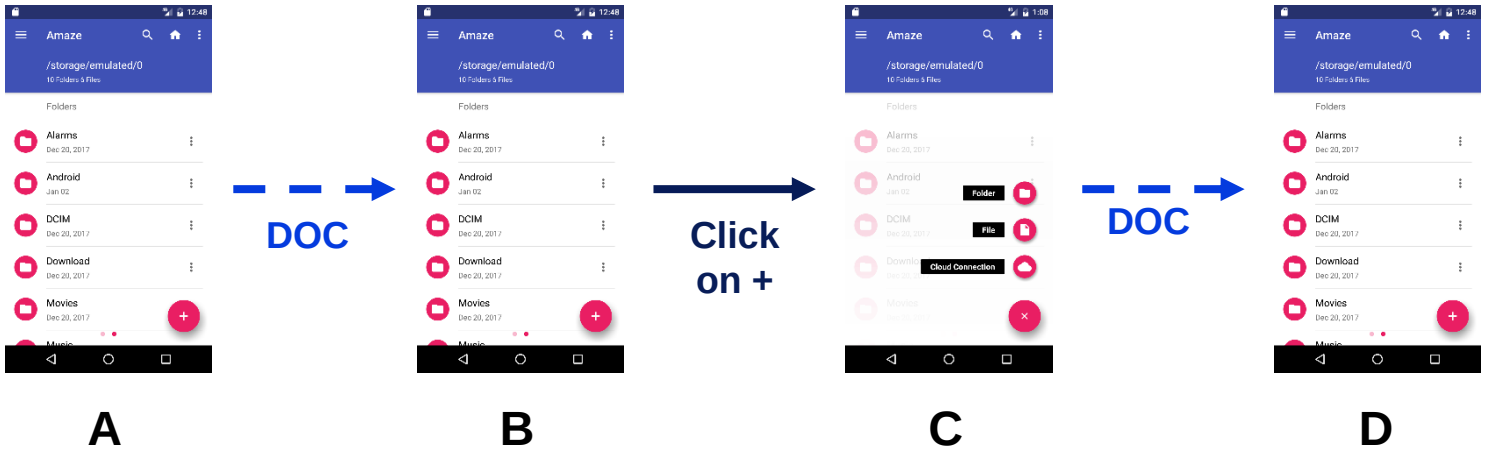
ALARic Workflow Example

B = A

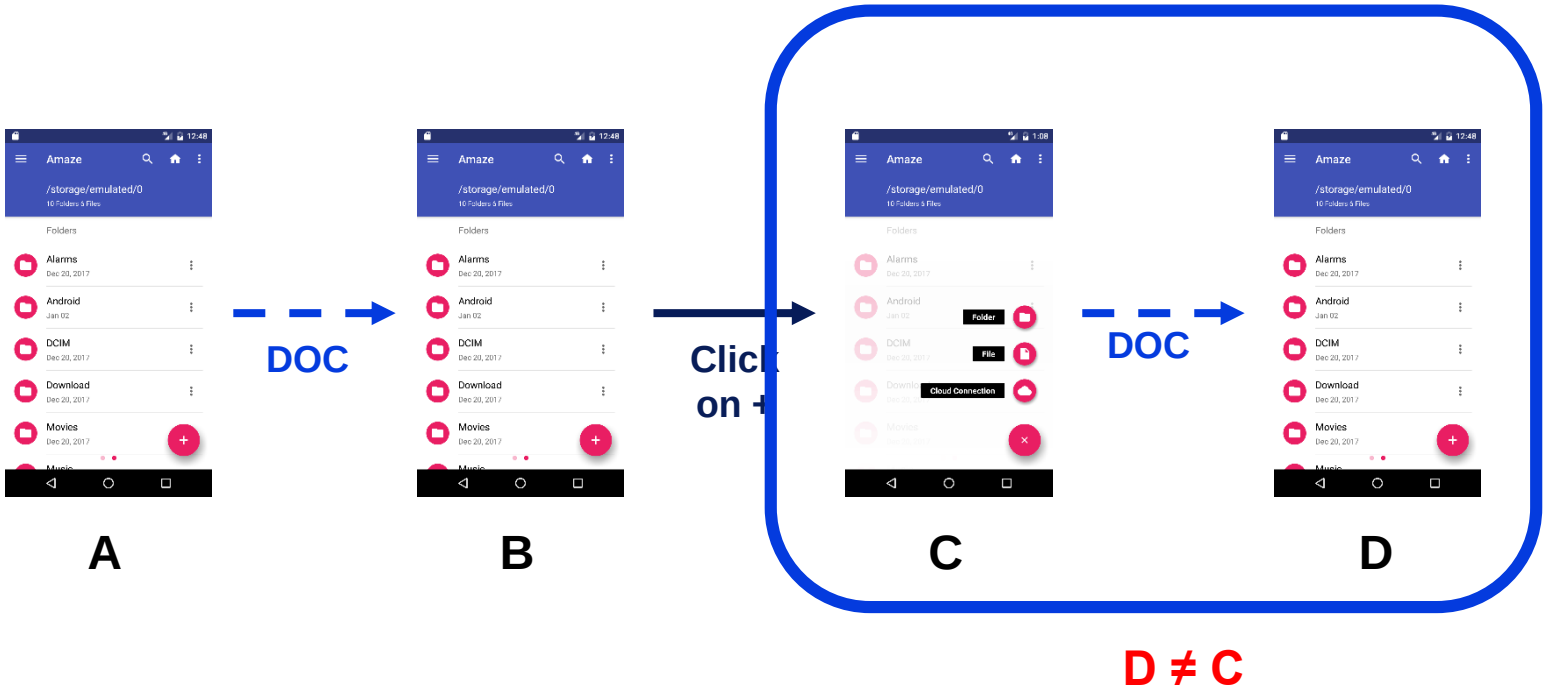
ALARic Workflow Example



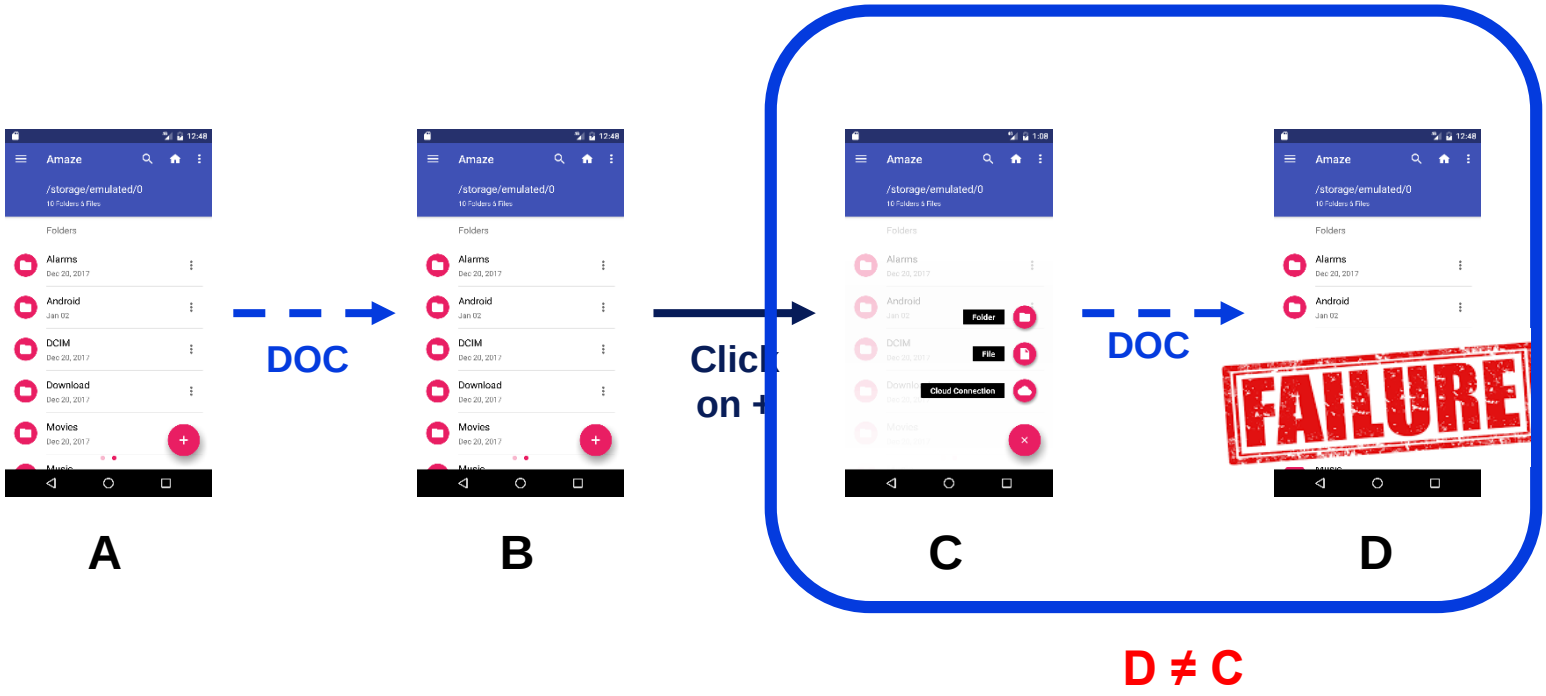
ALARic Workflow Example



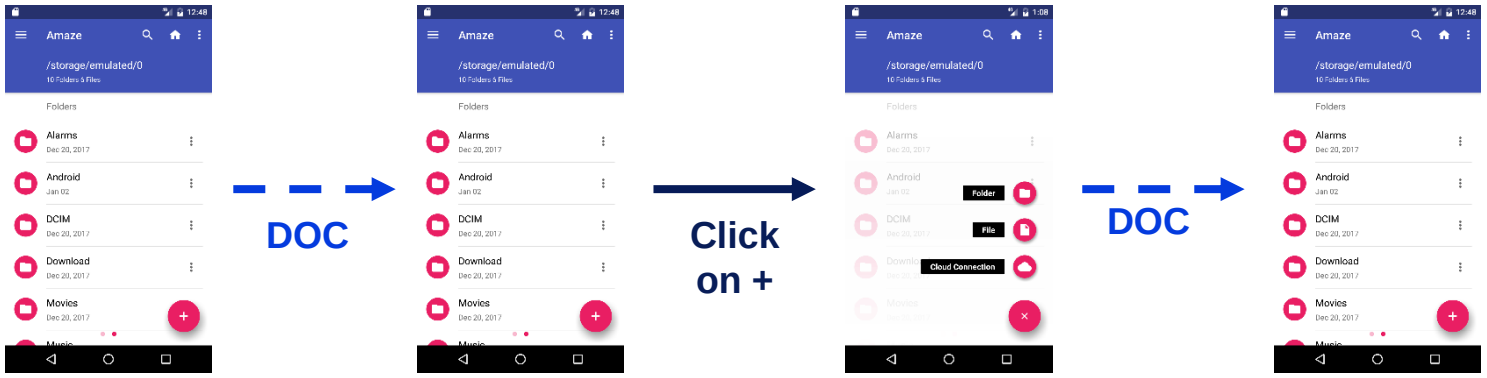
ALARic Workflow Example



ALARic Workflow Example



ALARic Workflow Example

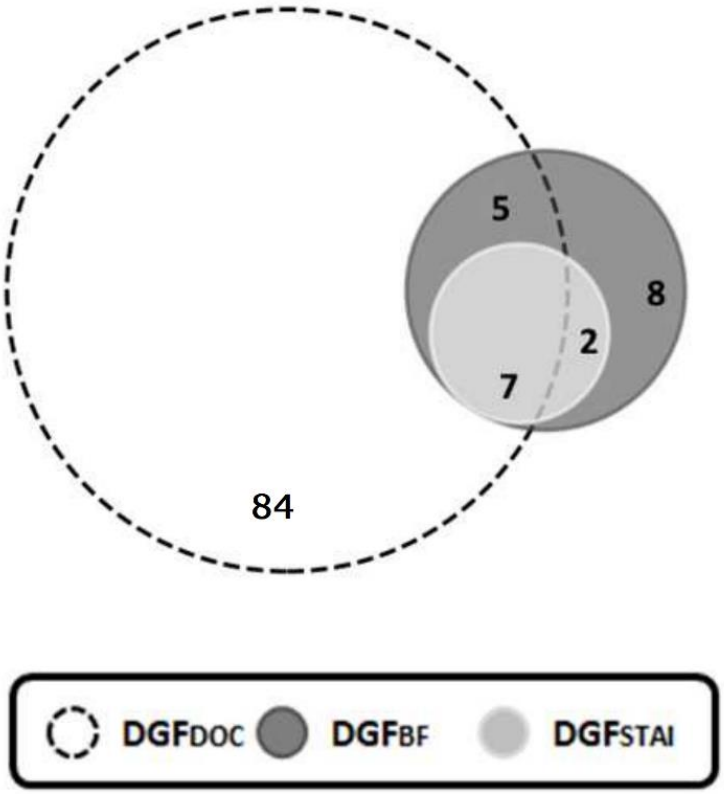


A **B** **C** **D**

D = A

Experimental Results

- ALARic detected 106 distinct GUI failures in 15 analyzed apps

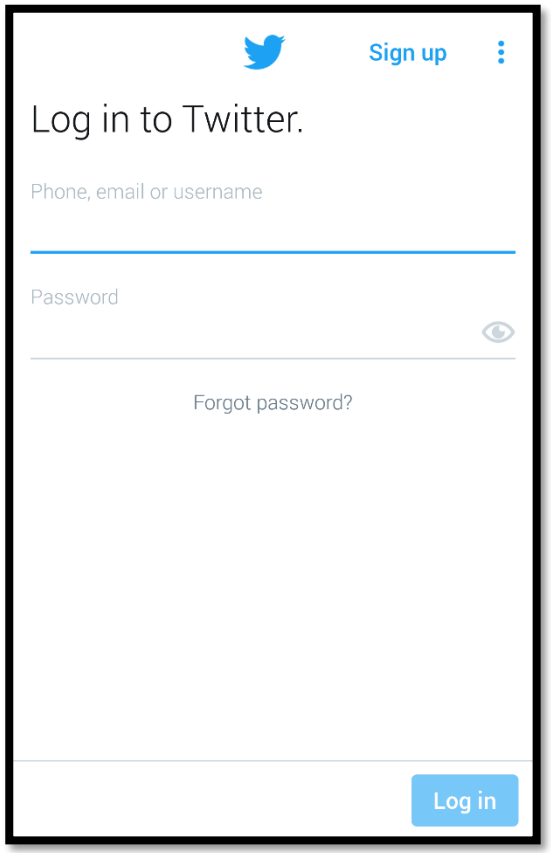


Challenge #2

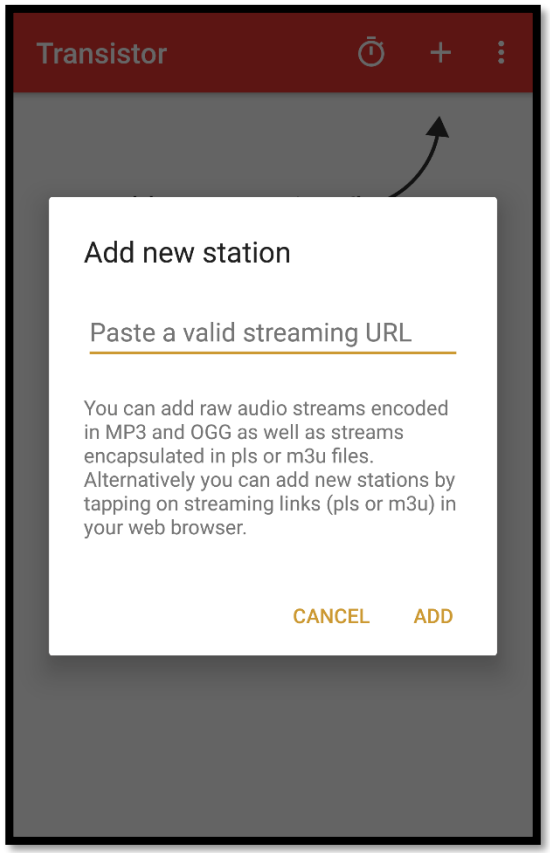
Exploiting app-specific knowledge that only human users can provide

Gate GUI Unlocking

Gate GUIs

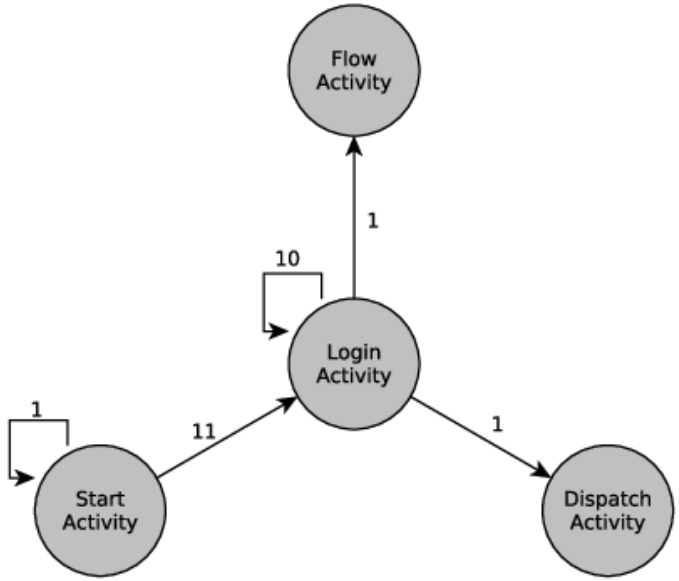


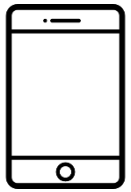
Login Gate GUI



Settings Gate GUI

Gate GUI Locked

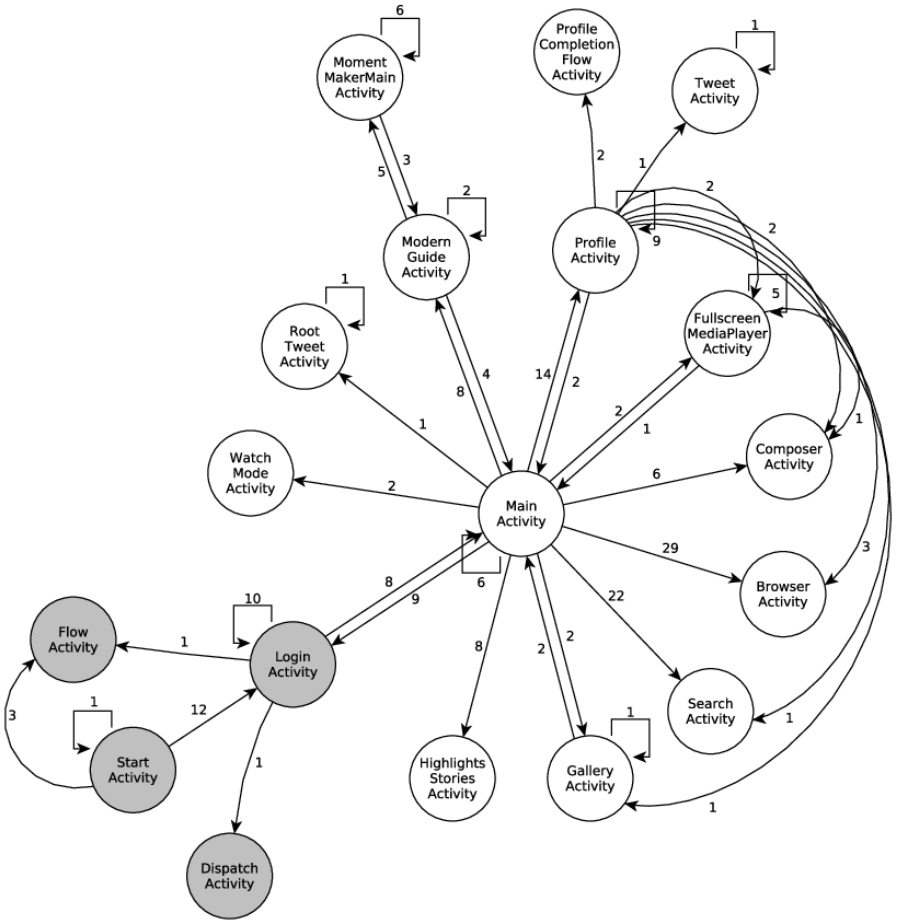


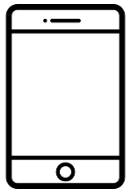


4 Activities

1 Mb

Gate GUI Unlocked

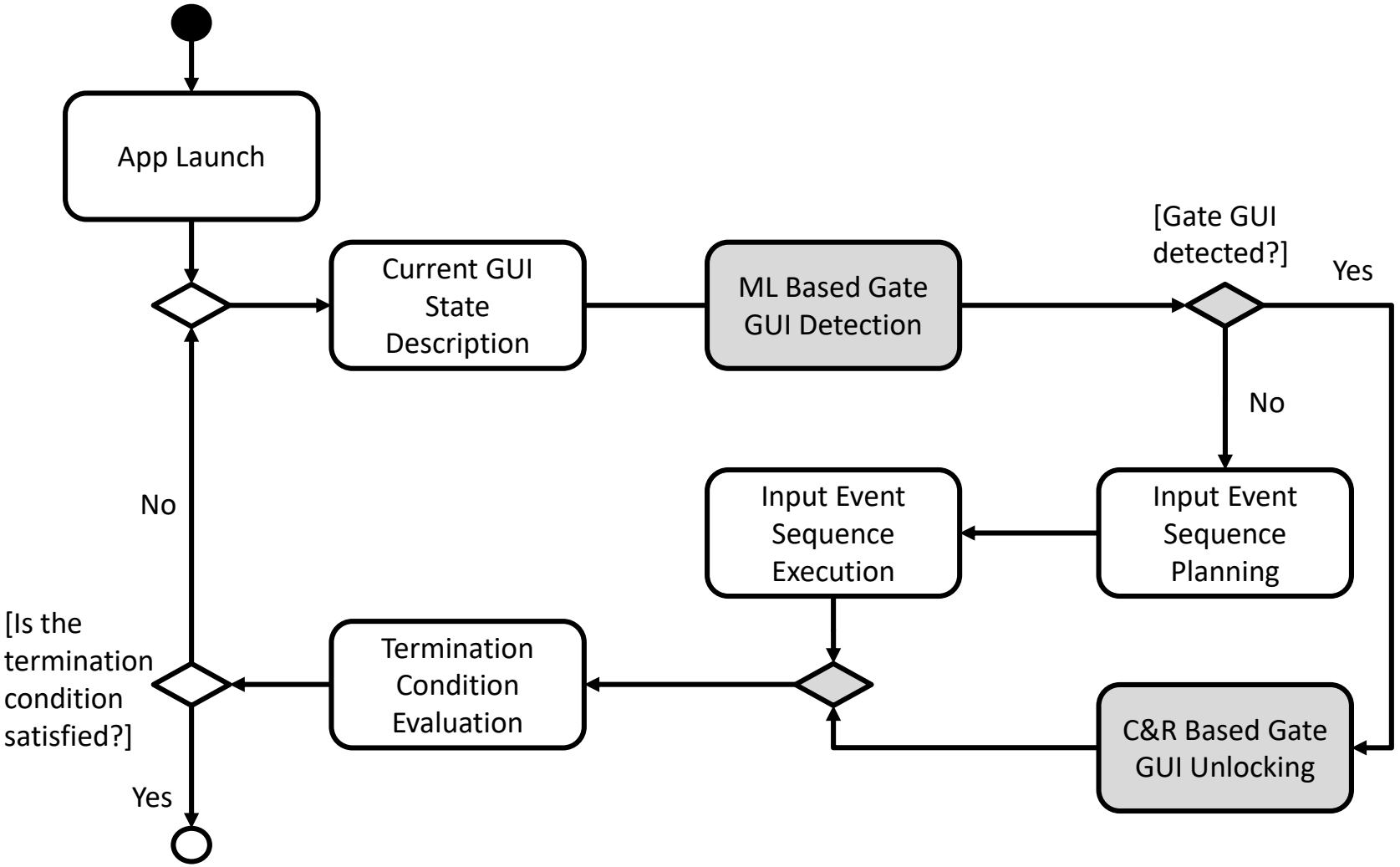




18 Activities

380 Mb

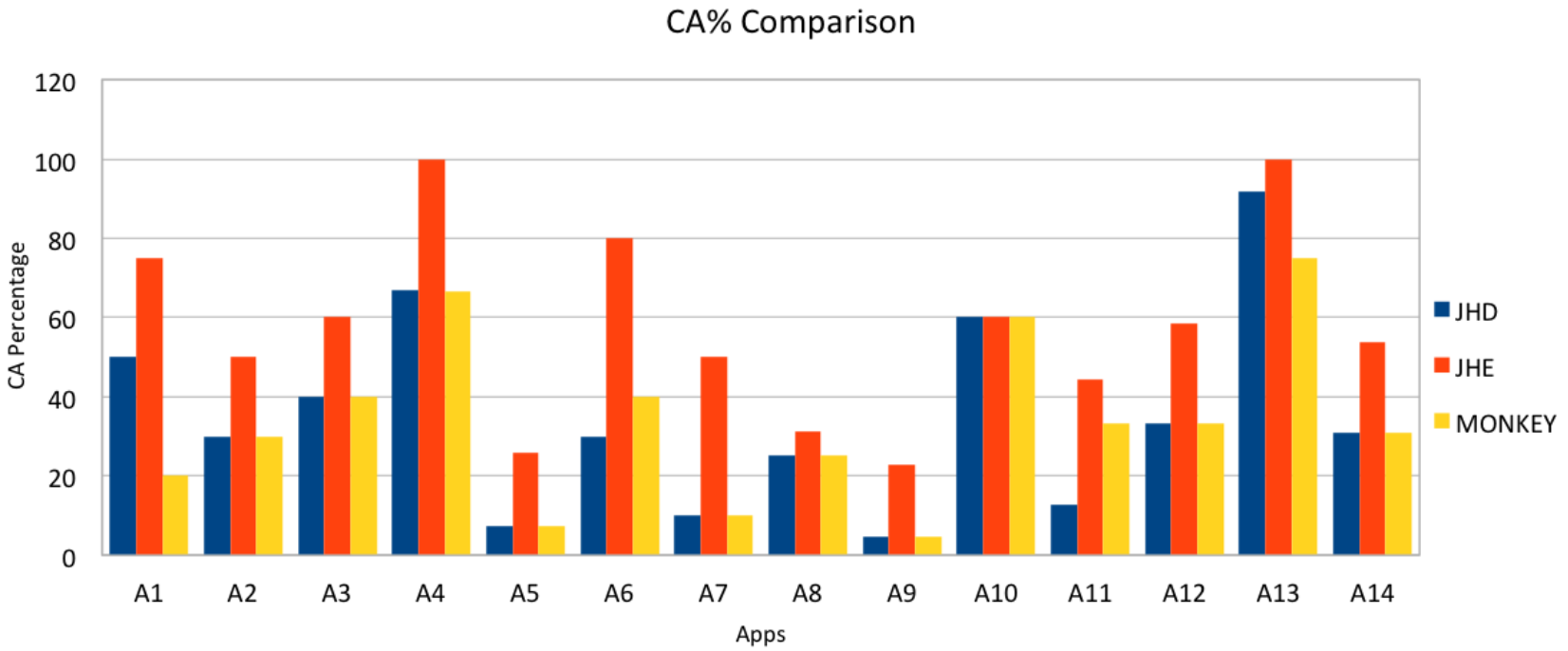
The juGULAR Approach



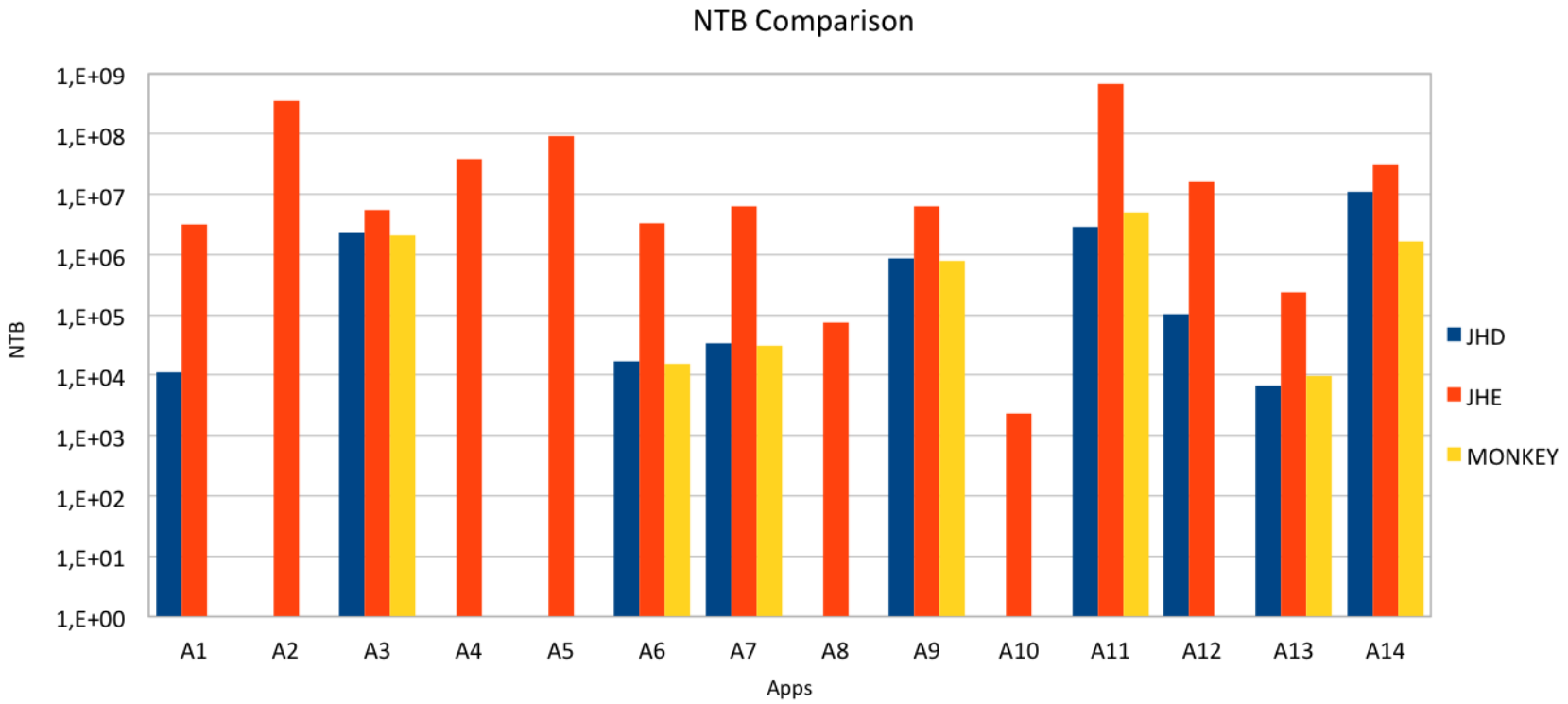
Experimental Evaluation

- Comparison between
  - juGULAR with Hybridization Disabled (**JHD**)
  - juGULAR with Hybridization Enabled (**JHE**)
  - The state-of-the-practice tool, **Monkey**

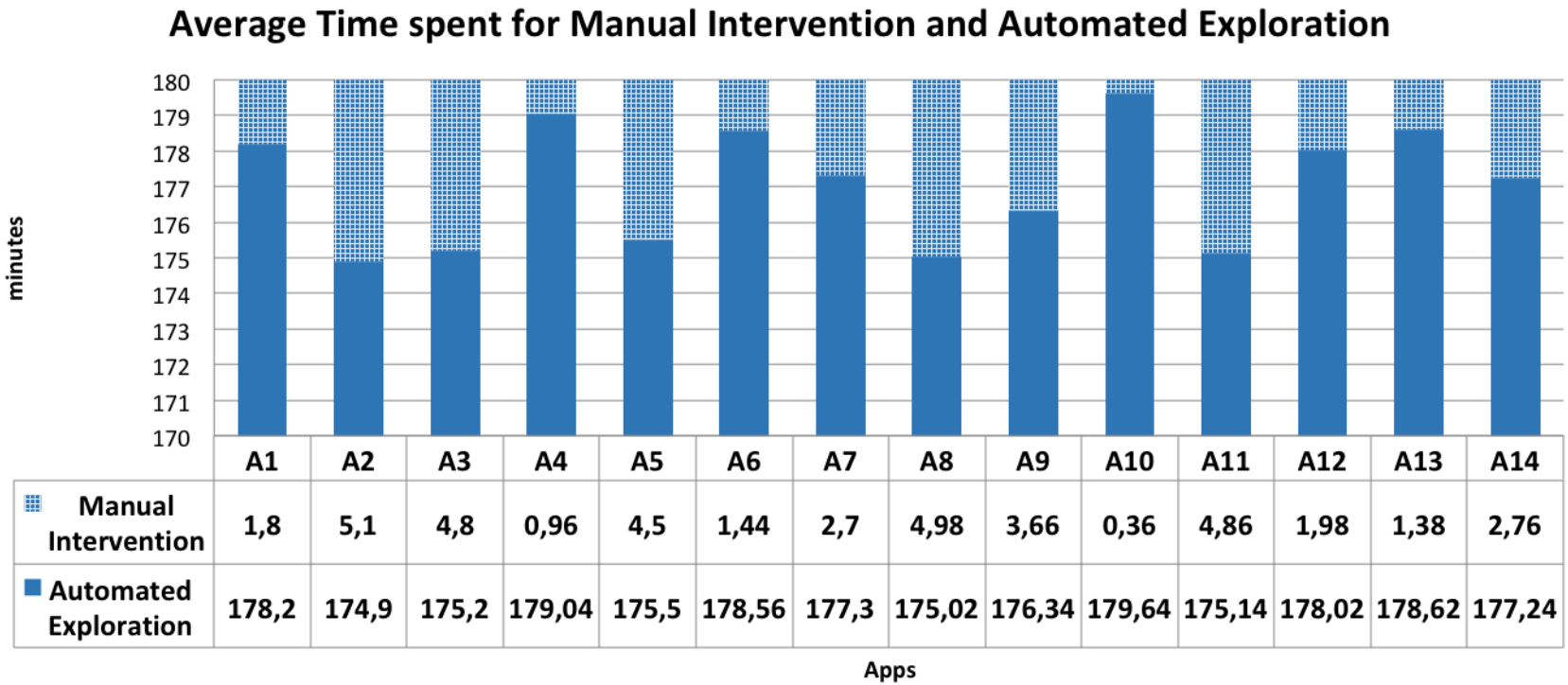
Covered Activities



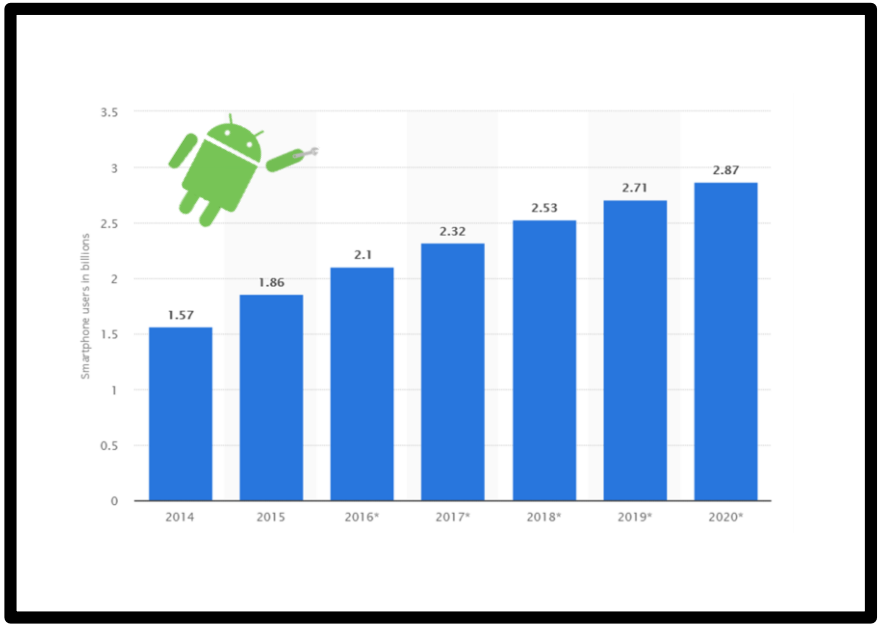
Network Traffic Bytes



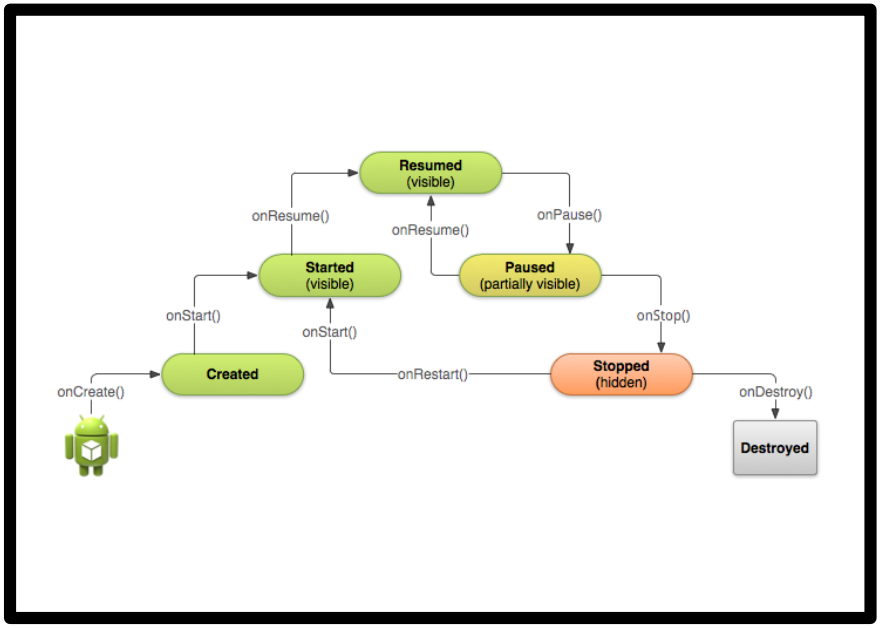
Manual Intervention Percentage

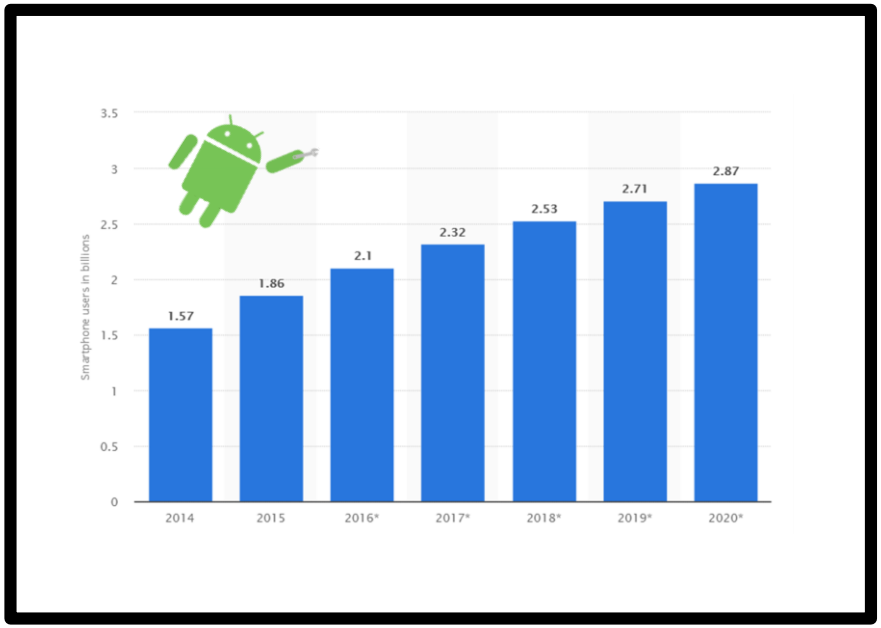


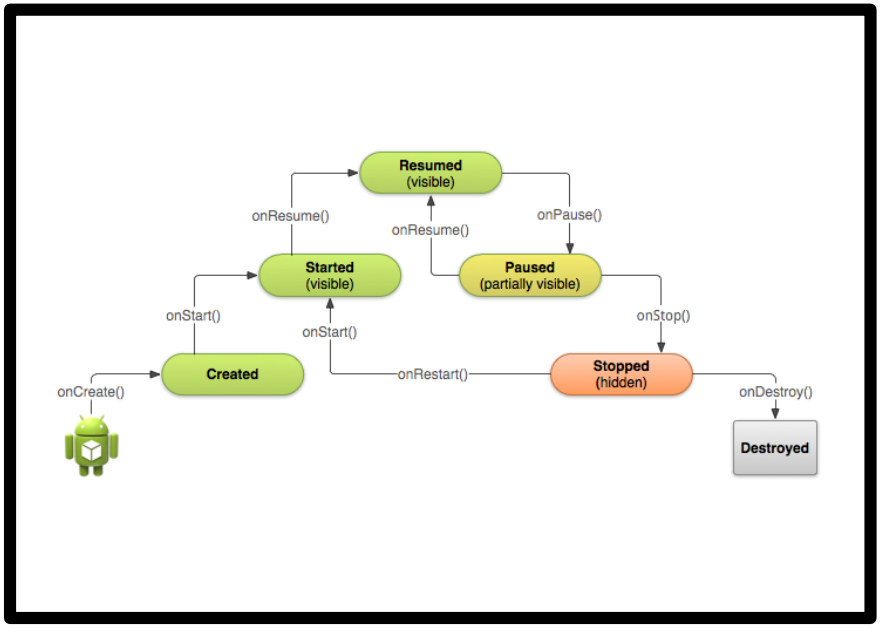
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
Manual Intervention	1,8	5,1	4,8	0,96	4,5	1,44	2,7	4,98	3,66	0,36	4,86	1,98	1,38	2,76
Automated Exploration	178,2	174,9	175,2	179,04	175,5	178,56	177,3	175,02	176,34	179,64	175,14	178,02	178,62	177,24

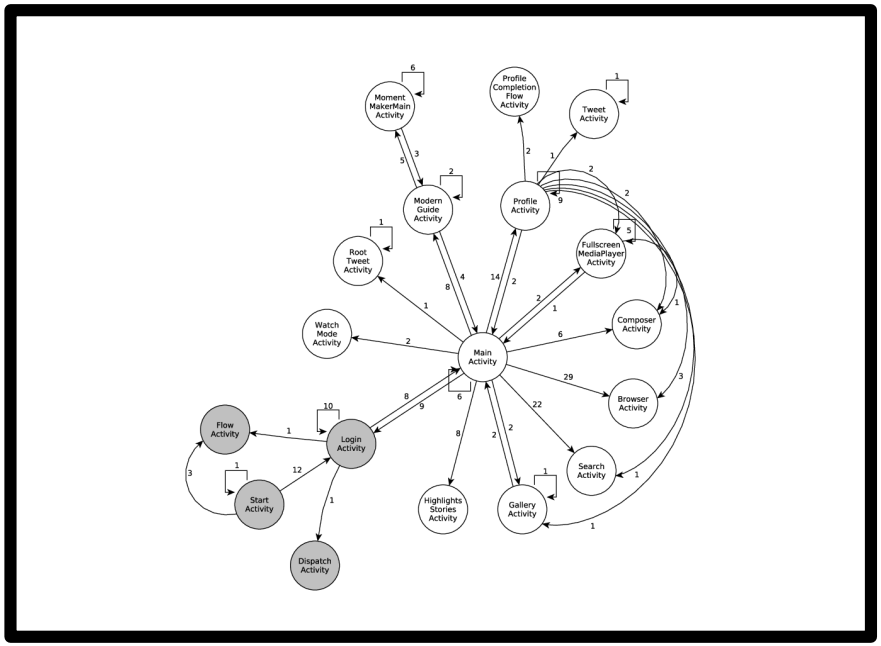

Smartphone users in billions
3.5
3
2.5
2
1.5
1
0.5
0
1.57
1.86
2.1
2.32
2.53
2.71
2.87
2014
2015
2016*
2017*
2018*
2019*
2020*

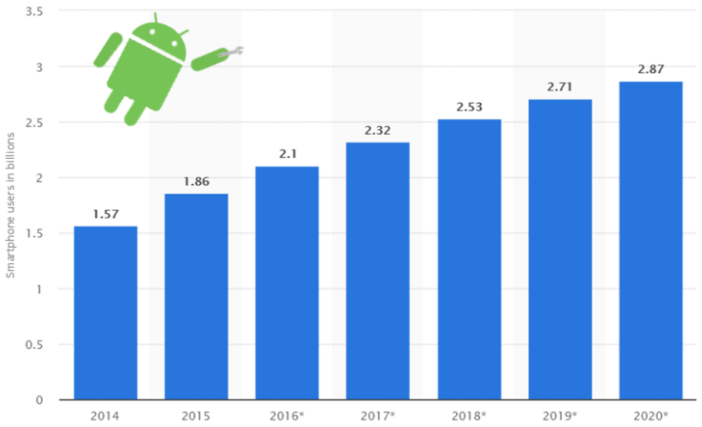

Smartphone users in billions
3.5
3
2.5
2
1.5
1
0.5
0
1.57
1.86
2.1
2.32
2.53
2.71
2.87
2014
2015
2016*
2017*
2018*
2019*
2020*

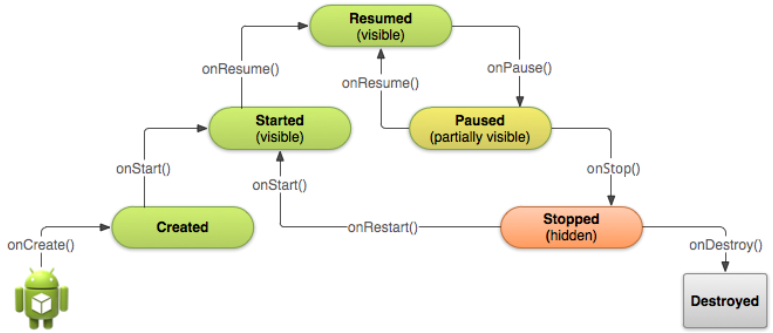

Resumed
(visible)
onResume()
onResume()
onPause()
Started
(visible)
Paused
(partially visible)
onStart()
onStart()
onStop()
Created
onRestart()
Stopped
(hidden)
onCreate()
onDestroy()
Destroyed

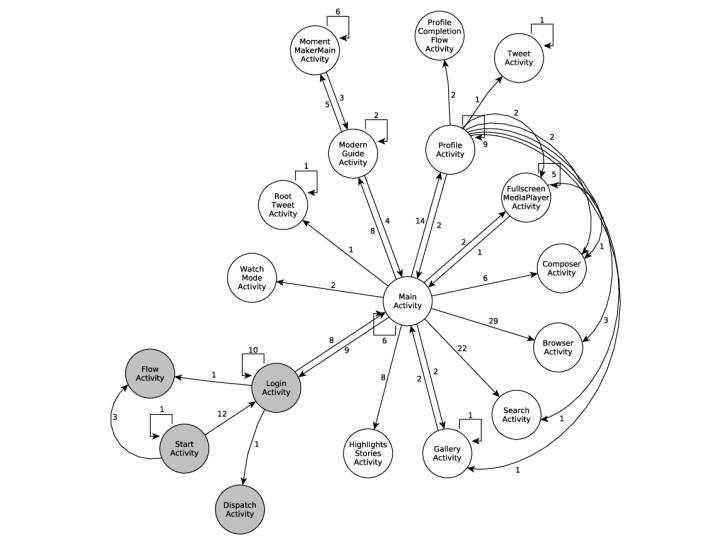

Smartphone users in billions
3.5
3
2.5
2
1.5
1
0.5
0
1.57
1.86
2.1
2.32
2.53
2.71
2.87
2014
2015
2016*
2017*
2018*
2019*
2020*

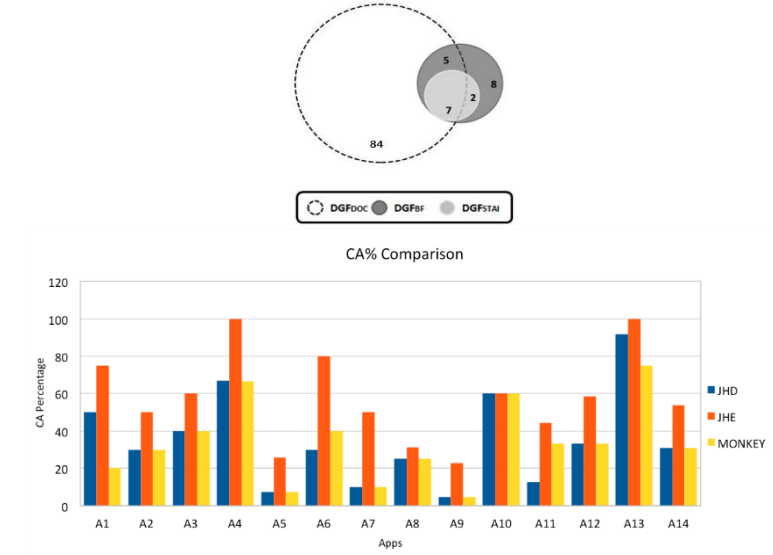

Resumed
(visible)
onResume()
onResume()
onPause()
Started
(visible)
Paused
(partially visible)
onStart()
onStart()
onStop()
Created
onRestart()
Stopped
(hidden)
onCreate()
onDestroy()
Destroyed


Moment MakerMain Activity
Profile Completion Flow Activity
Tweet Activity
Modern Guide Activity
Profile Activity
Root Tweet Activity
Fullscreen MediaPlayer Activity
Watch Mode Activity
Composer Activity
Main Activity
Browser Activity
Flow Activity
Login Activity
Search Activity
Start Activity
Highlights Stories Activity
Gallery Activity
Dispatch Activity


Smartphone users in billions
3.5
3
2.5
2
1.5
1
0.5
0
1.57
1.86
2.1
2.32
2.53
2.71
2.87
2014
2015
2016*
2017*
2018*
2019*
2020*


Resumed
(visible)
onResume()
onPause()
onResume()
Started
(visible)
Paused
(partially visible)
onStart()
onStop()
onStart()
Created
onRestart()
Stopped
(hidden)
onCreate()
onDestroy()
Destroyed


Moment MakerMain Activity
Profile Completion Flow Activity
Tweet Activity
Modern Guide Activity
Profile Activity
Root Tweet Activity
Fullscreen MediaPlayer Activity
Watch Mode Activity
Composer Activity
Main Activity
Browser Activity
Flow Activity
Login Activity
Search Activity
Start Activity
Highlights Stories Activity
Gallery Activity
Dispatch Activity


5
8
2
7
84
DGFDOC
DGFBF
DGFSTAI
CA% Comparison
CA Percentage
120
100
80
60
40
20
0
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
Apps
JHD
JHE
MONKEY

Products (1/2)

- **Journal Papers:**
  - D Amalfitano, V Riccio, ACR Paiva, and AR Fasolino (2018). **Why does the orientation change mess up my Android application? From GUI failures to code faults.** *Software Testing, Verification and Reliability, 28(1)*. Wiley. doi:10.1002/stvr.1654.
    - In collaboration with the University of Porto
    - **Wiley's #Top20Article:** Amongst articles published by Wiley between July 2016 and June 2018, this article received some of the highest downloads in the 12-months post online publication
  - D Amalfitano, V Riccio, N Amatucci, V De Simone, and AR Fasolino (2018) **Combining Automated GUI Exploration of Android apps with Capture and Replay through Machine Learning.** *Information and Software Technology, 105(1)*. Elsevier. doi:10.1016/j.infsof.2018.08.007.

Products (2/2)

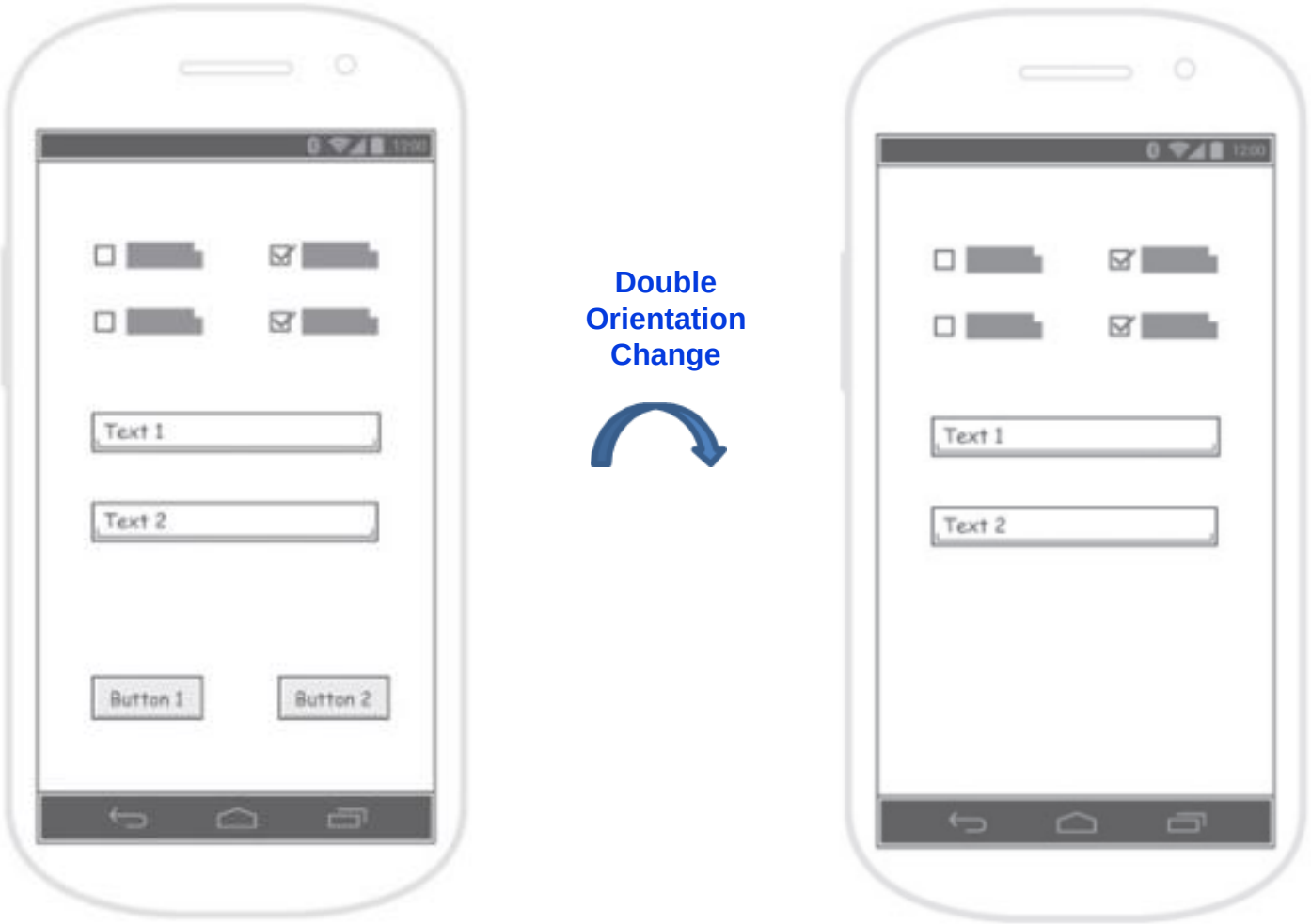
- **Conference Papers:**
  - D Amalfitano, V De Simone, A R Fasolino and V Riccio (2015). **Comparing Model Coverage and Code Coverage in Model Driven Testing: An Exploratory Study**, In *Proceedings of the 30th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW)*, Lincoln, NE, 2015, pp. 70-73. doi: 10.1109/ASEW.2015.18
  - Domenico Amalfitano, Nicola Amatucci, Vincenzo De Simone, Vincenzo Riccio, and Fasolino Anna Rita (2017). **Towards a Thing-In-the-Loop approach for the Verification and Validation of IoT systems.** In *Proceedings of the 1st ACM Workshop on the Internet of Safe Things (SafeThings'17)*, Rasit Eskicioglu (Ed.). ACM, New York, NY, USA, pp. 57-63. doi: 10.1145/3137003.3137007
  - Vincenzo Riccio, Domenico Amalfitano, and Anna Rita Fasolino (2018). **Is This the Lifecycle We Really Want? An Automated Black-Box Testing Approach for Android Activities.** *The Joint Workshop of 4th Workshop on UI Test Automation and 8th Workshop on TESting Techniques for eventBasED Software (INTUITESTBEDS 2018)*. ACM (In press).

Extra Slides

Motivating Example: Crash

- A crash occurs when an app stops functioning properly and exits unexpectedly

Missing GUI Failure



Extra GUI Failure

Wrong GUI Failure

ALARic Description

- **ALARic (Activity Lifecycle Android Ripper)**, a novel fully automated Black-Box Event-based testing technique to detect issues tied to the Activity lifecycle
- It combines:
  - The traditional testing approaches based on dynamic app exploration
  - A strategy that systematically exercises the Activity lifecycle on each GUI state encountered during the exploration
- It relies on:
  - **Lifecycle Event Sequences**, mobile-specific events able to exercise the Activity lifecycle
  - Testing oracles to detect crashes and GUI failures tied to the Activity lifecycle

Experimental Evaluation

- **GOAL:** Evaluate the ability of ALARic to automatically detect crashes and GUI failures tied to the Activity lifecycle
  - **RQ1:** How effective is the ALARic tool in detecting issues tied to the Activity lifecycle in real Android apps?
  - **RQ2:** How does the effectiveness of the ALARic tool in detecting crashes tied to the Activity lifecycle in real Android apps compare to the state-of-the-practice tool, Monkey?

Objects

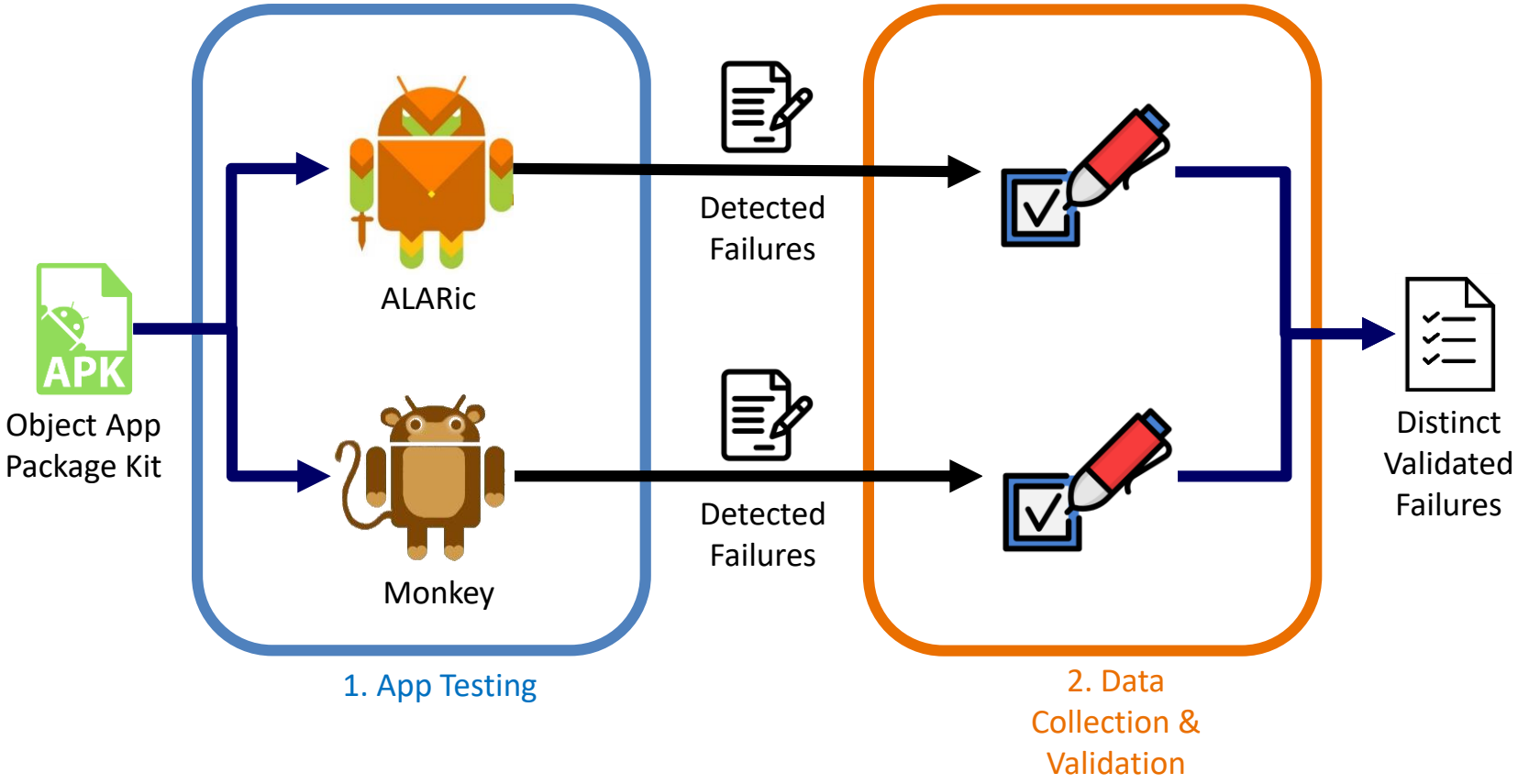
- 15 apps that are distributed by Google Play Store whose source code is available in the F-Droid repository

ID	App	Version	Activities
A1	A Time Tracker	0.21	5
A2	Port Knocker	1.0.9	6
A3	Who Has My Stuff?	1.0.27	4
A4	Agram	1.4.1	5
A5	Alarm Klock	1.9	5
A6	Padland	1.3	10
A7	Syncthing	0.9.1	12
A8	Anecdote	1.1.2	3
A9	Amaze File Manager	3.1.2 RC4	5
A10	Google Authenticator	2.21	5
A11	BeeCount	2.3.9	8
A12	FOSDEM companion	1.4.6	8
A13	Periodical	0.30	6
A14	Taskbar	3.0.2	23
A15	SpaRSS	1.11.8	8

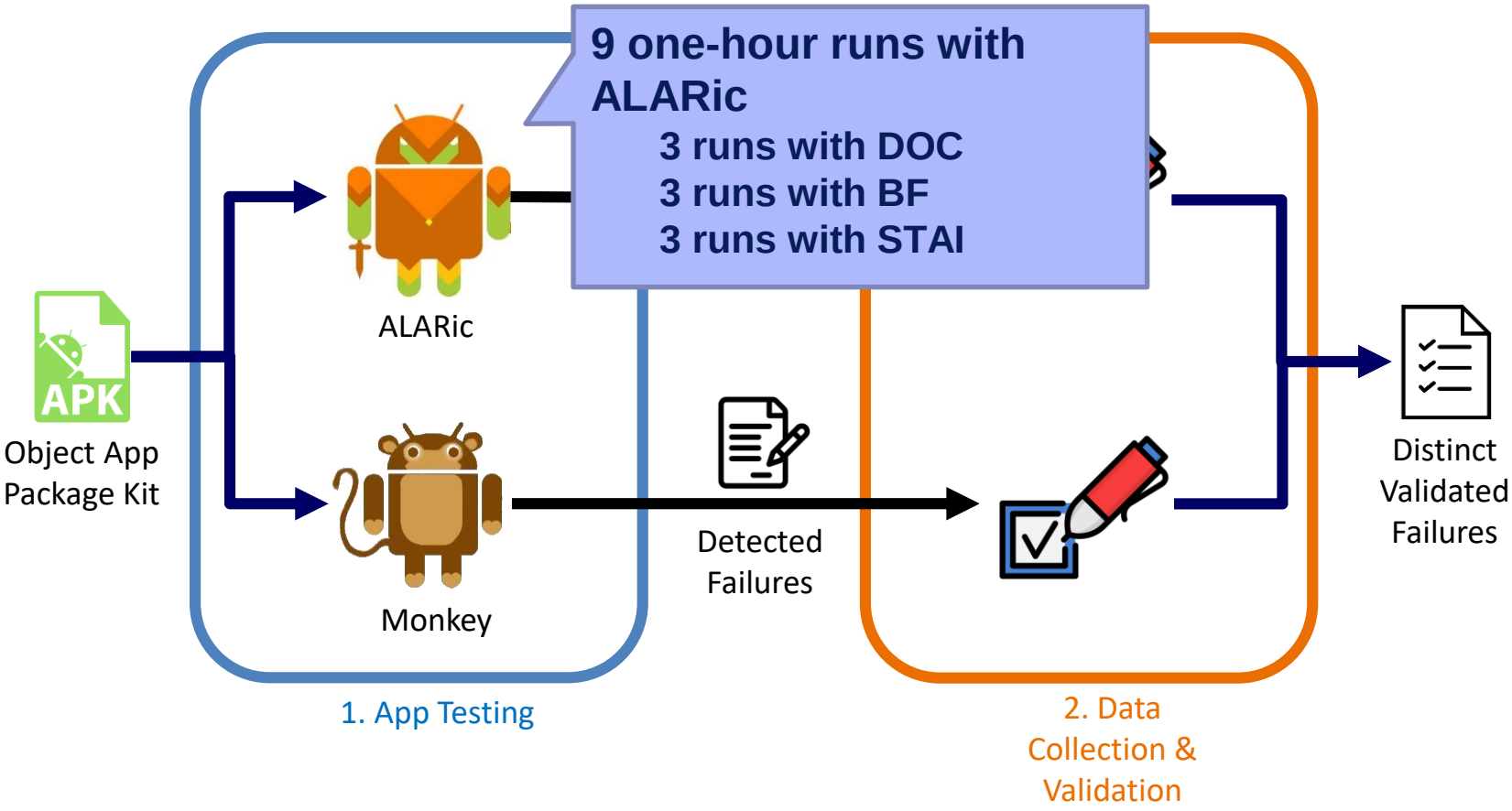
Metrics

- To evaluate the effectiveness of ALARic in detecting GUI failures:
  - **#DGFDOC** number of distinct GUI Failures triggered by DOC
  - **#DGFBF** number of distinct GUI Failures triggered by BF
  - **#DGFSTAI** number of distinct GUI Failures triggered by STAI
  - **#DGFTOTAL** number of distinct GUI Failures triggered by the DOC, BF, STAI
- To evaluate the effectiveness of both the tools in finding Crashes:
  - **#DCDOC** number of distinct crashes triggered by DOC
  - **#DCBF** number of distinct crashes triggered by BF
  - **#DCSTAI** number of distinct crashes triggered by STAI
  - **#DCTOTAL** number of distinct crashes triggered by the DOC, BF, STAI

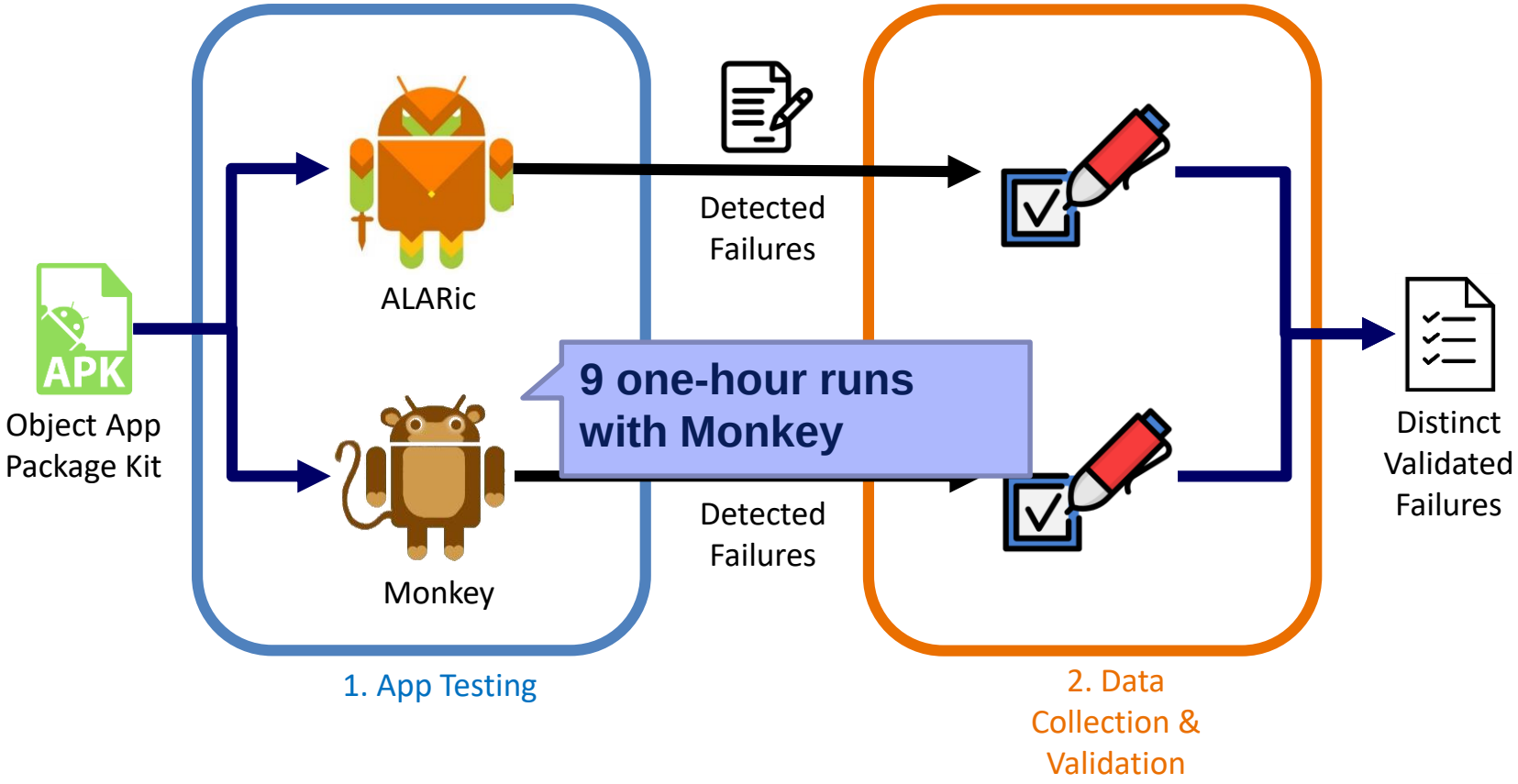
Experimental Procedure



Experimental Procedure

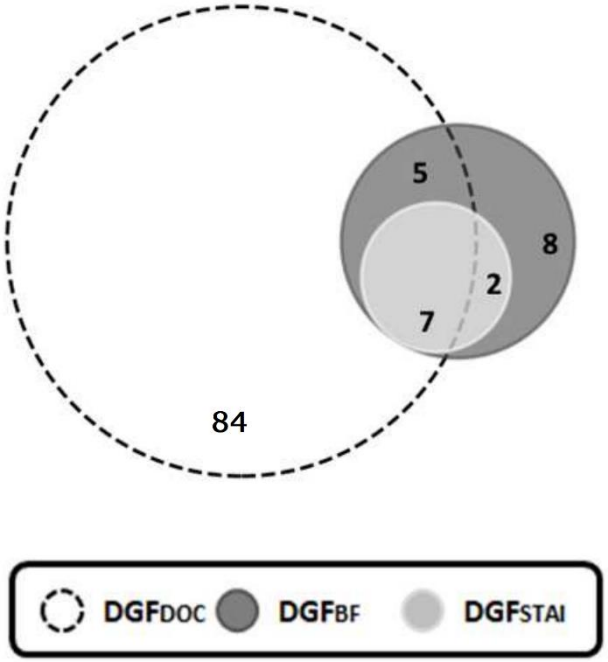


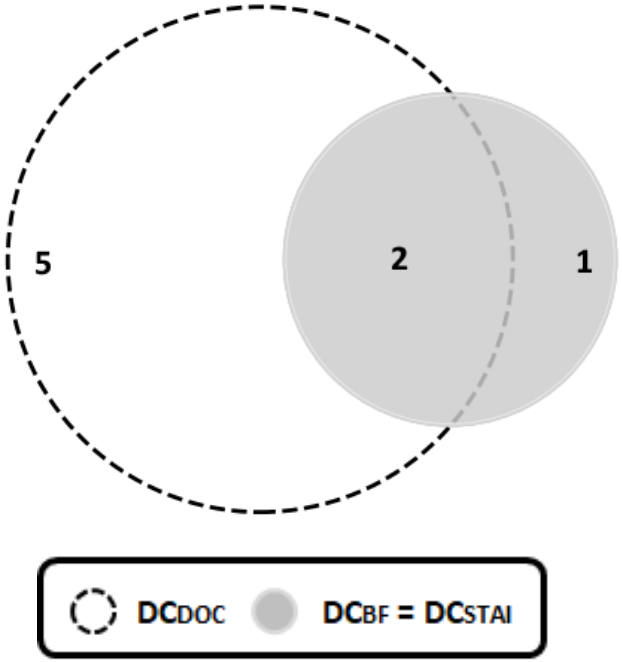
Experimental Procedure



Experimental Results: RQ1

- ALARic detected 106 distinct GUI failures and 8 crashes tied to the Activity lifecycle in the 15 analyzed apps





Experimental Results: RQ2

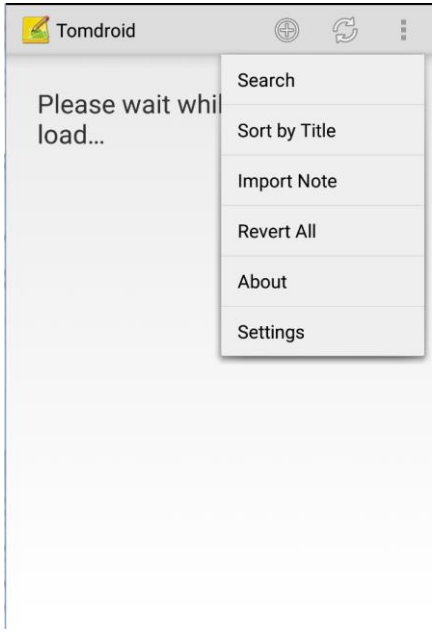
- ALARic outperformed Monkey in the ability to detect issues tied to the Activity lifecycle
  - In total ALARic triggered more crashes than Monkey
  - Monkey seeds events that exercise the Activity lifecycle, e.g. orientation changes, back button press, but it applies them without a proper strategy

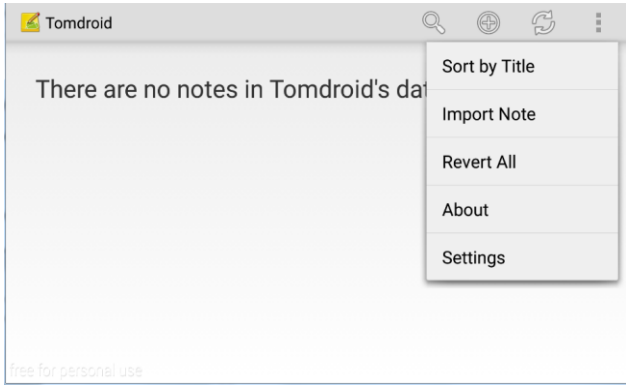
App	#DCALARic	#DCMonkey
A4	1	1
A6	1	0
A7	1	0
A9	2	0
A11	1	0
A15	2	1
Total	**8**	**2**

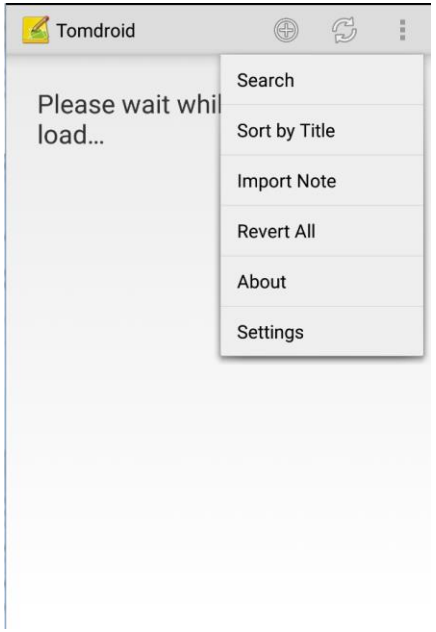
Lesson Learned

- The debugging activity we performed in the failure validation step showed us that the faults causing the failures were mostly located outside the code that overrides the lifecycle callback methods
  - Testers should look for faults that may affect the lifecycle of the Activities also outside the methods that override the lifecycle callbacks
  - Developers should correctly use the Android framework components since they may cause inconsistencies in the app behavior at runtime when Lifecycle Event Sequences occur

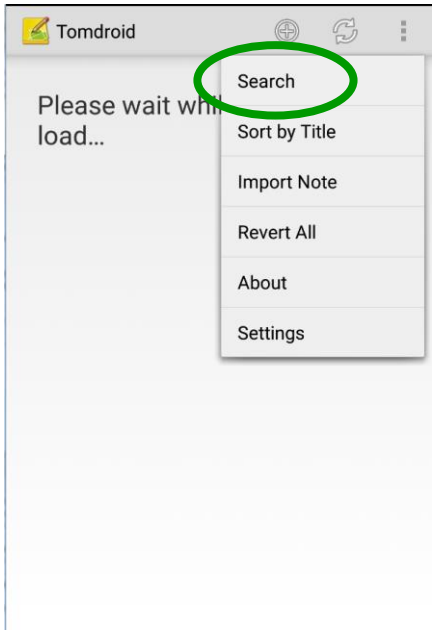
Lifecycle Event Sequences: DOC

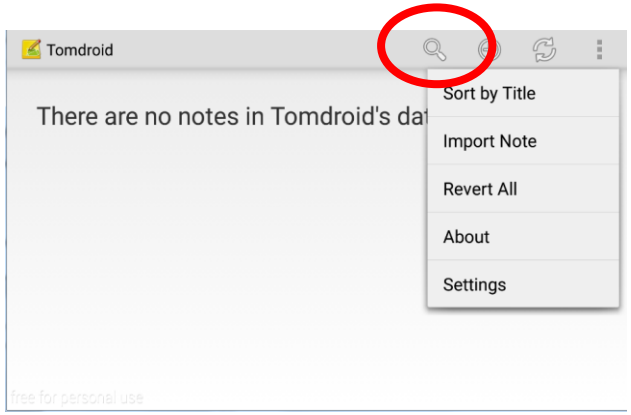


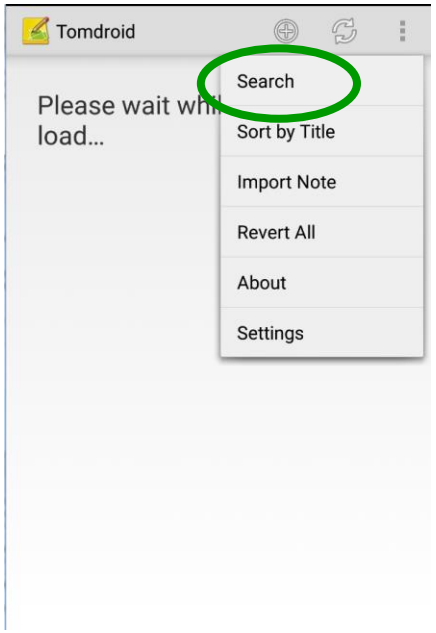




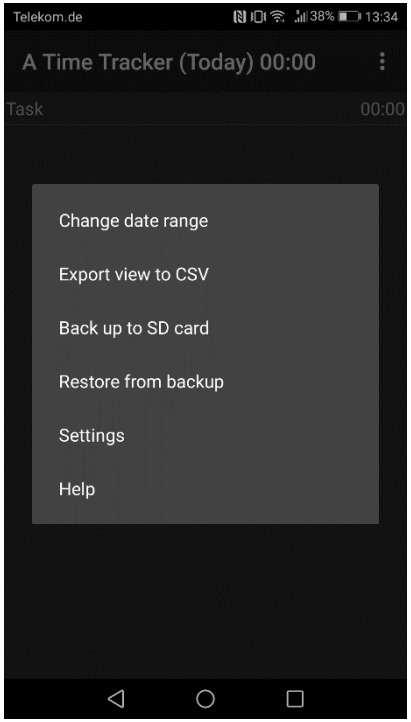
Lifecycle Event Sequences: DOC

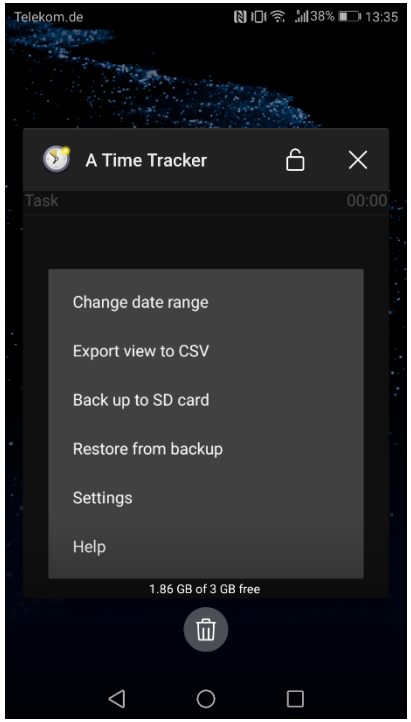


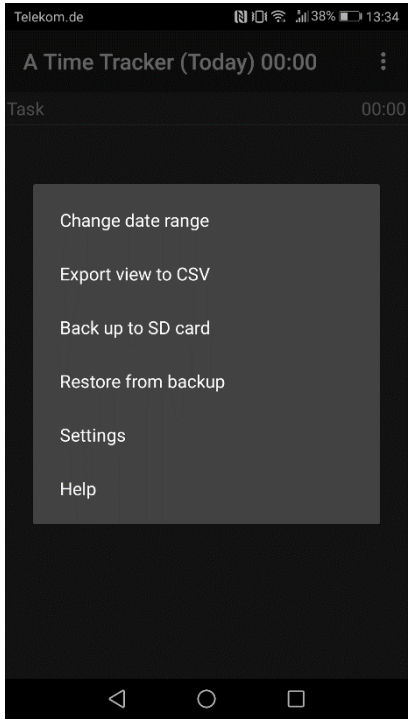




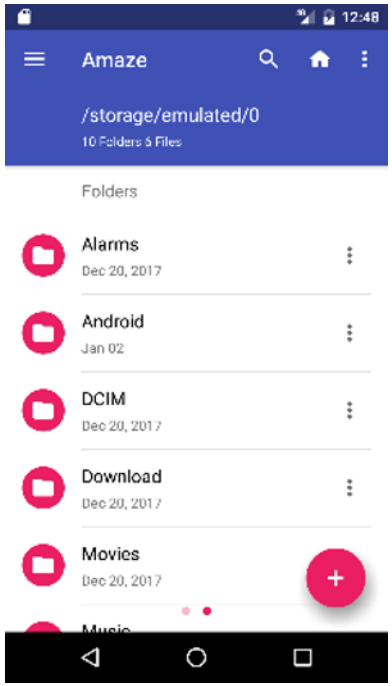
Lifecycle Event Sequences: BF

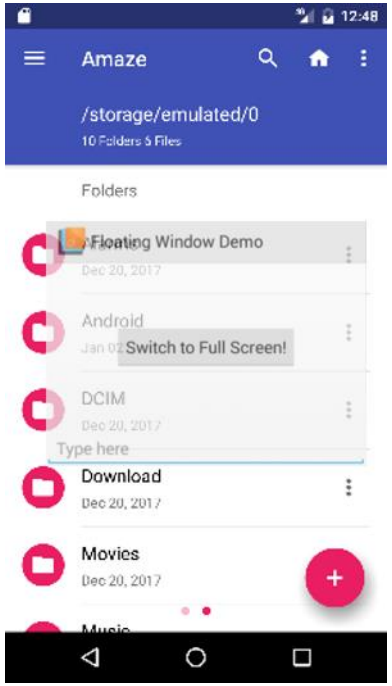


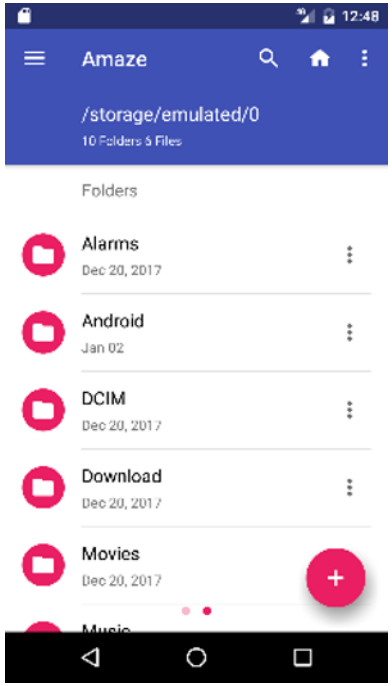




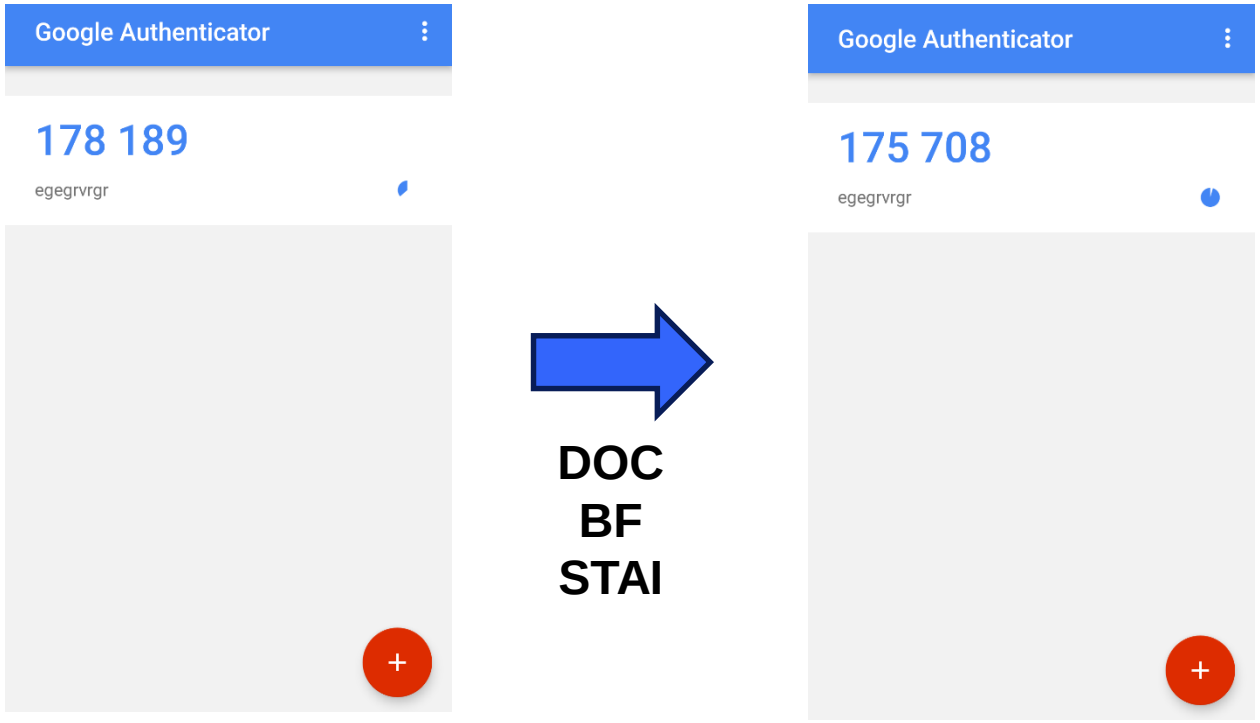
Lifecycle Event Sequences: STAI



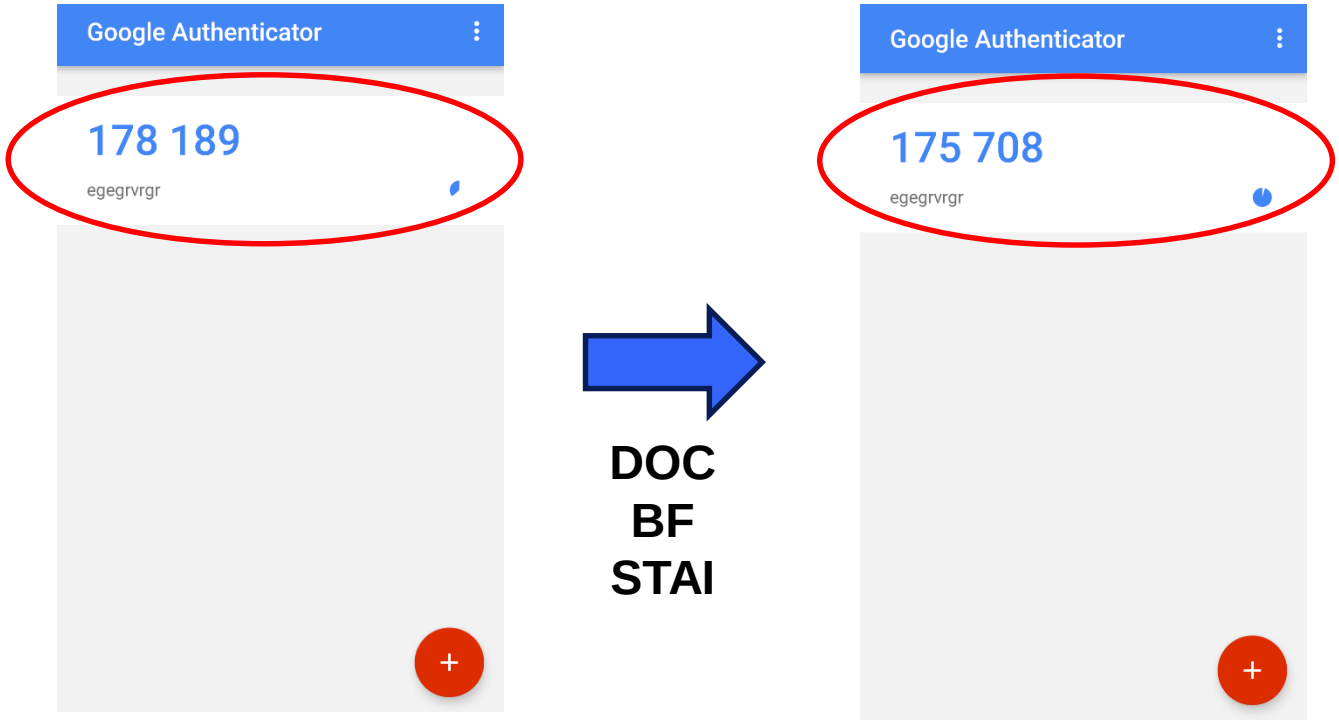




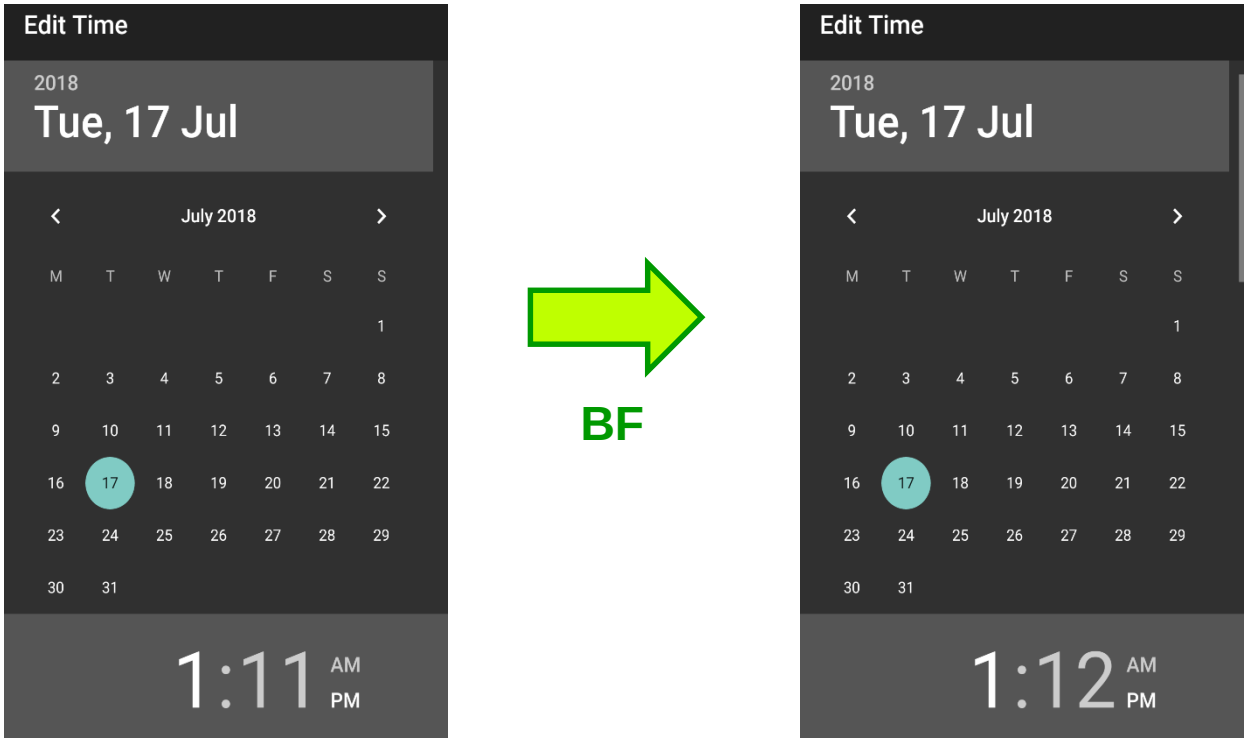
False Positive Example #1



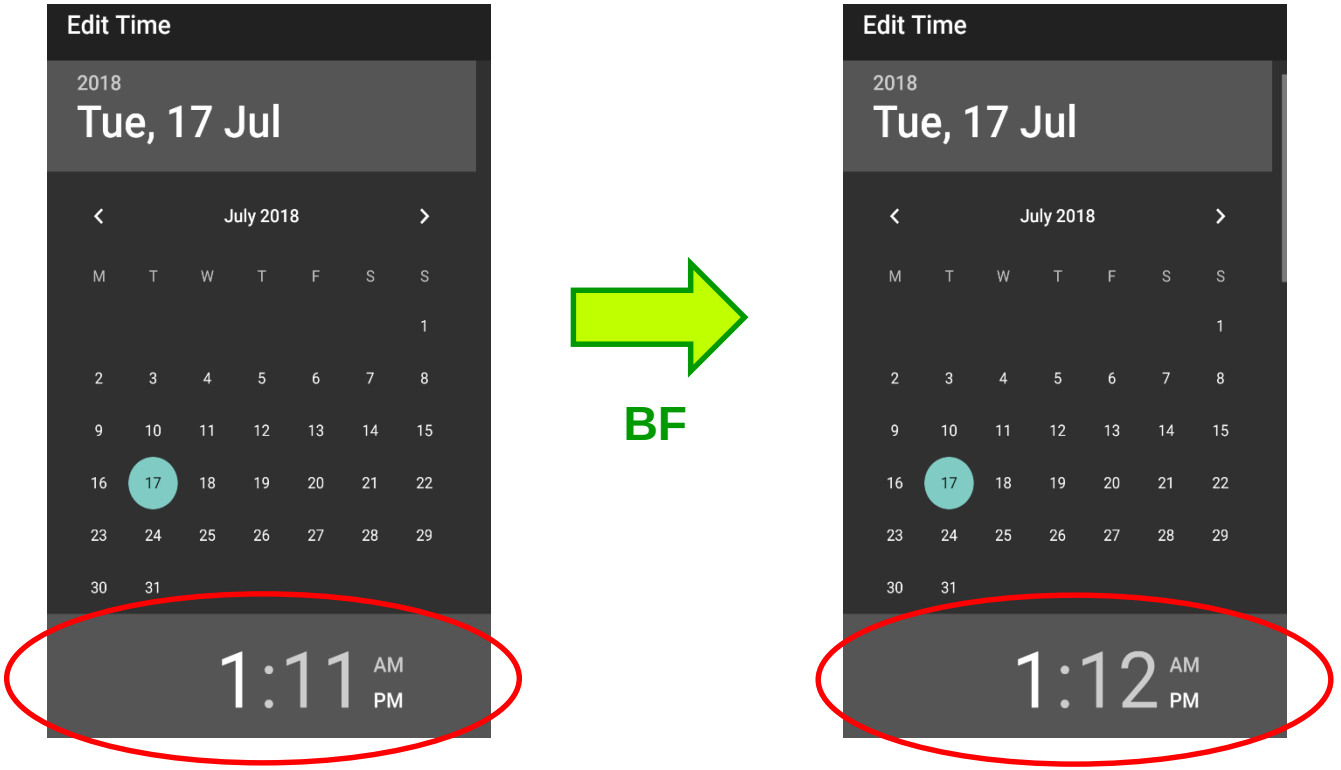
False Positive Example #1



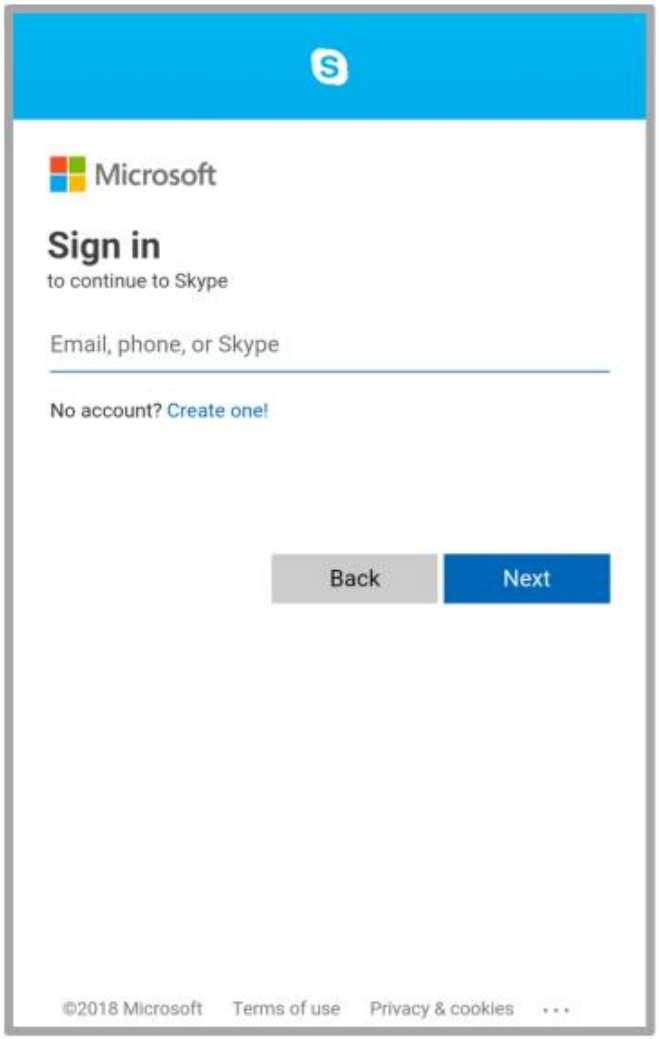
False Positive Example #2



False Positive Example #2



GUI XML Description

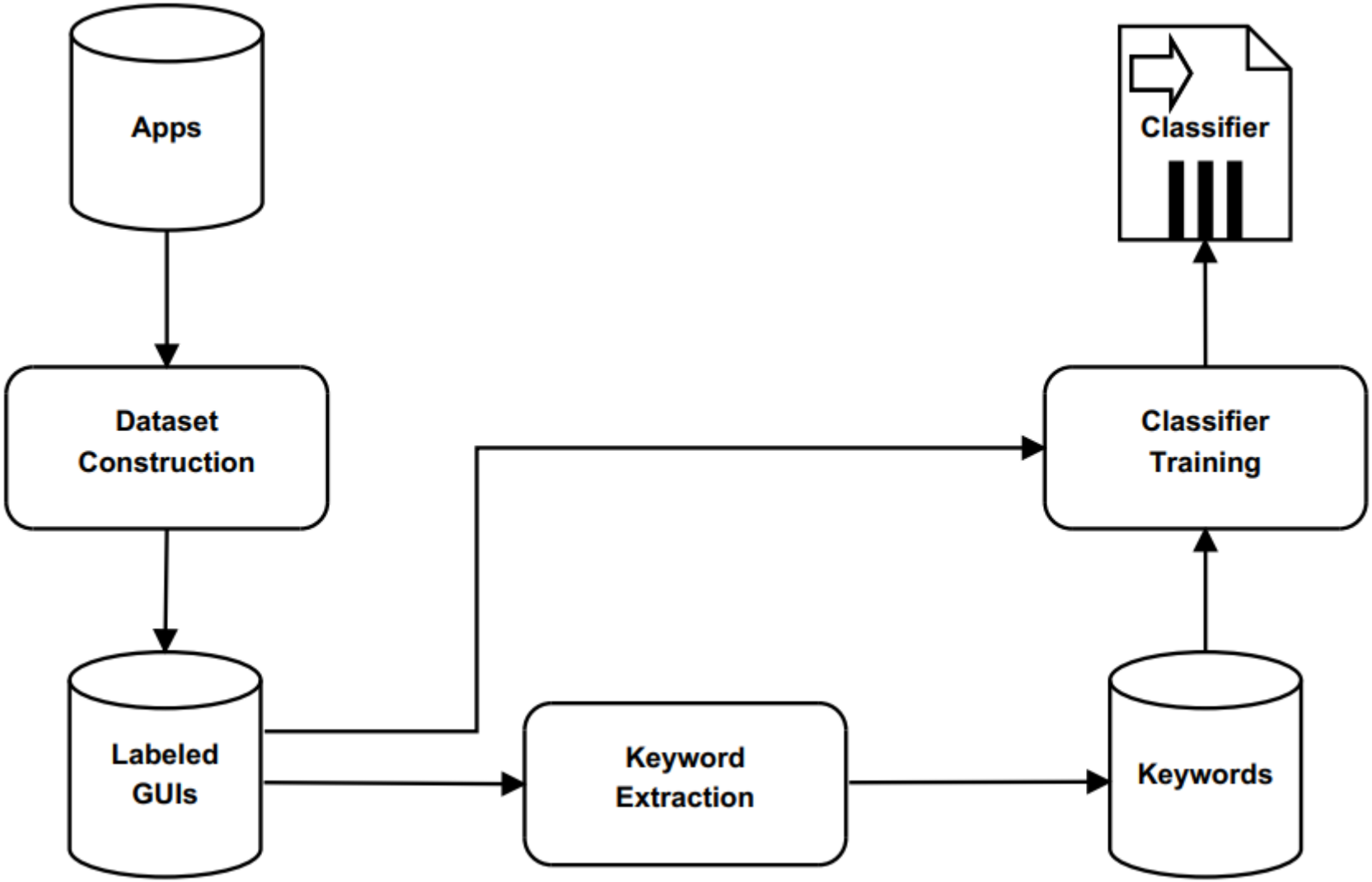



```
▼<node index="0" text="" resource-id="" class="android.webkit.WebView"
 package="com.skype.raider" content-desc="Sign in to Skype"
 checkable="false" checked="false" clickable="false" enabled="true"
 focusable="true" focused="true" scrollable="true" long-clickable="false"
 password="false" selected="false" bounds="[0,244][1080,1794]">
  ▼<node index="0" text="" resource-id="" class="android.view.View"
   package="com.skype.raider" content-desc="" checkable="false"
   checked="false" clickable="false" enabled="true" focusable="false"
   focused="false" scrollable="false" long-clickable="false"
   password="false" selected="false" bounds="[0,244][1080,1794]">
    <node index="0" text="" resource-id="" class="android.view.View"
    package="com.skype.raider" content-desc="" checkable="false"
    checked="false" clickable="false" enabled="true" focusable="false"
    focused="false" scrollable="false" long-clickable="false"
    password="false" selected="false" bounds="[63,307][1018,372]"/>
   ▼<node index="1" text="" resource-id="" class="android.view.View"
    package="com.skype.raider" content-desc="" checkable="false"
    checked="false" clickable="false" enabled="true" focusable="false"
    focused="false" scrollable="false" long-clickable="false"
    password="false" selected="false" bounds="[0,422][1080,1794]">
     <node index="0" text="" resource-id="" class="android.view.View"
     package="com.skype.raider" content-desc="Sign in to continue to
     Skype " checkable="false" checked="false" clickable="true"
     enabled="true" focusable="false" focused="false" scrollable="false"
     long-clickable="false" password="false" selected="false" bounds="
     [57,422][1023,551]"/>
     <node index="1" text="" resource-id="" class="android.view.View"
     package="com.skype.raider" content-desc="" checkable="false"
     checked="false" clickable="false" enabled="true" focusable="false"
     focused="false" scrollable="false" long-clickable="false"
     password="false" selected="false" bounds="[0,0][0,0]"/>
     <node index="2" text="" resource-id=""
     class="android.widget.EditText" package="com.skype.raider" content-
     desc="Enter your email, phone, or Skype." checkable="false"
     checked="false" clickable="true" enabled="true" focusable="true"
     focused="false" scrollable="false" long-clickable="false"
     password="false" selected="false" bounds="[63,580][1018,677]"/>
```

GUI Textual Information Content

Keyword	(a)	(b)	(c)
forgot	✓	✓	✗
login	✓	✗	✗
password	✓	✓	✗
facebook	✓	✓	✗

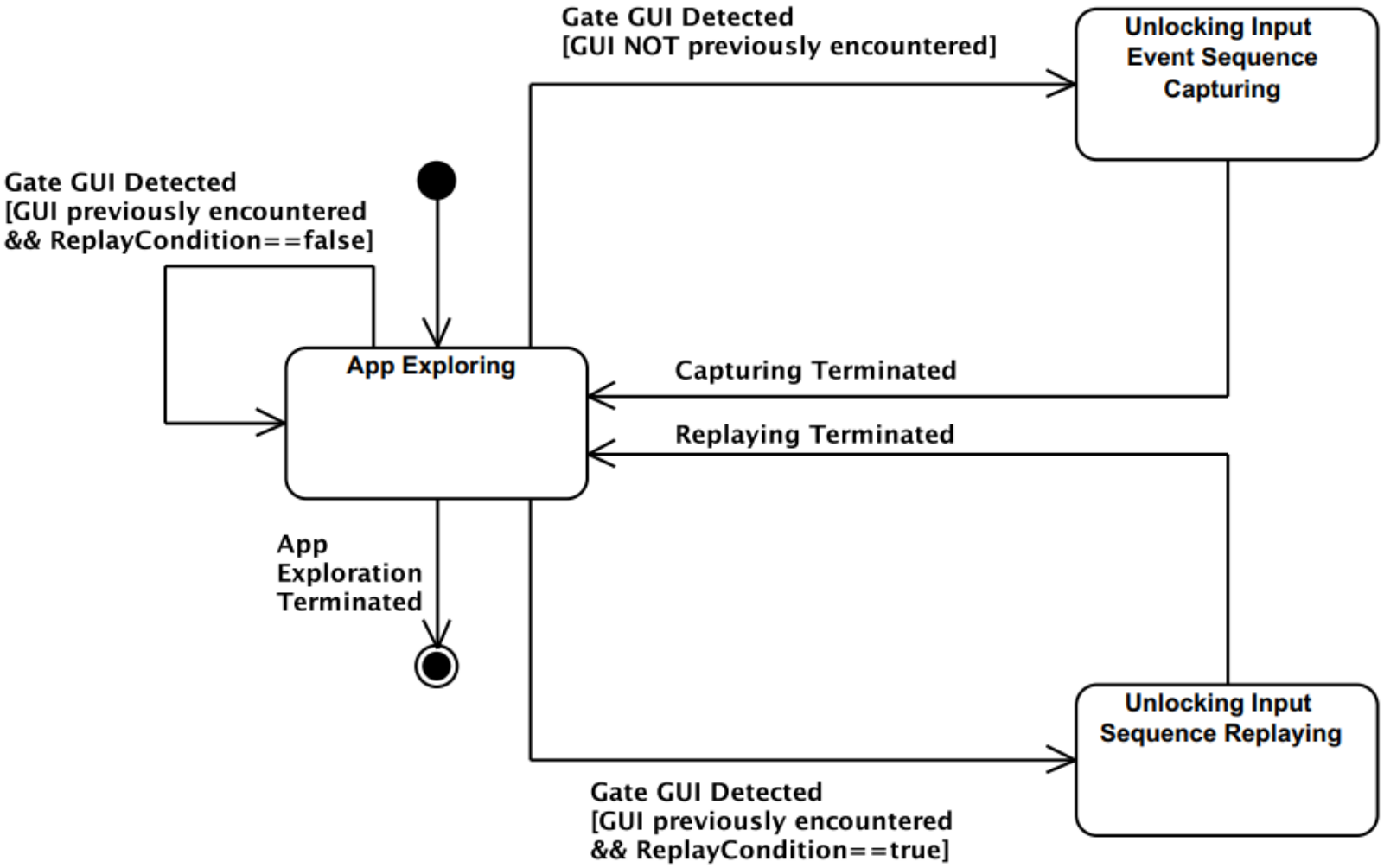
ML-based Classifier Training Process



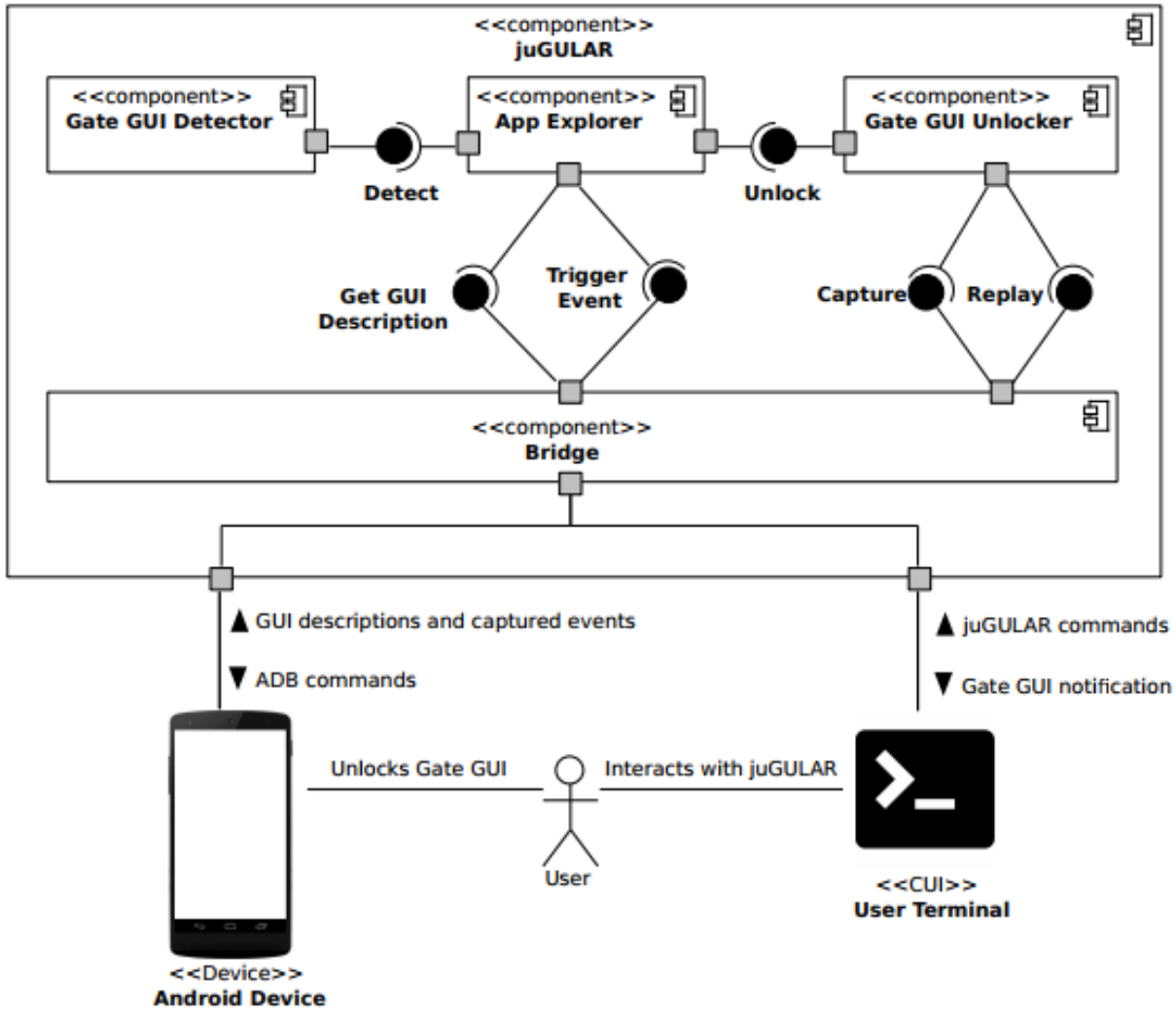
Gate GUI Classifiers' Performance

	Login Gate GUI	Network Settings Gate GUI
Precision	0.814	0.751
Recall	0.807	0.900
F-measure	0.807	0.813

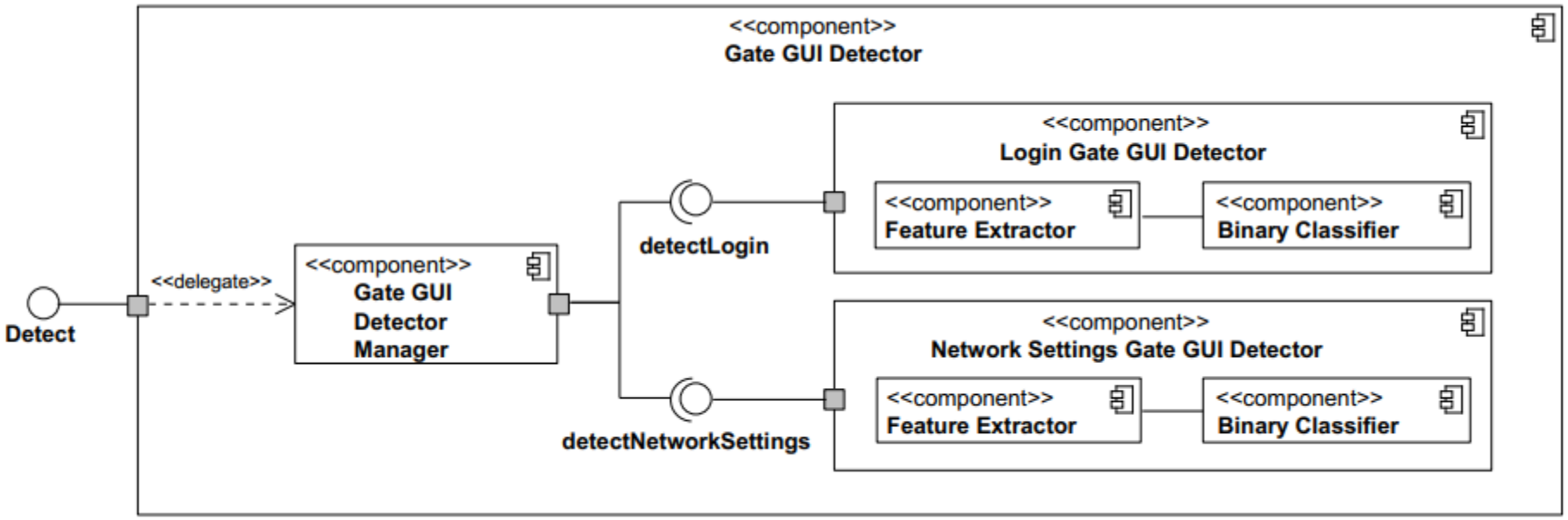
Combining AGET and C&R



The juGULAR Platform



Gate GUI Detector



Experimental Evaluation

- **GOAL:** Understand how the hybridization proposed by juGULAR does impact the ability of fully automated GUI exploration techniques in analyzing apps and at what cost.

  - **RQ1:** How does the hybridization introduced by juGULAR affect the effectiveness of an automated exploration technique?
  - **RQ2:** How does the manual intervention required by juGULAR affect the costs of the hybrid exploration approach?
  - **RQ3:** How does the exploration effectiveness of juGULAR compare to the effectiveness of the AGET implemented by the state-of-the-practice Monkey tool?

Objects

App ID	App Name	Package Name	Version	App Description
A1	Flym News Reader	net.fred.feedex	1.9.0	Simple, modern and totally free RSS reader.
A2	Conversations	eu.siacs.conversations	1.19.5	Jabber/XMPP client for Android.
A3	DAVdroid	at.bitfire.davdroid	1.5.0.3-ose	Calendar synchronization app.
A4	Transistor Radio	org.y20k.transistor	2.2.0	App for listening to radio over internet.
A5	k9-Mail	com.fsck.k9	5.206	Email client supporting multiple accounts.
A6	mGit	com.manichord.mgit	1.5.0	Git client and text editor.
A7	Muspy	com.danielme.muspyforandroid	3.4.48	Client for Muspy.com.
A8	OpenRedmine	jp.redmine.redmineclient	3.20	Android Redmine client.
A9	OwnCloud	com.owncloud.android	2.3.0	Android client for private ownCloud Server.
A10	PortKnocker	com.xargsgrep.portknocker	1.0.11	App that pings a specific TCP/UDP port.
A11	LibreTorrent	org.proninyaroslav.libretorrent	1.4	Original Free torrent client.
A12	Connectbot	org.connectbot	1.9.2-oss	Powerful open-source Secure Shell (SSH) client.
A13	PodListen	com.einmalfel.podlisten	1.3.6	Free Podcast app.
A14	ServeStream	net.sourceforge.servestream	0.7.3	Open source HTTP streaming media player and media server browser.

Metrics

$$CA\% = \frac{\#\,Covered\ Activities}{\#\,App\ Activities} * 100$$

$$CLOC\% = \frac{\#\,Covered\ LOCs}{\#\,App\ LOCs} * 100$$

$$NTB = \#\,App\ Sent\ Bytes + \#\,App\ Received\ Bytes$$

$$MIT\% = \frac{\sum_i CaptureTime_i}{TotalExplorationTime} * 100$$

Covered Lines Of Code