Automated Cross-Platform Inconsistency Detection for Mobile Apps

Mattia Fazzini

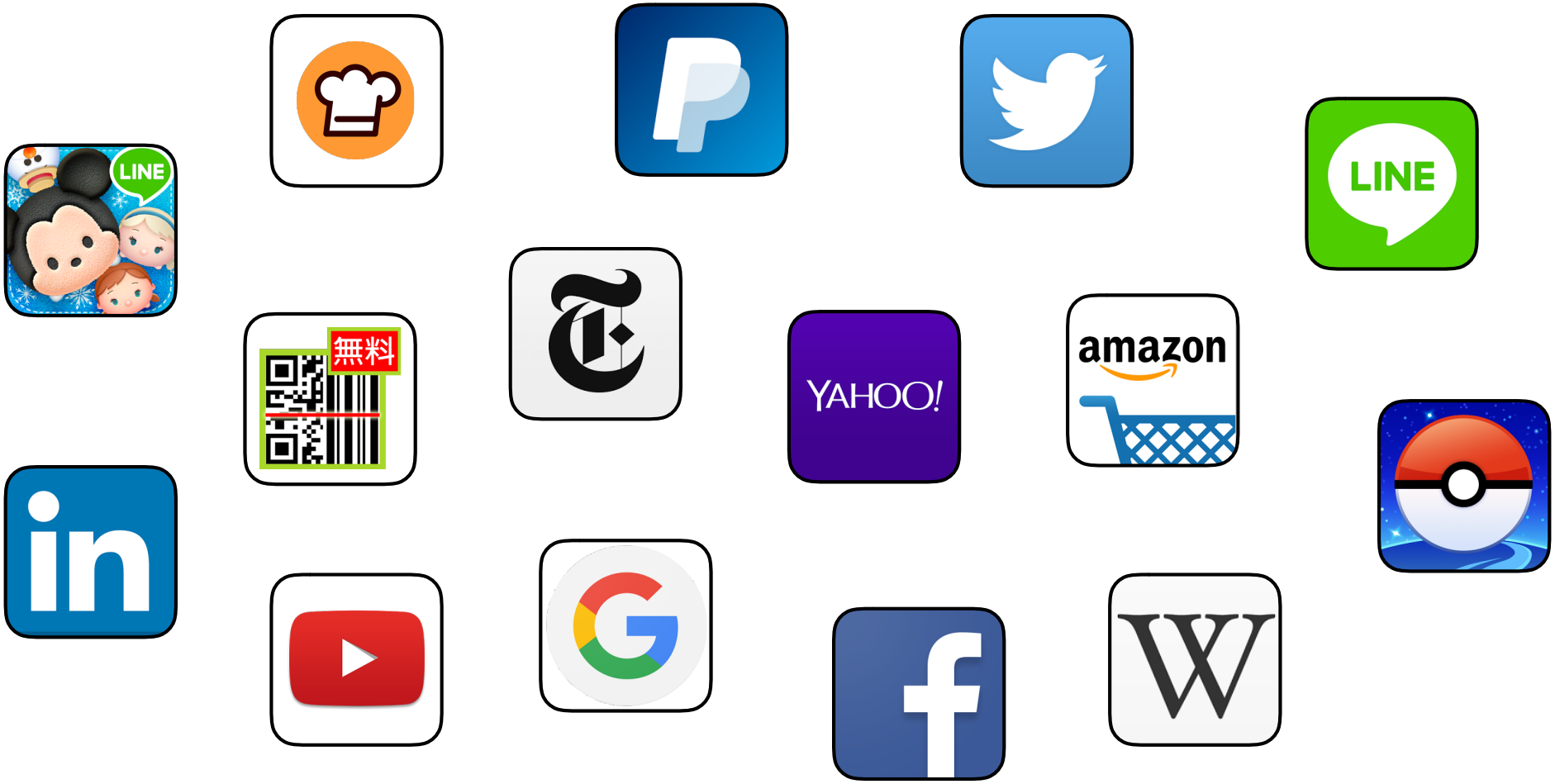
Georgia Tech

Alessandro Orso

Georgia Tech

Mobile Applications

Mobile Applications



Fragmentation

Device

System

Screen

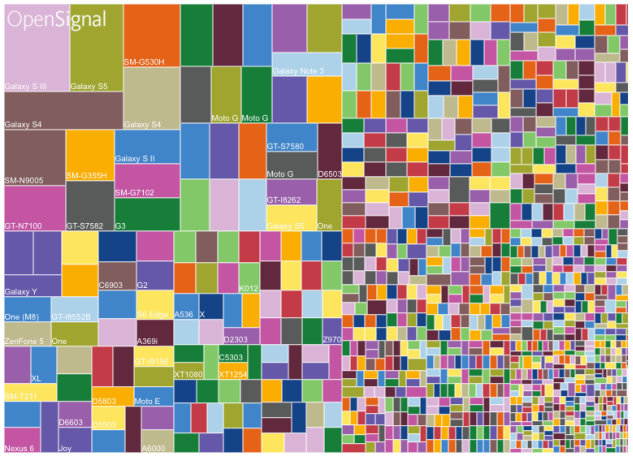
News

Jul 2016

"Due to the wide array of available Android devices, **we are targeting our support to a select number of Android devices** to continue improving our overall Salesforce1 for Android user experience"

Fragmentation

Device



System

Screen

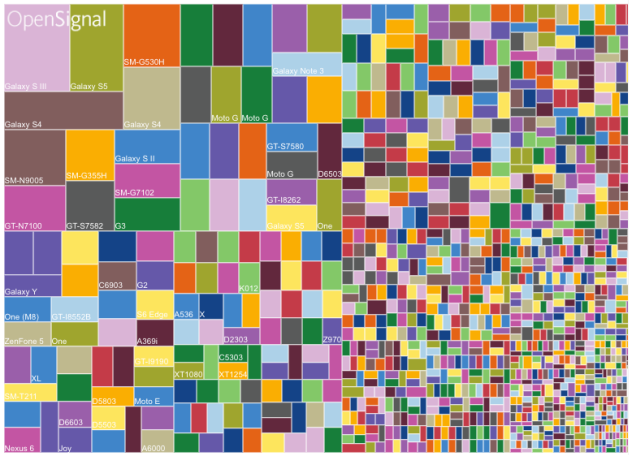
News

Jul 2016

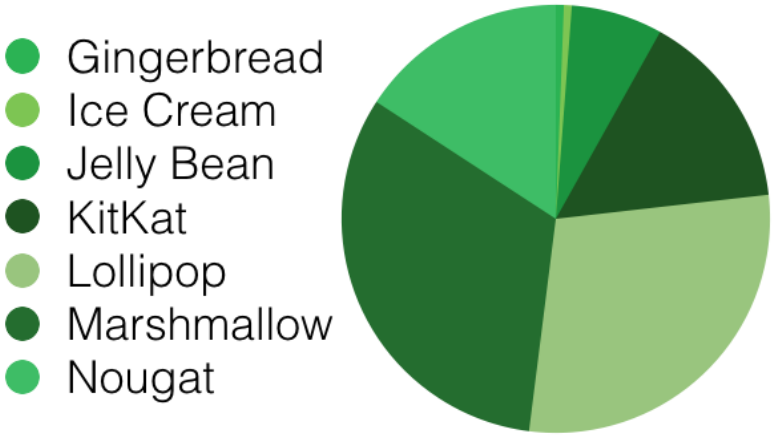
"Due to the wide array of available Android devices, **we are targeting our support to a select number of Android devices** to continue improving our overall Salesforce1 for Android user experience"

Fragmentation

Device



System



Screen

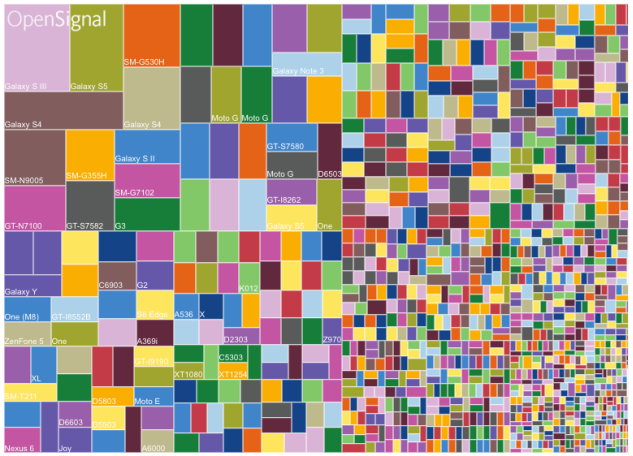
News

Jul 2016

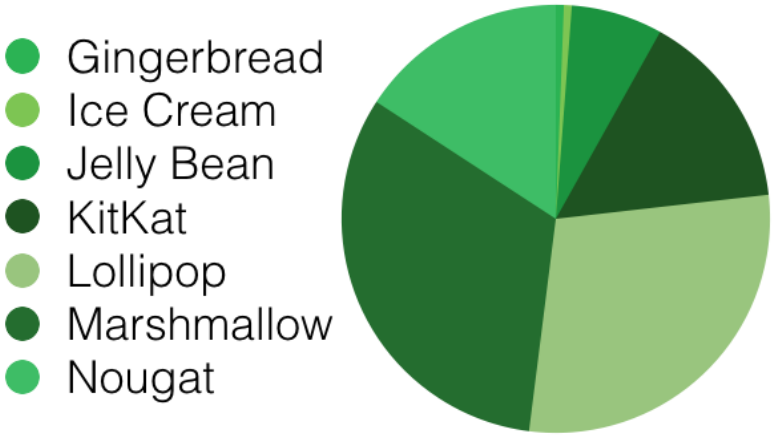
"Due to the wide array of available Android devices, **we are targeting our support to a select number of Android devices** to continue improving our overall Salesforce1 for Android user experience"

Fragmentation

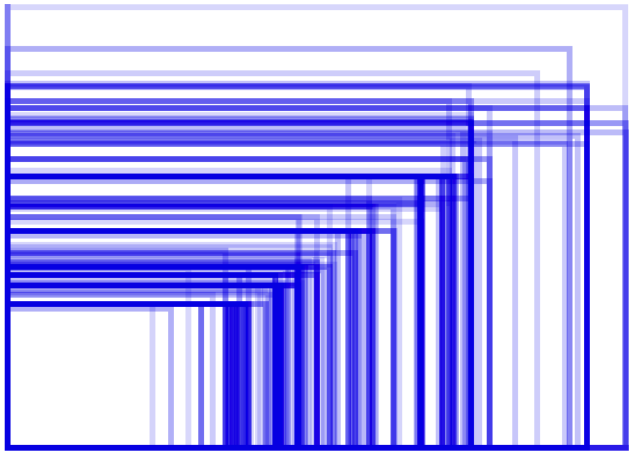
Device



System



Screen



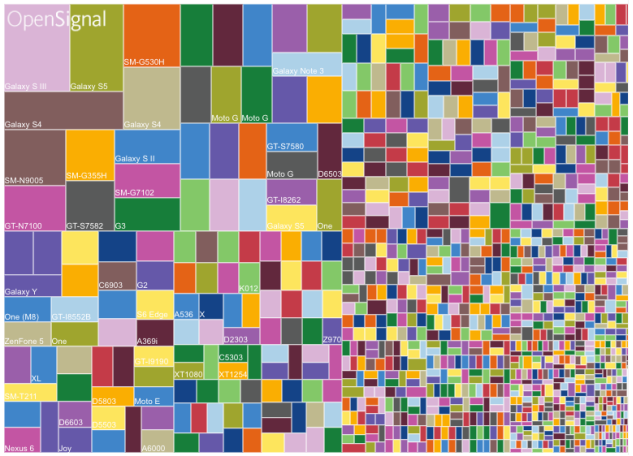
News

Jul 2016

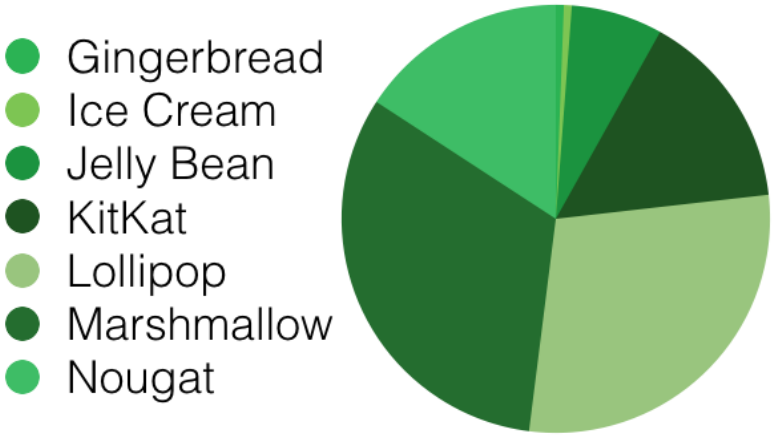
"Due to the wide array of available Android devices, **we are targeting our support to a select number of Android devices** to continue improving our overall Salesforce1 for Android user experience"

Fragmentation

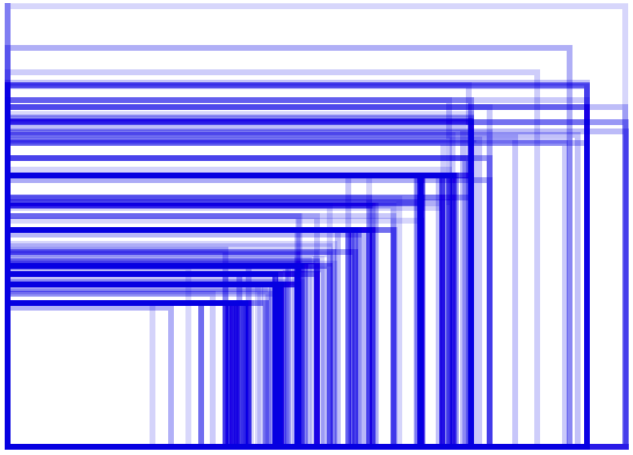
Device



System



Screen



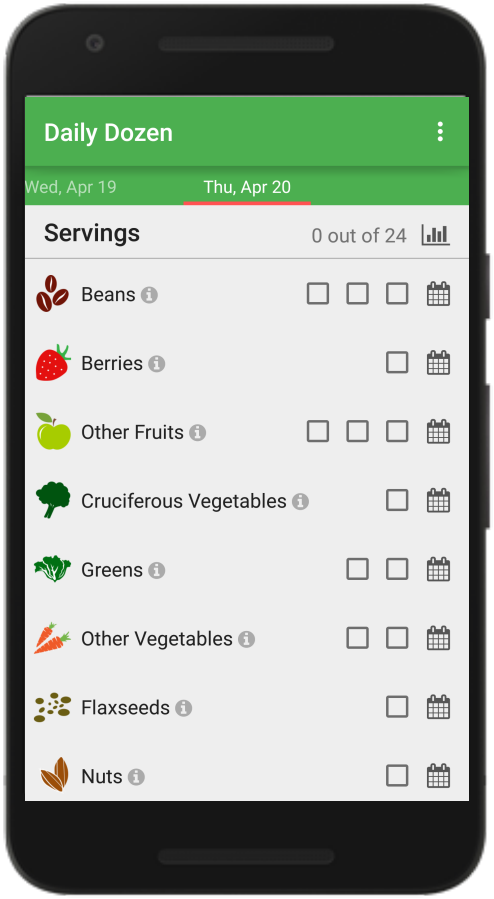
News

Jul 2016

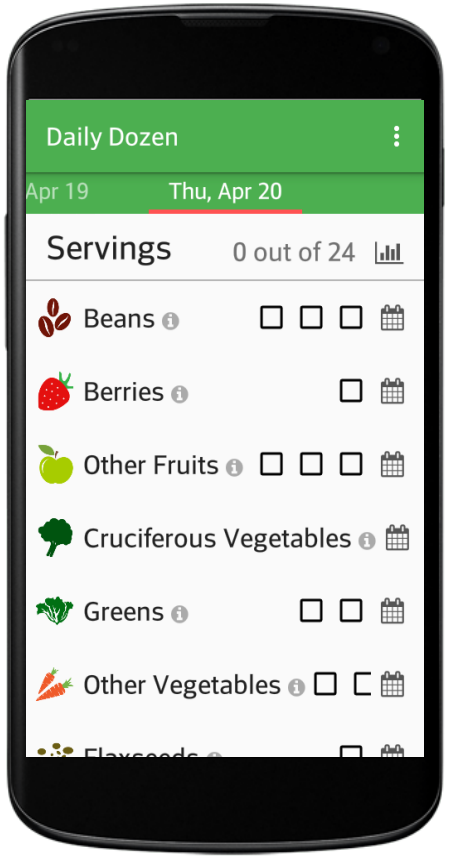
"Due to the wide array of available Android devices, **we are targeting our support to a select number of Android devices** to continue improving our overall Salesforce1 for Android user experience"

Motivating Example

LG G3

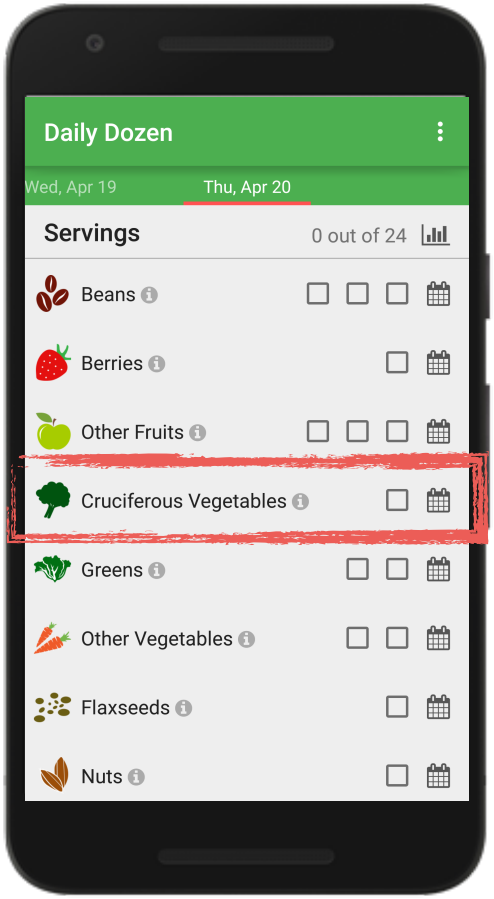


LG Optimus L70

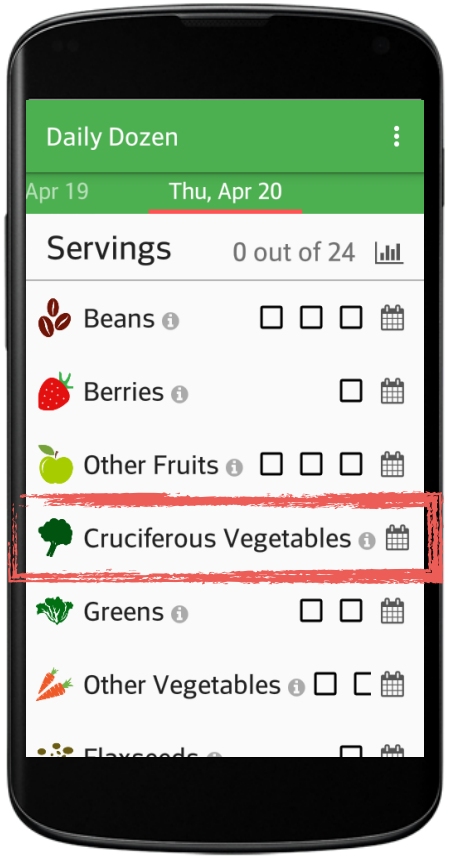


Motivating Example

LG G3

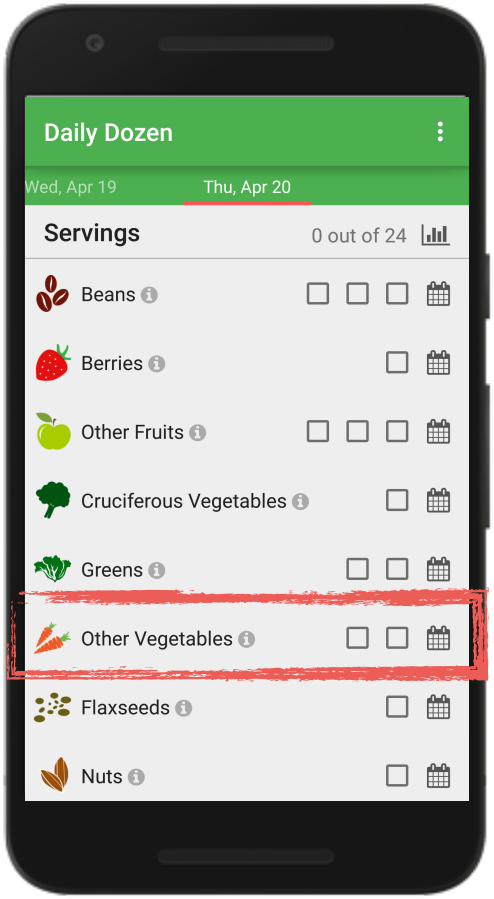


LG Optimus L70

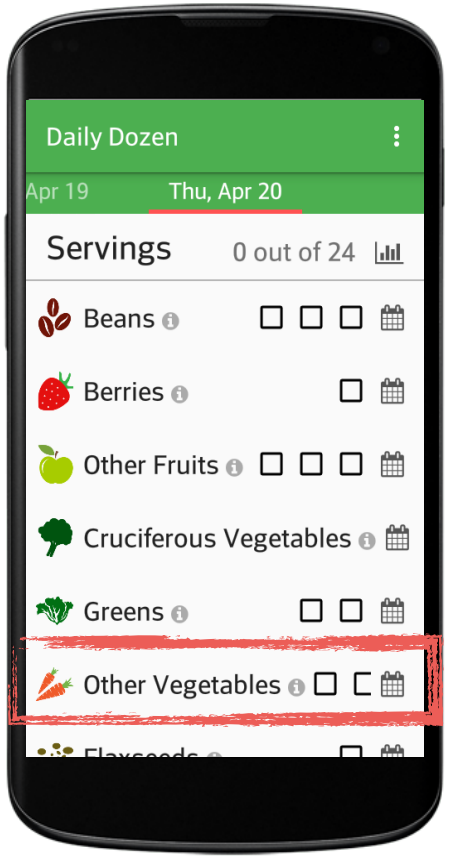


Motivating Example

LG G3

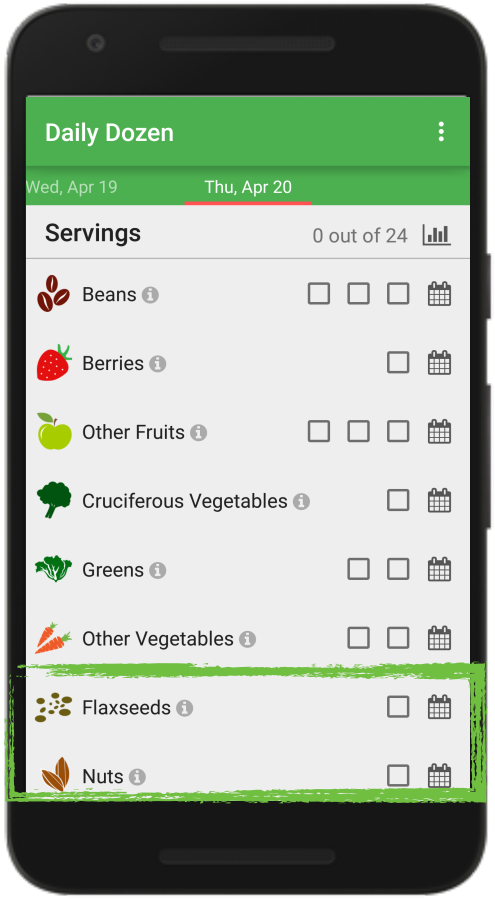


LG Optimus L70

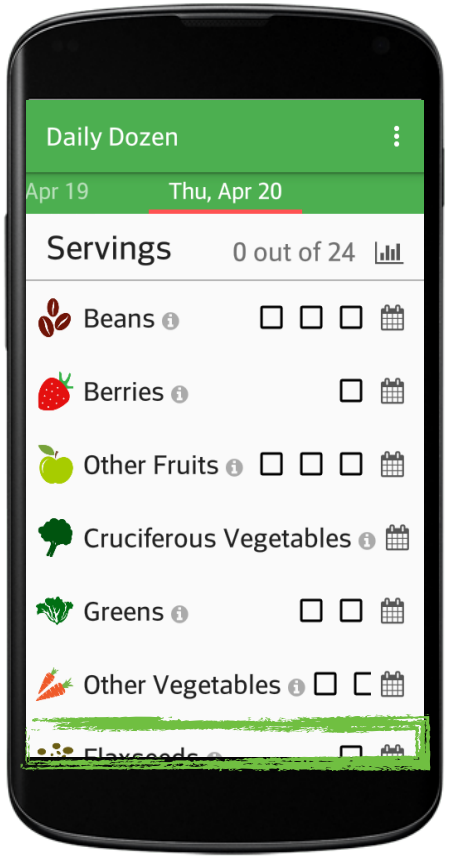


Motivating Example

LG G3

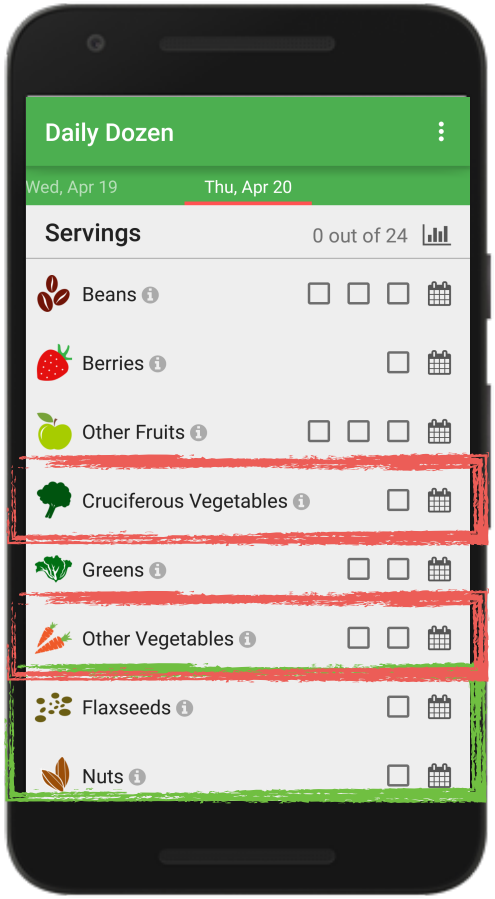


LG Optimus L70

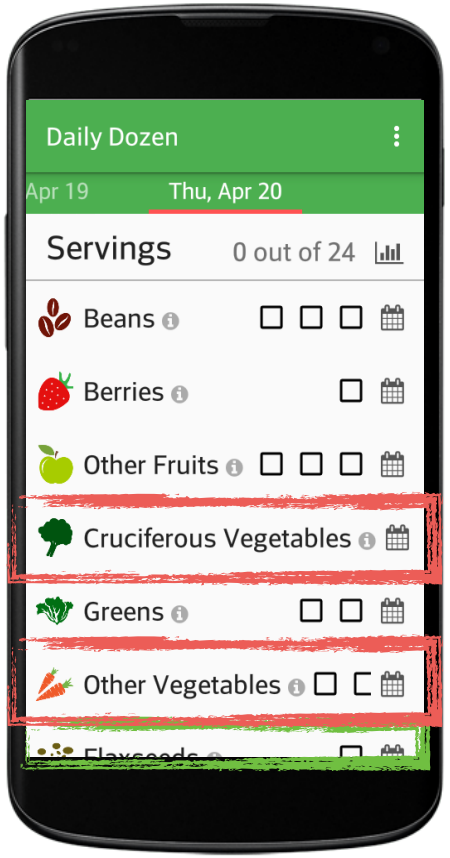


Motivating Example

LG G3

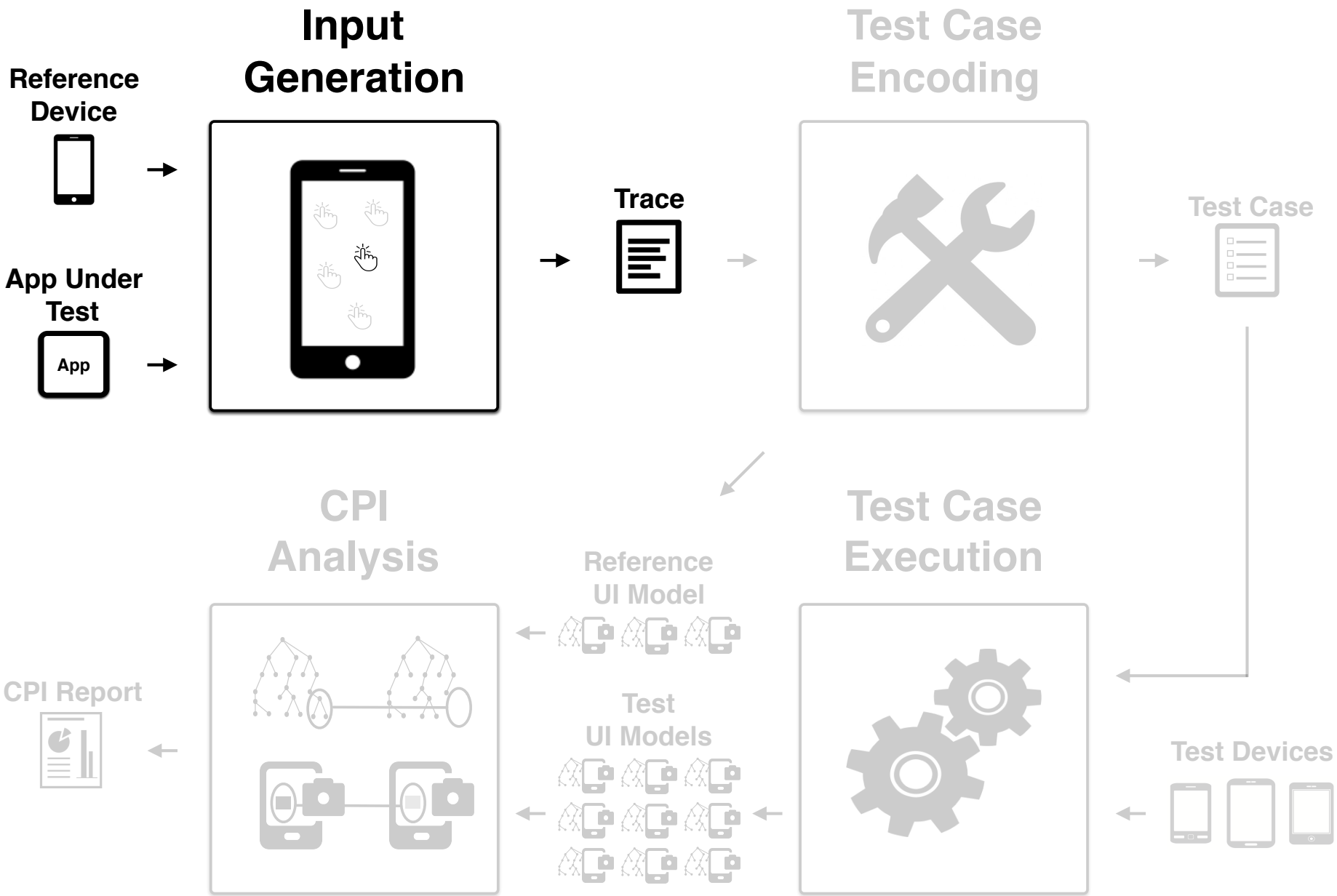


LG Optimus L70

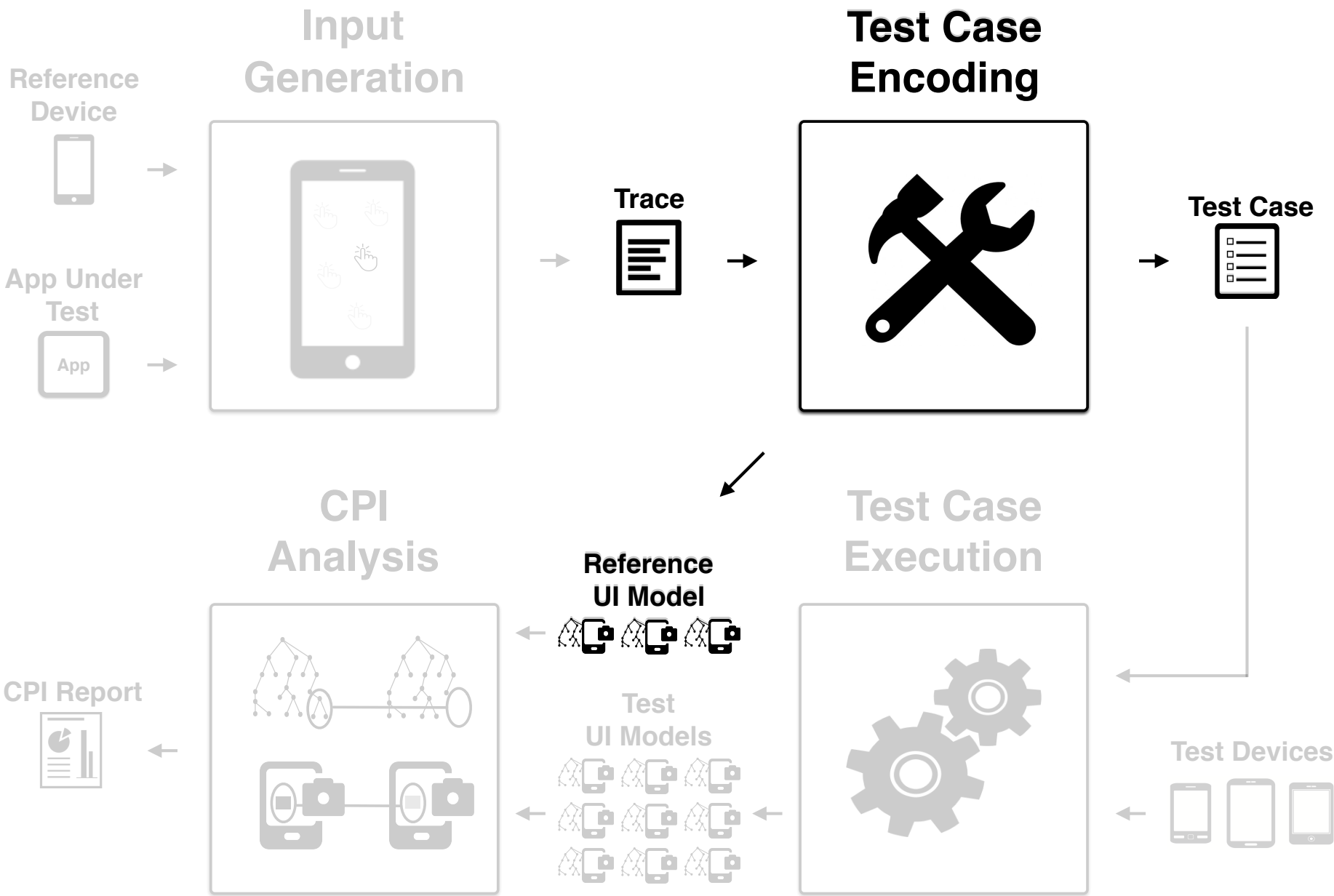


DIFFDROID Overview

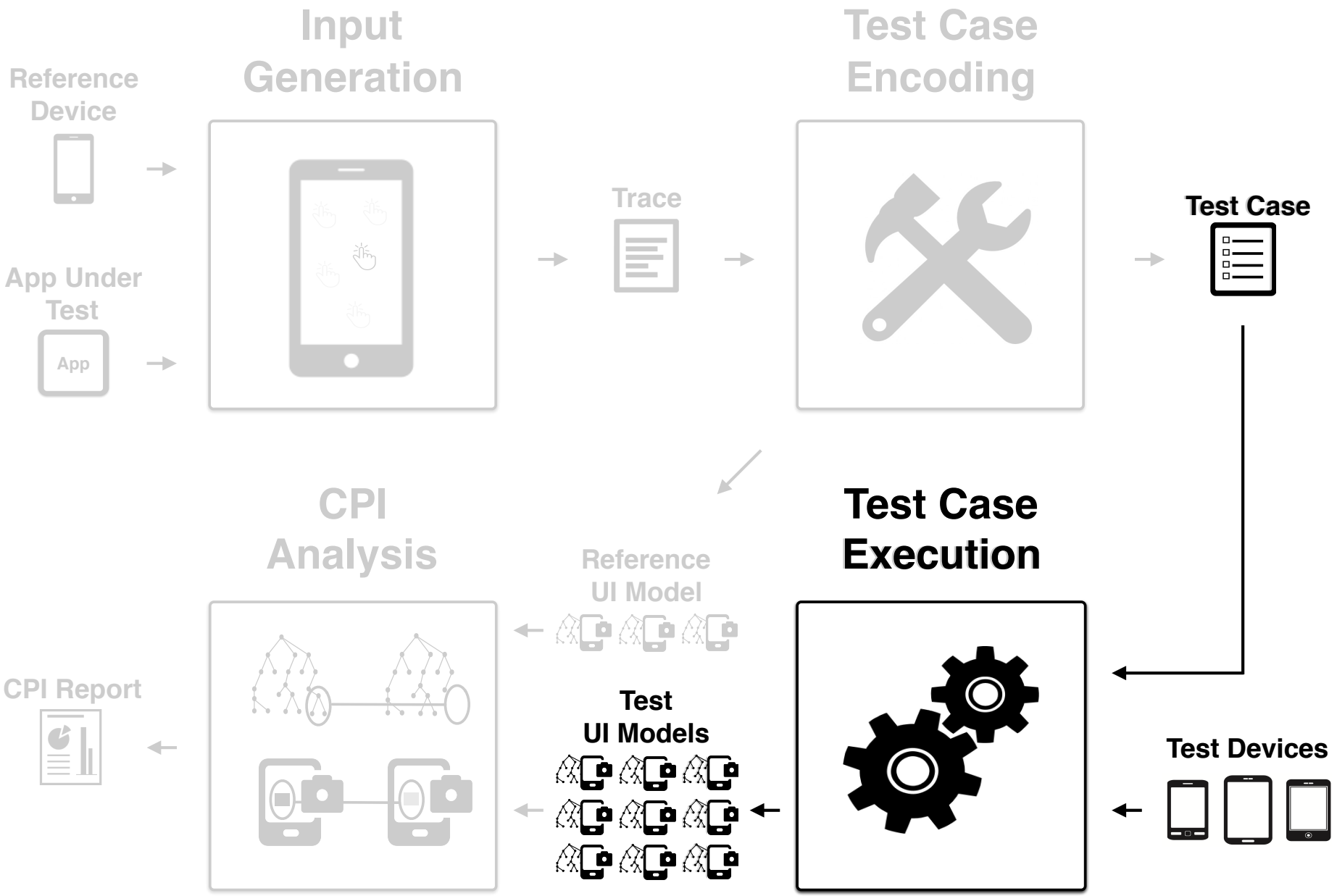
DIFFDROID Overview



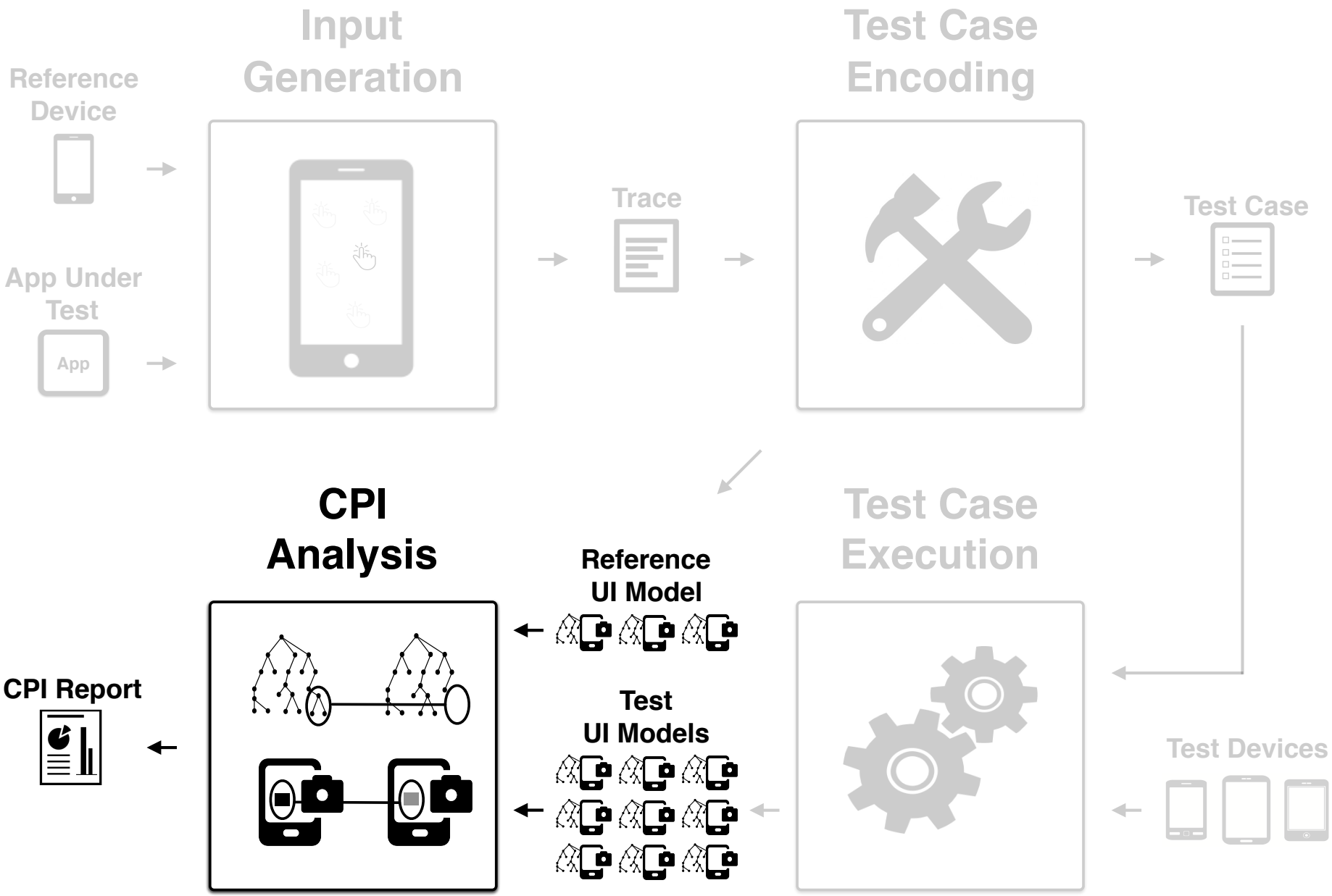
DIFFDROID Overview



DIFFDROID Overview



DIFFDROID Overview



Input Generation

Reference Device

Inputs

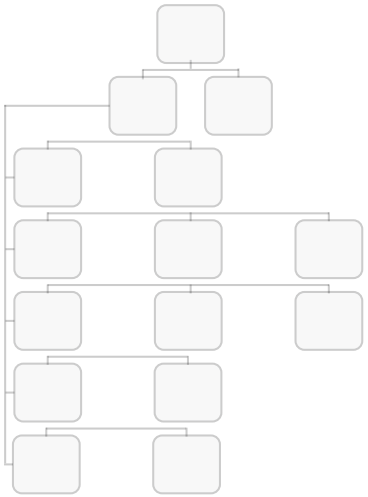
Key

System

Touch

Window Model

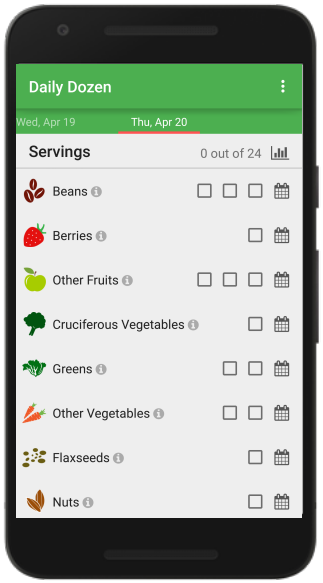
UI Hierarchy



Screenshot

Input Generation

Reference Device



Inputs

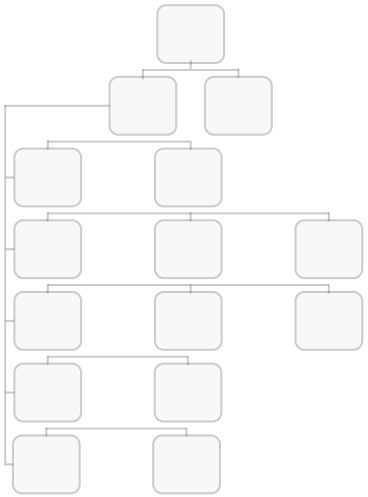
Key

System

Touch

Window Model

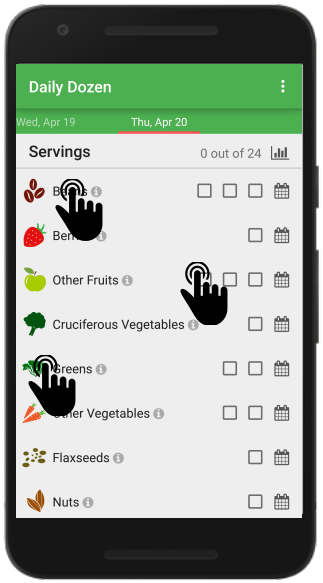
UI Hierarchy



Screenshot

Input Generation

Reference Device

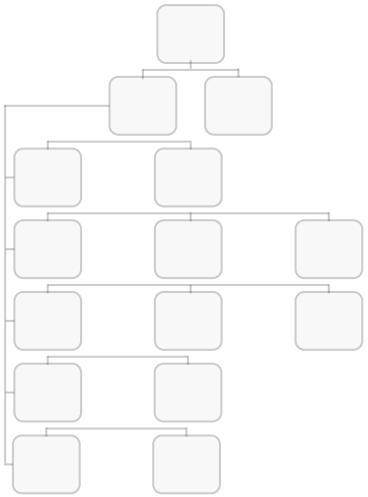


Inputs

- Key
- System
- Touch

Window Model

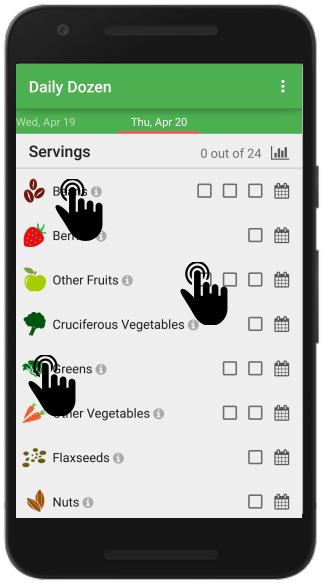
UI Hierarchy



Screenshot

Input Generation

Reference Device



Inputs

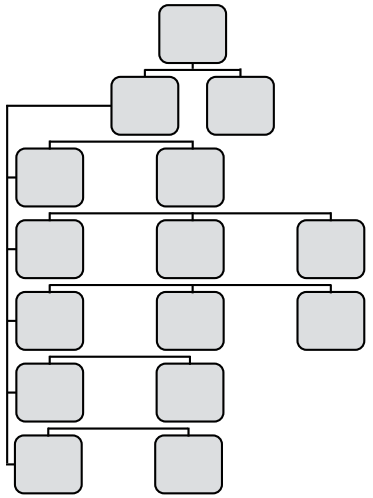
Key

System

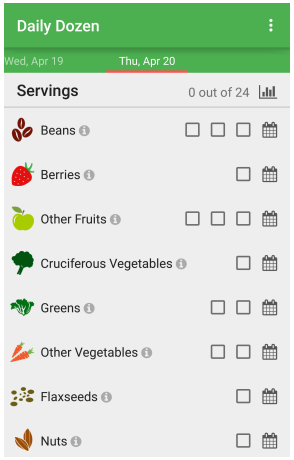
Touch

Window Model

UI Hierarchy

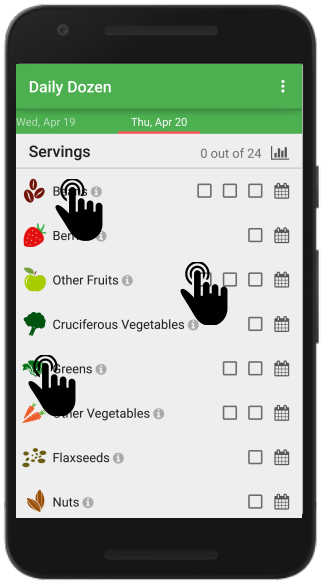


Screenshot



Input Generation

Reference Device

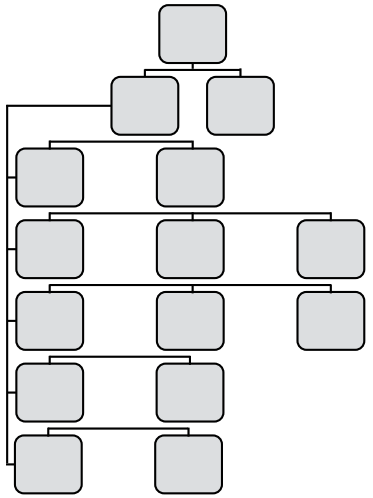


Inputs

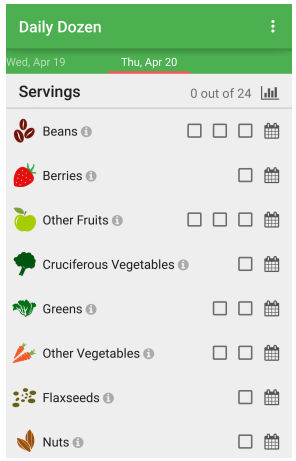
- Key
- System
- Touch

Window Model

UI Hierarchy



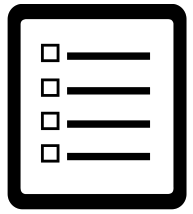
Screenshot



Trace

Test Case Encoding

Test Case

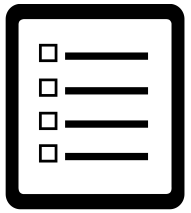


Trace

Reference
UI Model

Test Case Encoding

Test Case



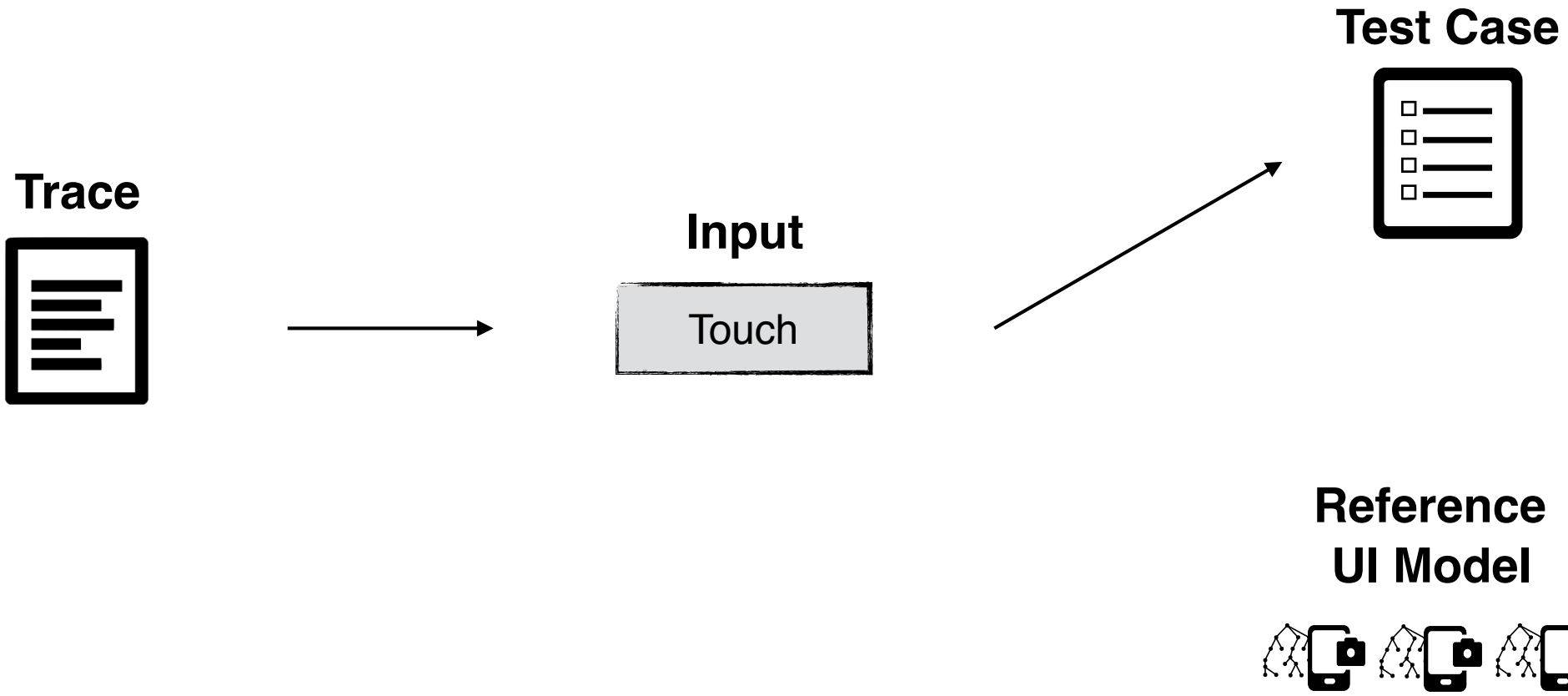
Trace

Input

Touch

Reference
UI Model

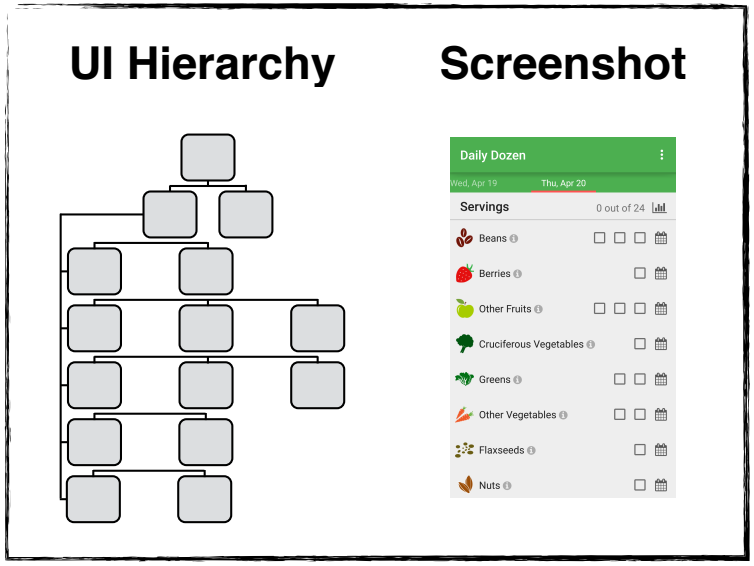
Test Case Encoding



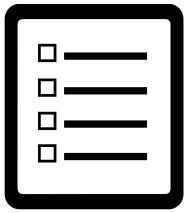
Test Case Encoding

Trace

Window Model

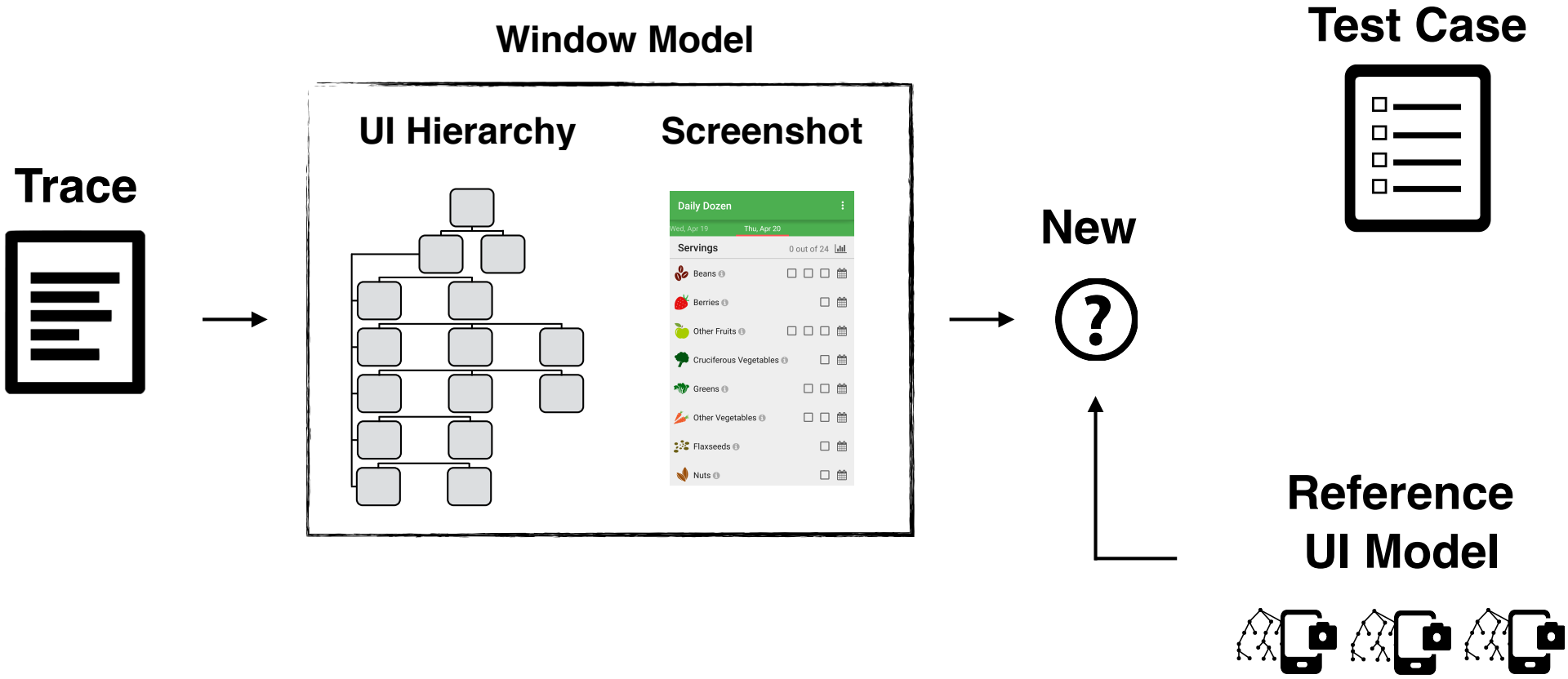


Test Case

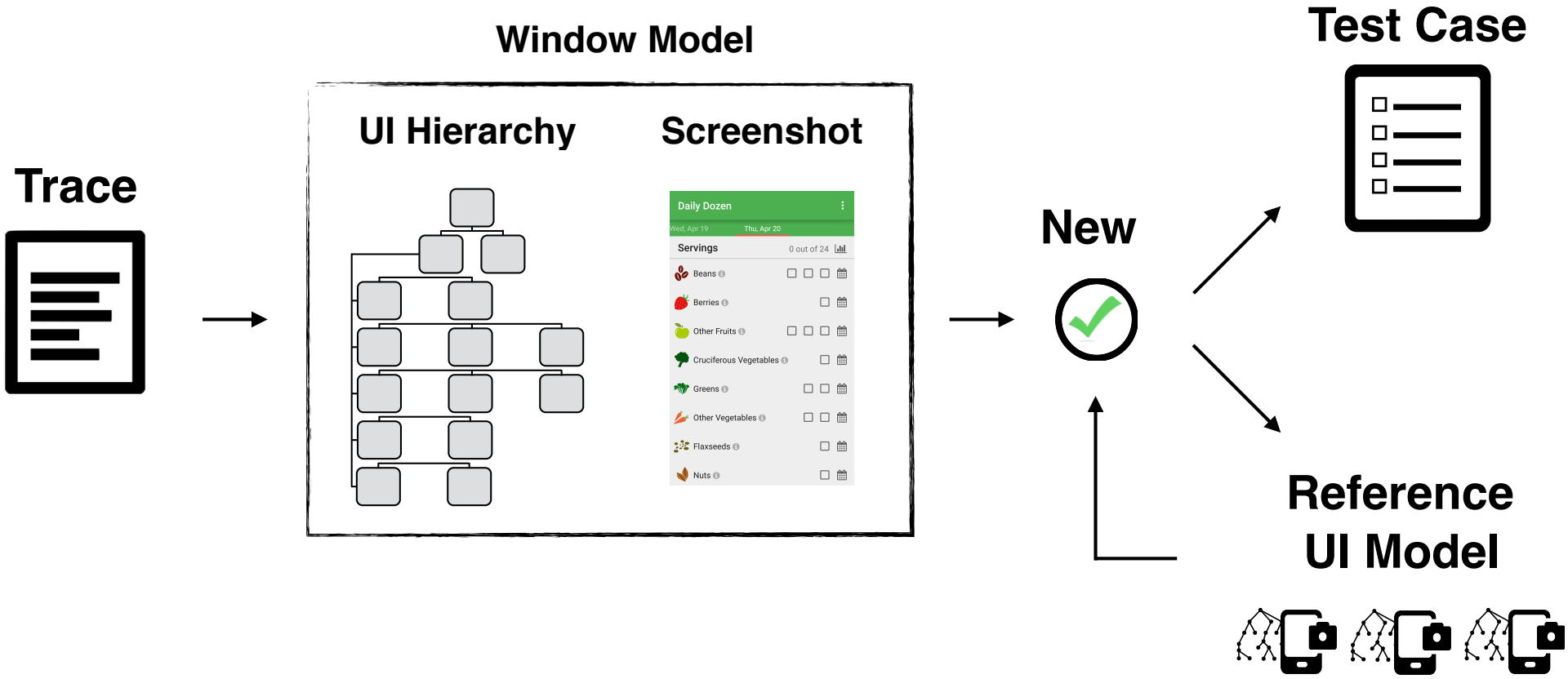


Reference UI Model

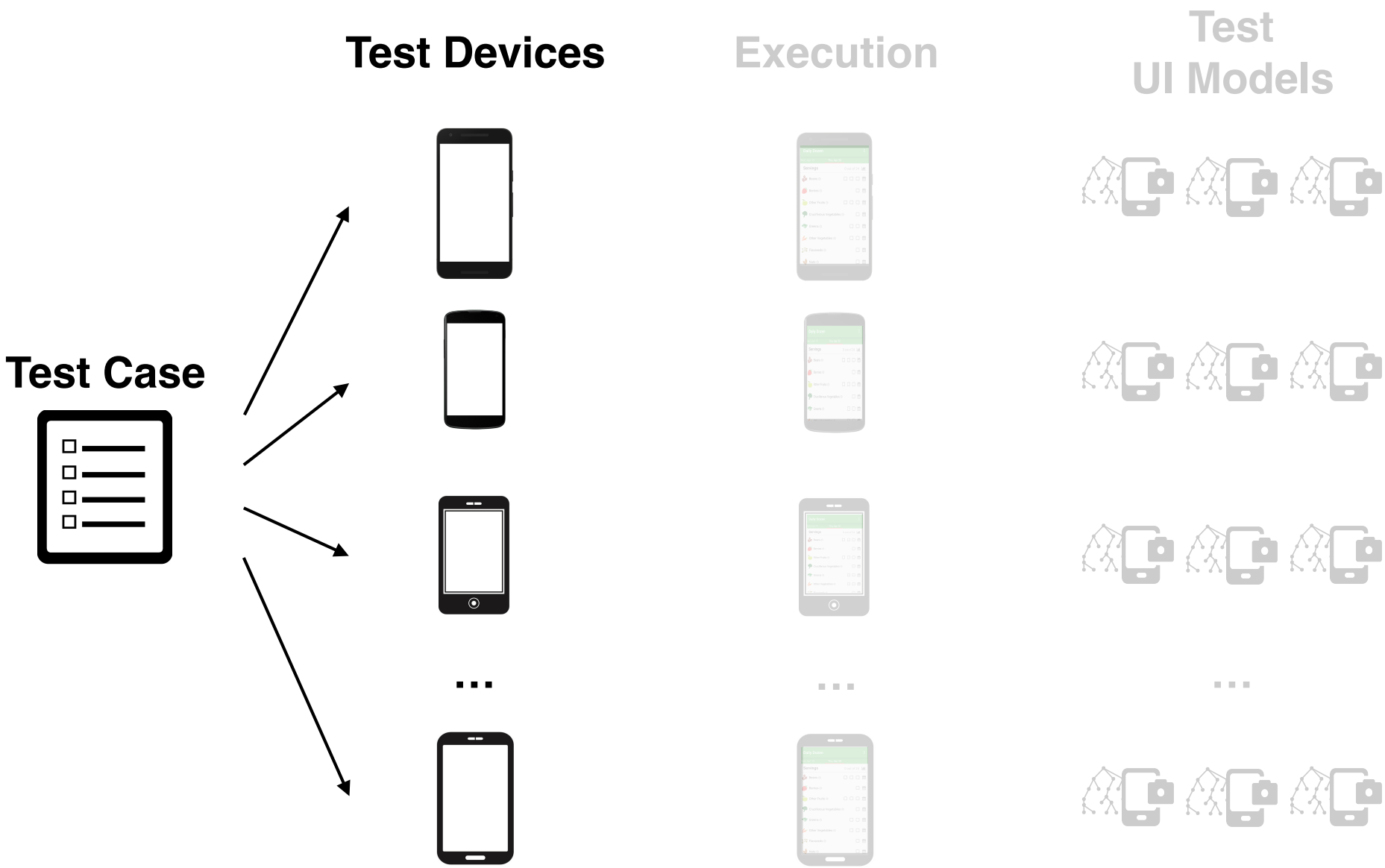
Test Case Encoding



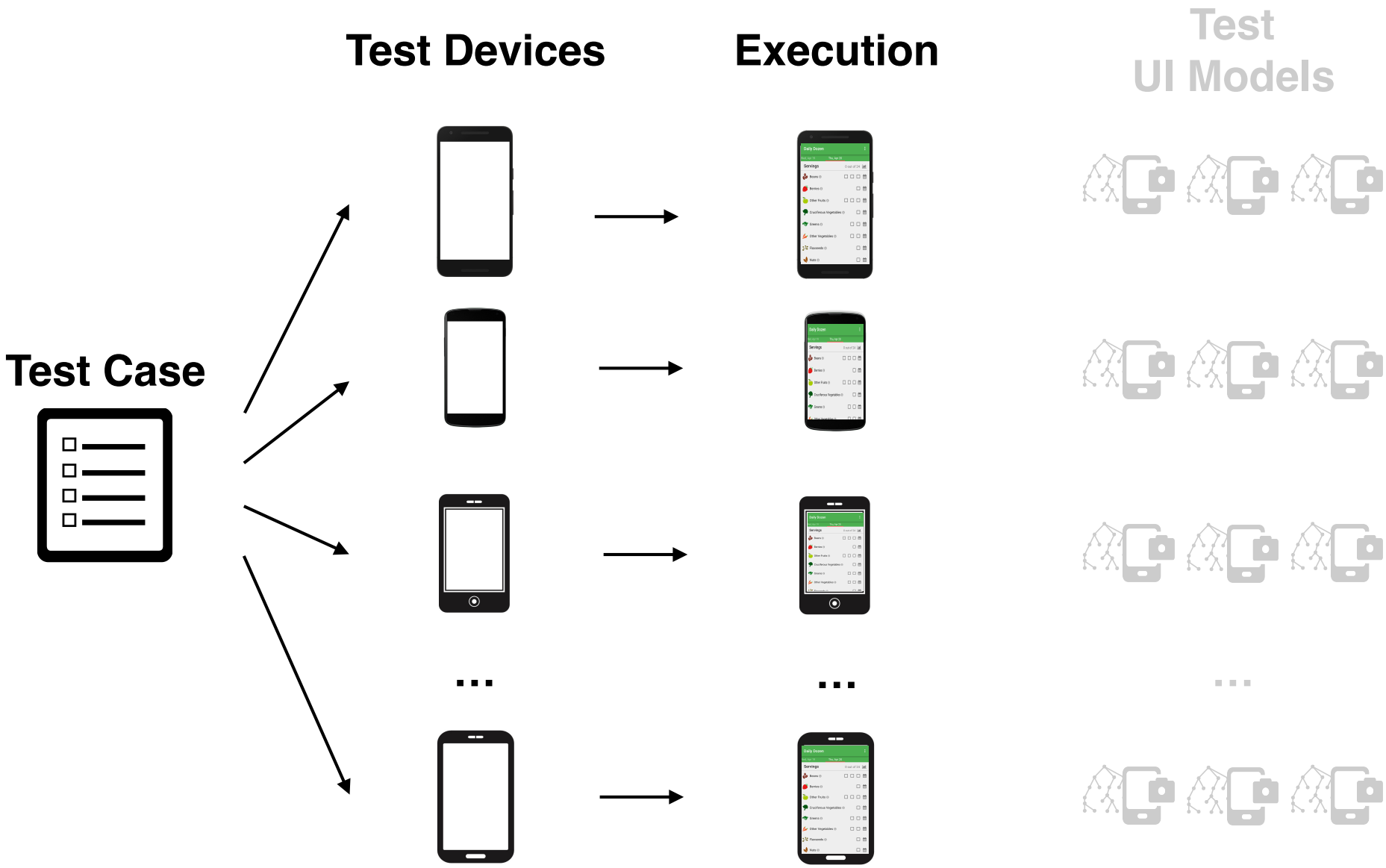
Test Case Encoding



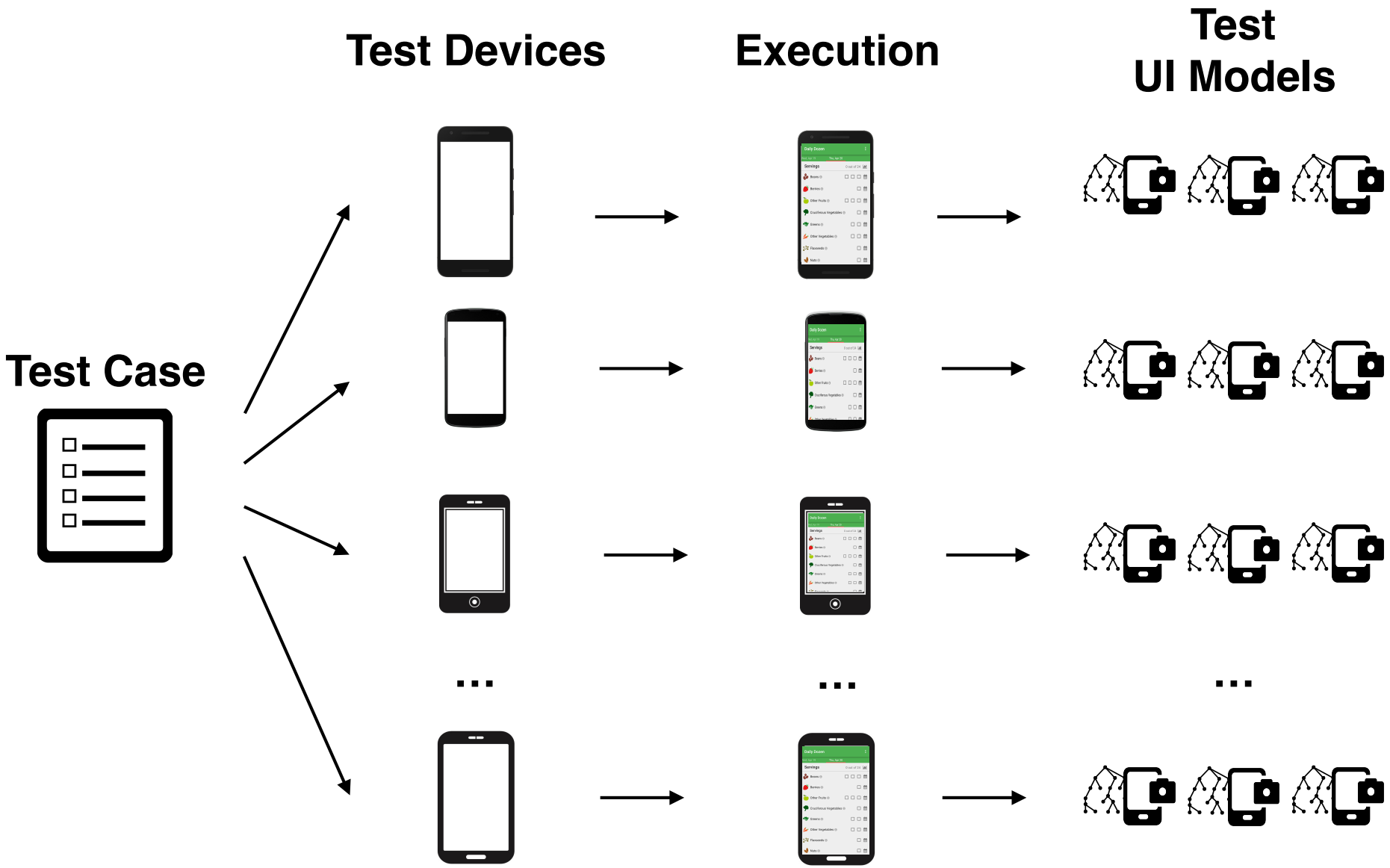
Test Case Execution



Test Case Execution



Test Case Execution

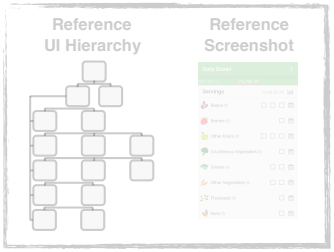


CPI Analysis

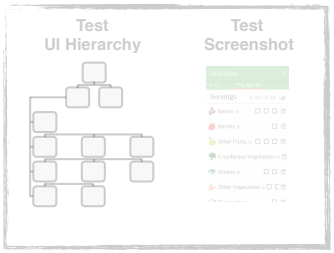
Reference UI Model

Test UI Model

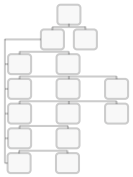
Reference Window Model



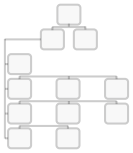
Test Window Model



Reference UI Hierarchy



Test UI Hierarchy

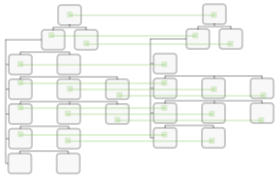


Reference Screenshot

Test Screenshot

Structural Analysis

Node Matching



Visual Analysis

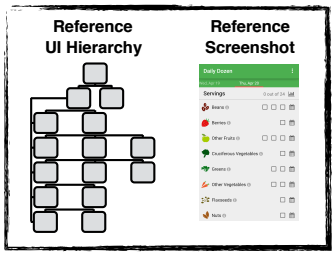
CPI Report

CPI Analysis

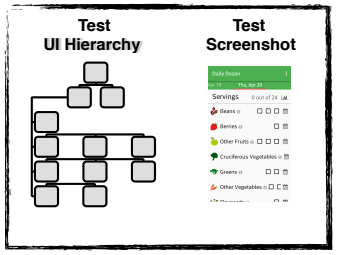
Reference
UI Model

Test
UI Model

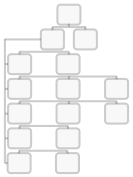
Reference Window Model



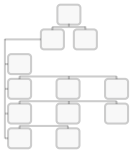
Test Window Model



Reference
UI Hierarchy



Test
UI Hierarchy

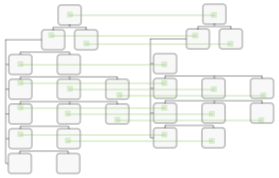


Reference
Screenshot

Test
Screenshot

Structural Analysis

Node Matching



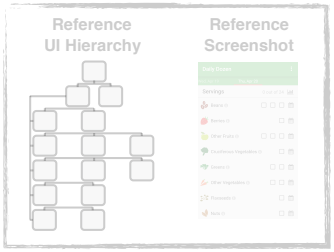
Visual Analysis

CPI
Report

CPI Analysis

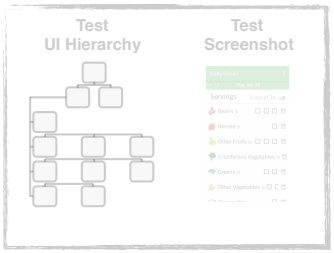
Reference
UI Model

Reference
Window Model



Test
UI Model

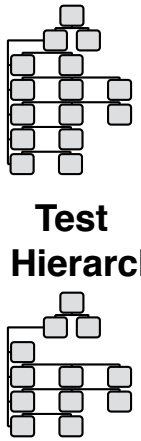
Test
Window Model

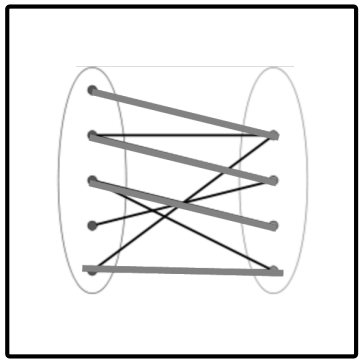


Structural Analysis

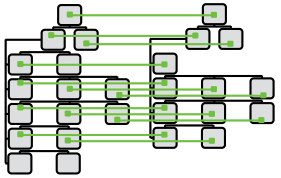
Reference
UI Hierarchy

Test
UI Hierarchy





Node Matching



CPI
Report

Reference
Screenshot

Test
Screenshot

Visual Analysis

CPI Analysis

Structural Analysis

Reference
UI Model

Reference
Window Model

Reference
UI Hierarchy

Reference
Screenshot

Reference
UI Hierarchy

Test
UI Hierarchy

Test
UI Model

Test
Window Model

Test
UI Hierarchy

Test
Screenshot

Reference
Screenshot

Test
Screenshot

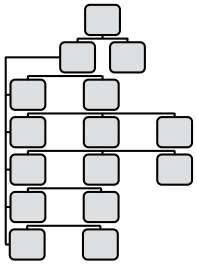
Node Matching

Visual Analysis

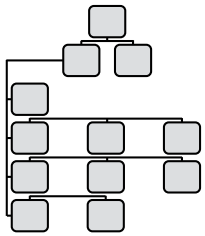
CPI
Report

Structural Analysis

Reference UI Hierarchy



Test UI Hierarchy



Node Similarity

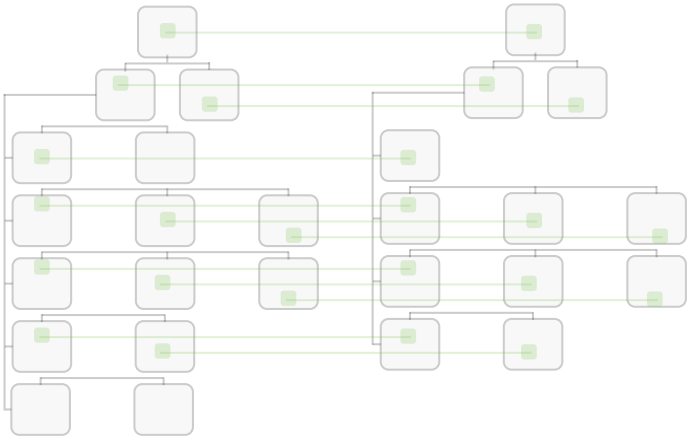
Resource ID

XPath

Properties

checkable	checked
focusable	focused
clickable	selected
long-clickable	enabled
scrollable	text

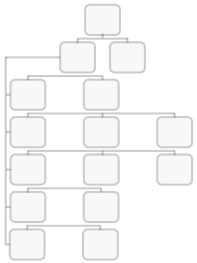
Node Matching



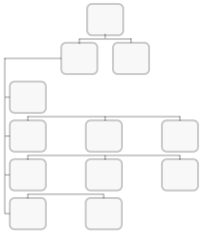
Structural Analysis

Node Similarity

Reference
UI Hierarchy



Test
UI Hierarchy



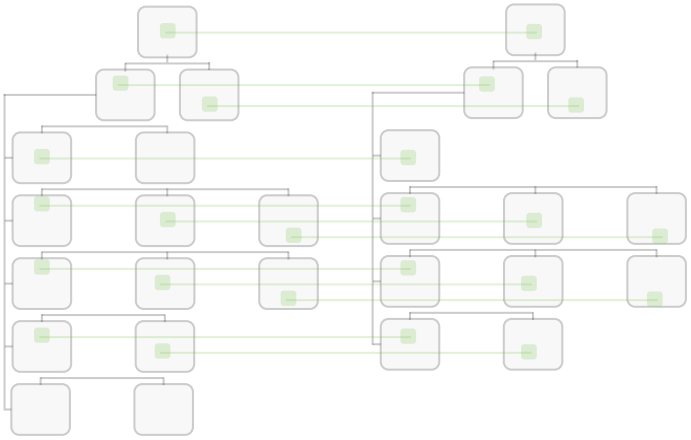
Resource ID

XPath

Properties

checkable	checked
focusable	focused
clickable	selected
long-clickable	enabled
scrollable	text

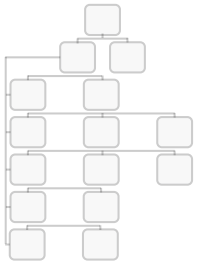
Node Matching



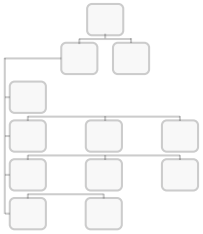
Structural Analysis

Node Similarity

Reference
UI Hierarchy



Test
UI Hierarchy



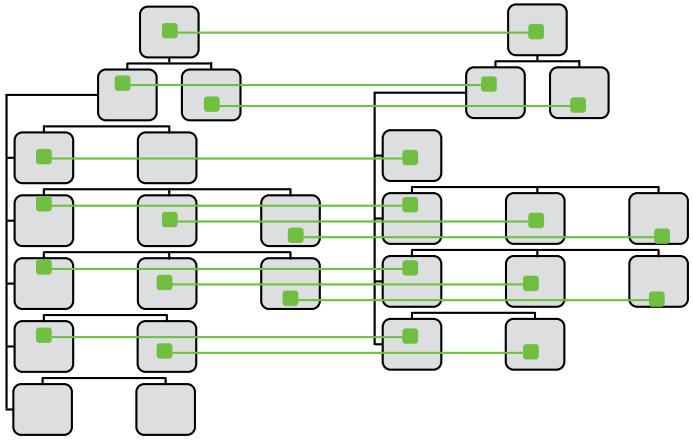
Resource ID

XPath

Properties

checkable	checked
focusable	focused
clickable	selected
long-clickable	enabled
scrollable	text

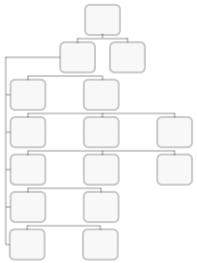
Node Matching



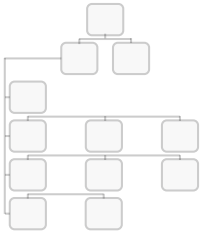
Structural Analysis

Node Similarity

Reference
UI Hierarchy



Test
UI Hierarchy



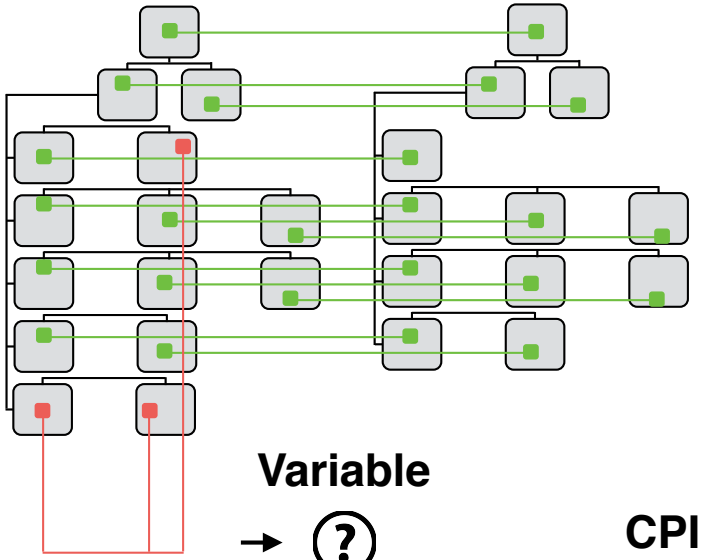
Resource ID

XPath

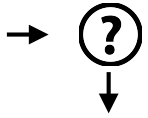
Properties

checkable	checked
focusable	focused
clickable	selected
long-clickable	enabled
scrollable	text

Node Matching



Variable

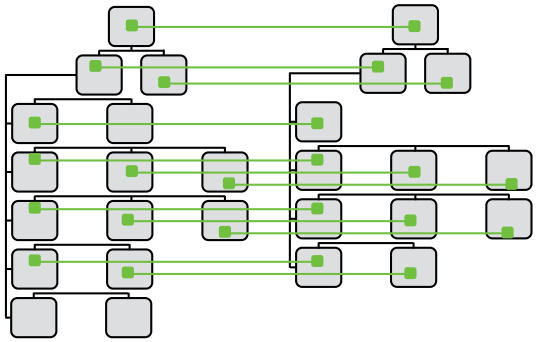


Structural
Inconsistency

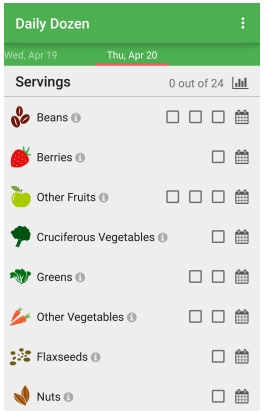
CPI
Report

Visual Analysis

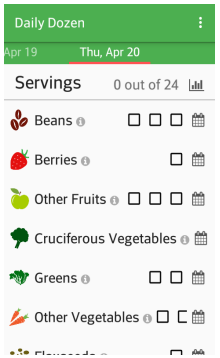
Node Matching



Reference Screenshot



Test Screenshot



Complex-Wavelet Structural Similarity Index (CW-SSIM)

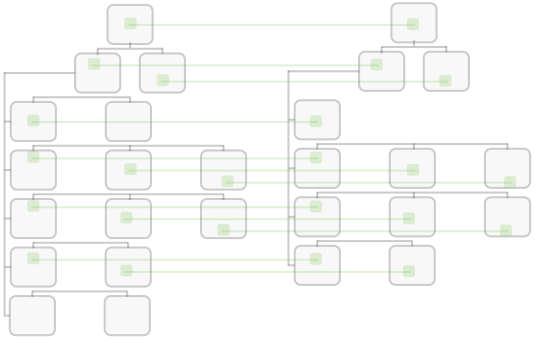
Earth Mover Distance of Color Histogram (EMD)

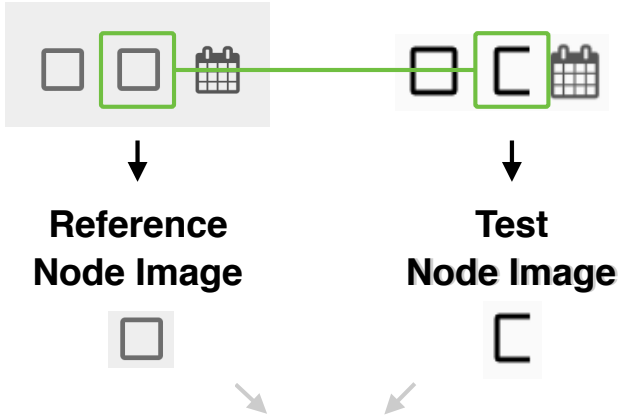
Relative Ratio Change (RRC)

Optical Character Recognition (OCR)

Visual Analysis

Node Matching





Complex-Wavelet Structural Similarity Index (CW-SSIM)

Earth Mover Distance of Color Histogram (EMD)

Relative Ratio Change (RRC)

Optical Character Recognition (OCR)

Reference Screenshot

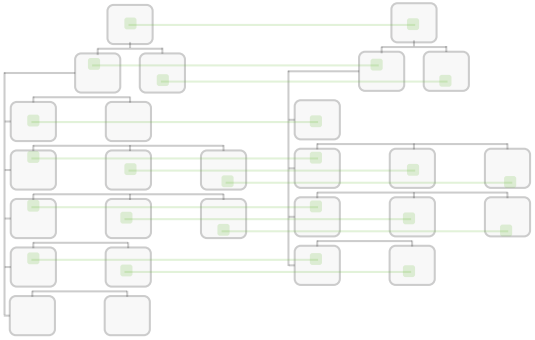
Test Screenshot

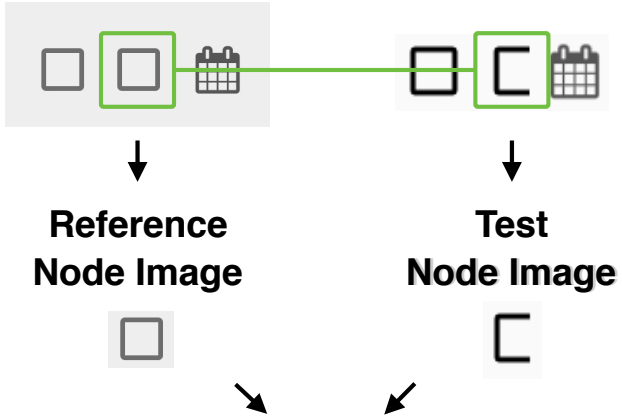
CW-SSIM EMD RRC OCR

C4.5 Decision Tree Classifier

Visual Analysis

Node Matching



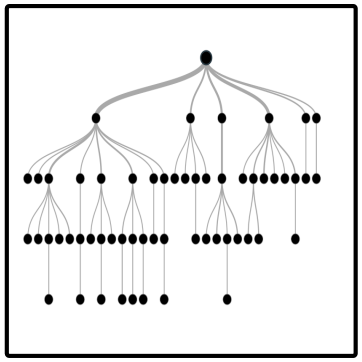


Reference Screenshot

Test Screenshot

CW-SSIM EMD RRC OCR

C4.5 Decision Tree Classifier



Complex-Wavelet Structural Similarity Index (CW-SSIM)

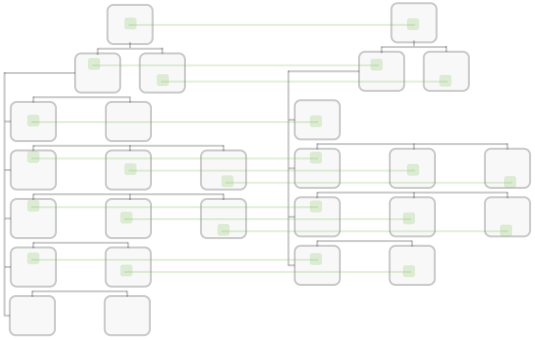
Earth Mover Distance of Color Histogram (EMD)

Relative Ratio Change (RRC)

Optical Character Recognition (OCR)

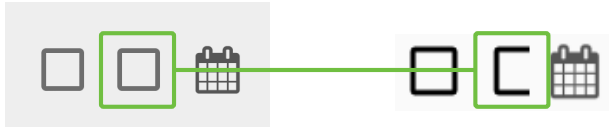
Visual Analysis

Node Matching



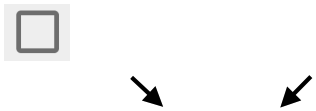
Reference Screenshot

Test Screenshot



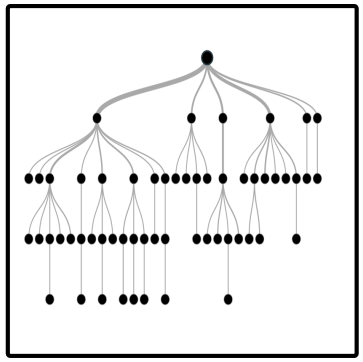
Reference Node Image

Test Node Image



CW-SSIM EMD RRC OCR

C4.5 Decision Tree Classifier



Complex-Wavelet Structural Similarity Index (CW-SSIM)

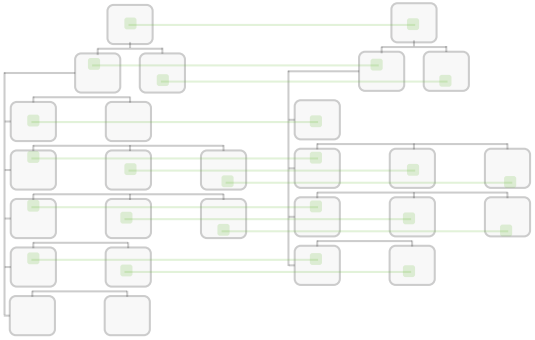
Earth Mover Distance of Color Histogram (EMD)

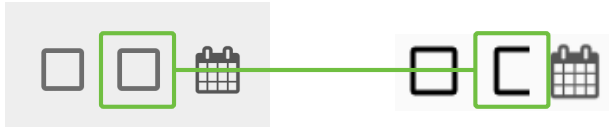
Relative Ratio Change (RRC)

Optical Character Recognition (OCR)

Visual Analysis

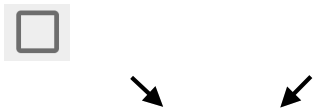
Node Matching





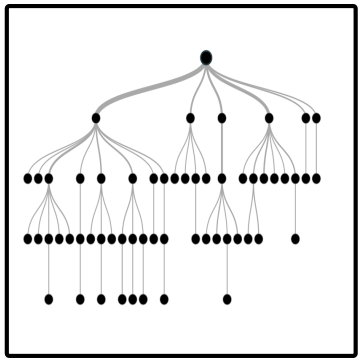
Reference Node Image

Test Node Image



CW-SSIM EMD RRC OCR

C4.5 Decision Tree Classifier



Visual Inconsistency

CPI Report

Complex-Wavelet Structural Similarity Index (CW-SSIM)

Earth Mover Distance of Color Histogram (EMD)

Relative Ratio Change (RRC)

Optical Character Recognition (OCR)

Reference Screenshot

Test Screenshot

Implementation

Input Generation

Monkey

Test Case Encoding

Espresso

Test Case Execution

AWS Device Farm

CPI Analysis

WEKA

Implementation

Input Generation

Monkey

Test Case Encoding

Espresso

Test Case Execution

AWS Device Farm

CPI Analysis

WEKA

Implementation

Input Generation

Monkey

Test Case Encoding

Espresso

Test Case Execution

AWS Device Farm

CPI Analysis

WEKA

Implementation

Input Generation

Monkey

Test Case Encoding

Espresso

Test Case Execution

AWS Device Farm

CPI Analysis

WEKA

Implementation

Input Generation

Monkey

Test Case Encoding

Espresso

Test Case Execution

AWS Device Farm

CPI Analysis

WEKA

Empirical Evaluation

Research Questions:

RQ1: Can DIFFDROID detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DIFFDROID?

RQ3: Are there similarities among devices exhibiting CPIs?

Empirical Evaluation

Research Questions:

RQ1: Can DIFFDROID detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DIFFDROID?

RQ3: Are there similarities among devices exhibiting CPIs?

Empirical Evaluation

Research Questions:

RQ1: Can DIFFDROID detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DIFFDROID?

RQ3: Are there similarities among devices exhibiting CPIs?

Empirical Evaluation

Research Questions:

RQ1: Can DIFFDROID detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DIFFDROID?

RQ3: Are there similarities among devices exhibiting CPIs?

Empirical Evaluation

Research Questions:

RQ1: Can DIFFDROID detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DIFFDROID?

RQ3: Are there similarities among devices exhibiting CPIs?

Benchmarks and Setup

ID	Name	Category	Version	LOC (#K)
A1	BuildmLearn	Education	2.5.0	23.6
A2	Daily Dozen	Health	10.3	6.3
A3	Kitchen Timer	Tools	1.1.6	4.3
A4	Outlay	Finance	1.1.3	8
A5	Translation Studio	Books	9.0	51.2

Reference Device

LG G3, Android 22

Test Devices

147

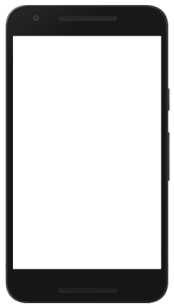
Resolution	OS Version
720x1280	19
768x1280	21
1080x1920	22
1440x2560	23
480x800	24
540x960	25
480x854	

Benchmarks and Setup

ID	Name	Category	Version	LOC (#K)
A1	BuildmLearn	Education	2.5.0	23.6
A2	Daily Dozen	Health	10.3	6.3
A3	Kitchen Timer	Tools	1.1.6	4.3
A4	Outlay	Finance	1.1.3	8
A5	Translation Studio	Books	9.0	51.2

Reference Device

LG G3, Android 22



Test Devices

147

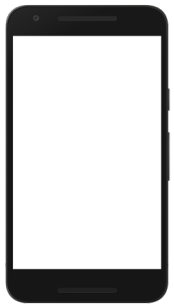
Resolution	OS Version
720x1280	19
768x1280	21
1080x1920	22
1440x2560	23
480x800	24
540x960	25
480x854	

Benchmarks and Setup

ID	*Name*	*Category*	*Version*	*LOC (#K)*
A1	BuildmLearn	Education	2.5.0	23.6
A2	Daily Dozen	Health	10.3	6.3
A3	Kitchen Timer	Tools	1.1.6	4.3
A4	Outlay	Finance	1.1.3	8
A5	Translation Studio	Books	9.0	51.2

Reference Device

LG G3, Android 22



Test Devices

147

Resolution	OS Version
720x1280	19
768x1280	21
1080x1920	22
1440x2560	23
480x800	24
540x960	25
480x854	

RQ1

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

ID	*Test Device*	*Window Model*	*Structural CPI*	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
A1	135	19	0	2	7	14	1
A2	138	22	2	0	0	22	4
A3	129	13	2	3	0	2	1
A4	125	14	0	1	0	17	2
A5	136	17	2	3	0	19	8
			6	*9*	*7*	*74*	*16*

RQ1

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

ID	*Test Device*	*Window Model*	*Structural CPI*	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
A1	135	19	0	2	7	14	1
A2	138	22	2	0	0	22	4
A3	129	13	2	3	0	2	1
A4	125	14	0	1	0	17	2
A5	136	17	2	3	0	19	8
			6	*9*	*7*	*74*	*16*

RQ1

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

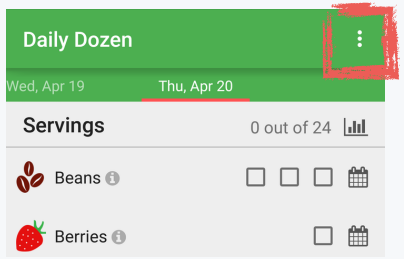
ID	*Test Device*	*Window Model*	*Structural CPI*	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
A1	135	19	0	2	7	14	1
A2	138	22	2	0	0	22	4
A3	129	13	2	3	0	2	1
A4	125	14	0	1	0	17	2
A5	136	17	2	3	0	19	8
			6	*9*	*7*	*74*	*16*

RQ1

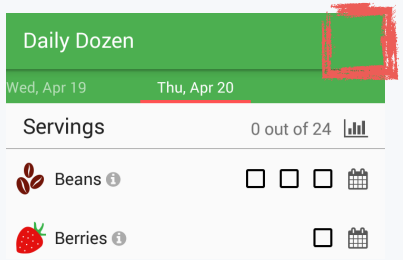
RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

Example

Reference Device



Test Device



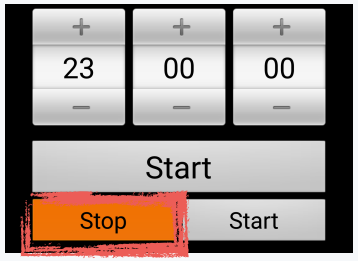
Structural CPI	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
0	2	7	14	1
2	0	0	22	4
2	3	0	2	1
0	1	0	17	2
2	3	0	19	8
6	*9*	*7*	*74*	*16*

RQ1

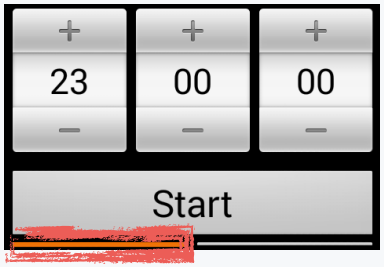
RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

Example

Reference Device



Test Device



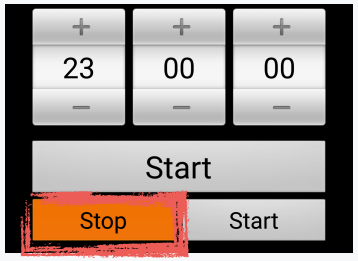
Structural CPI	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
0	2	7	14	1
2	0	0	22	4
2	3	0	2	1
0	1	0	17	2
2	3	0	19	8
6	*9*	*7*	*74*	*16*

RQ1

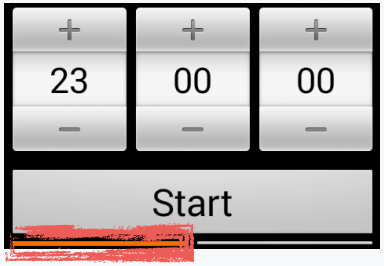
RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

Example

Reference Device



Test Device



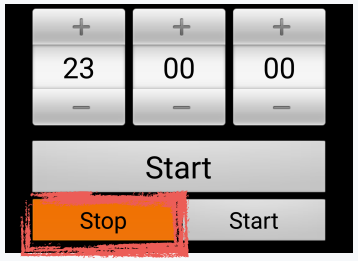
Structural CPI	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
0	2	7	14	1
2	0	0	22	4
2	3	0	2	1
0	1	0	17	2
2	3	0	19	8
6	*9*	*7*	*74*	*16*

RQ1

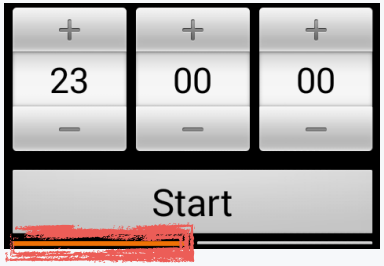
RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

Example

Reference Device



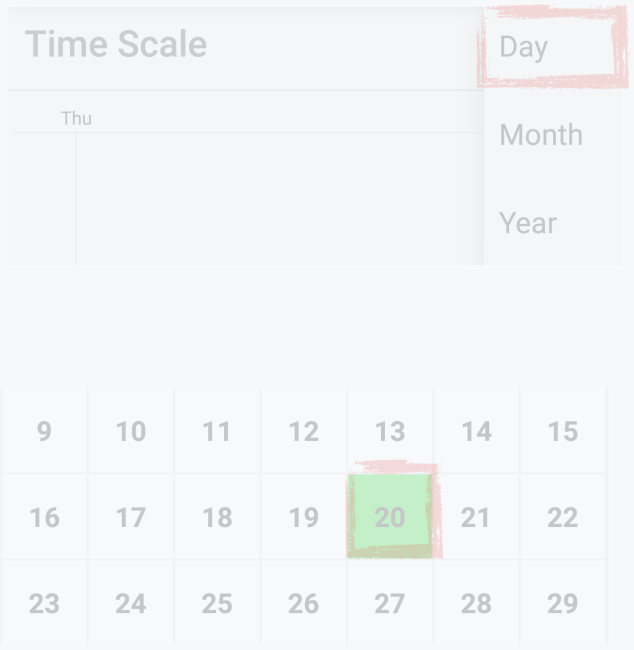
Test Device

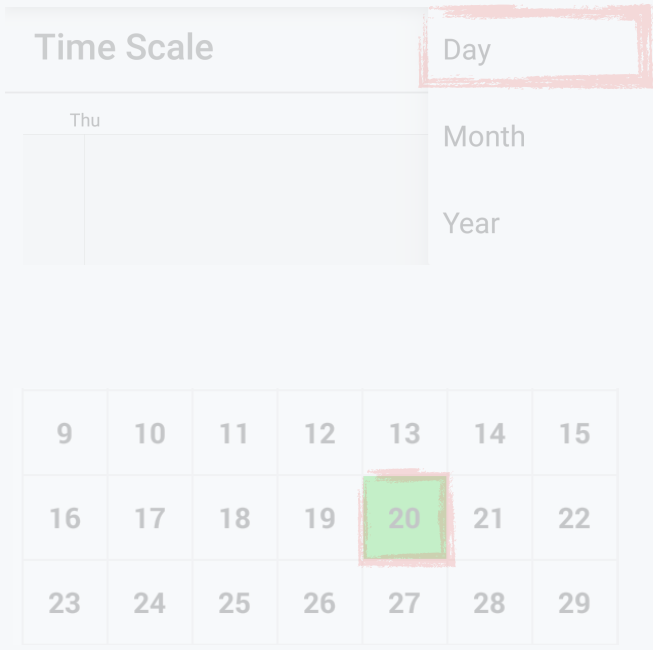


Structural CPI	*Functional CPI*	*Version CPI*	*Cosmetic CPI*	*False Positive*
0	2	7	14	1
2	0	0	22	4
2	3	0	2	1
0	1	0	17	2
2	3	0	19	8
6	*9*	*7*	*74*	*16*

RQ1

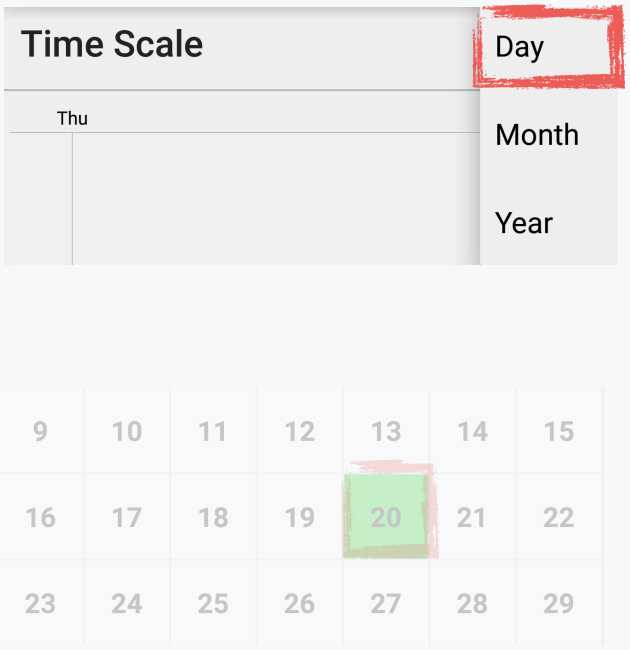
RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

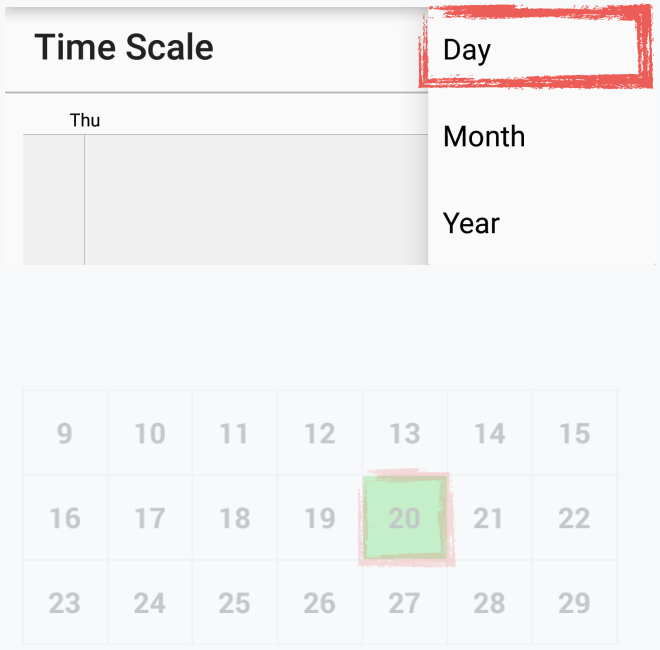




RQ1

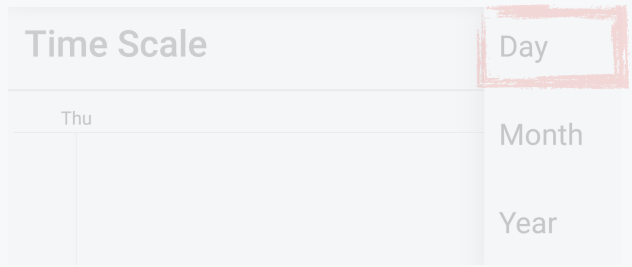
RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?



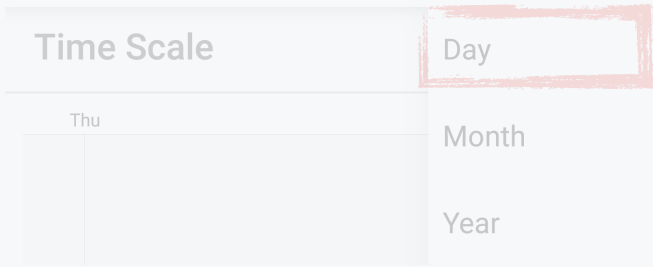


RQ1

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?



9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29



9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

Evaluation Summary

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

DIFFDROID can detect CPIs in mobile applications while reporting a limited number of false positives

RQ2: What is the cost of running DiffDroid?

DIFFDROID can run overnight for the cases considered

RQ3: Are there similarities among devices exhibiting CPIs?

Devices that are more problematic have low values for resolution and density. However, considering testing devices solely based on resolution and density would have not allowed us to identify all the inconsistencies reported.

Evaluation Summary

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

DIFFDROID can detect CPIs in mobile applications while reporting a limited number of false positives

RQ2: What is the cost of running DiffDroid?

DIFFDROID can run overnight for the cases considered

RQ3: Are there similarities among devices exhibiting CPIs?

Devices that are more problematic have low values for resolution and density. However, considering testing devices solely based on resolution and density would have not allowed us to identify all the inconsistencies reported.

Evaluation Summary

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

DIFFDROID can detect CPIs in mobile applications while reporting a limited number of false positives

RQ2: What is the cost of running DiffDroid?

DIFFDROID can run overnight for the cases considered

RQ3: Are there similarities among devices exhibiting CPIs?

Devices that are more problematic have low values for resolution and density. However, considering testing devices solely based on resolution and density would have not allowed us to identify all the inconsistencies reported.

Evaluation Summary

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

DIFFDROID can detect CPIs in mobile applications while reporting a limited number of false positives

RQ2: What is the cost of running DiffDroid?

DIFFDROID can run overnight for the cases considered

RQ3: Are there similarities among devices exhibiting CPIs?

Devices that are more problematic have low values for resolution and density. However, considering testing devices solely based on resolution and density would have not allowed us to identify all the inconsistencies reported.

Future Work

Additional user study

API based differential analysis

Multi-class classifier approach

CPIs repair technique

Future Work

Additional user study

API based differential analysis

Multi-class classifier approach

CPIs repair technique

Future Work

Additional user study

API based differential analysis

Multi-class classifier approach

CPIs repair technique

Future Work

Additional user study

API based differential analysis

Multi-class classifier approach

CPIs repair technique

Future Work

Additional user study

API based differential analysis

Multi-class classifier approach

CPIs repair technique

Summary

Fragmentation

Device

System

Screen

News

Jul 2016

"Due to the wide array of available Android devices, we are targeting our support to a select number of Android devices to continue improving our overall Salesforce1 for Android user experience"

DiffDroid Overview

Empirical Evaluation

Research Questions:

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DiffDroid?

RQ3: Are there similarities among devices exhibit- ing CPIs?

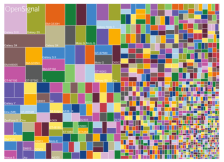
Future Work

- Additional user study
- API based differential analysis
- Multi-class classifier approach
- CPIs repair technique

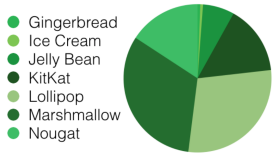
Summary

Fragmentation

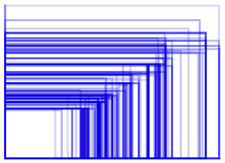
Device



System



Screen



News

Jul 2016

"Due to the wide array of available Android devices, we are targeting our support to a select number of Android devices to continue improving our overall Salesforce1 for Android user experience"

DiffDroid Overview

Empirical Evaluation

Research Questions:

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DiffDroid?

RQ3: Are there similarities among devices exhibit- ing CPIs?

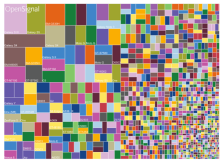
Future Work

- Additional user study
- API based differential analysis
- Multi-class classifier approach
- CPIs repair technique

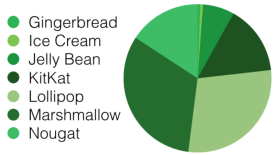
Summary

Fragmentation

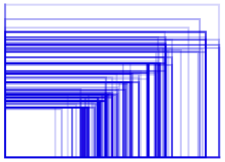
Device



System



Screen

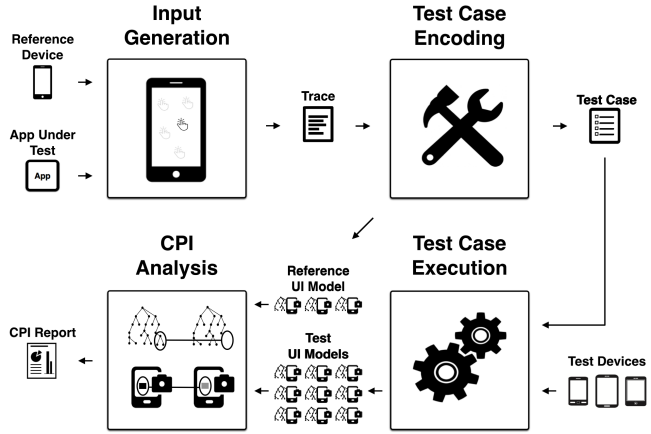


News

Jul 2016

"Due to the wide array of available Android devices, we are targeting our support to a select number of Android devices to continue improving our overall Salesforce1 for Android user experience"

DiffDroid Overview



Empirical Evaluation

Research Questions:

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DiffDroid?

RQ3: Are there similarities among devices exhibit- ing CPIs?

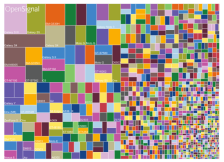
Future Work

- Additional user study
- API based differential analysis
- Multi-class classifier approach
- CPIs repair technique

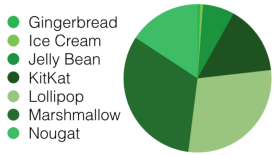
Summary

Fragmentation

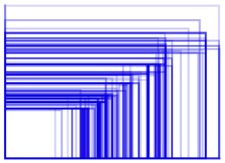
Device



System



Screen

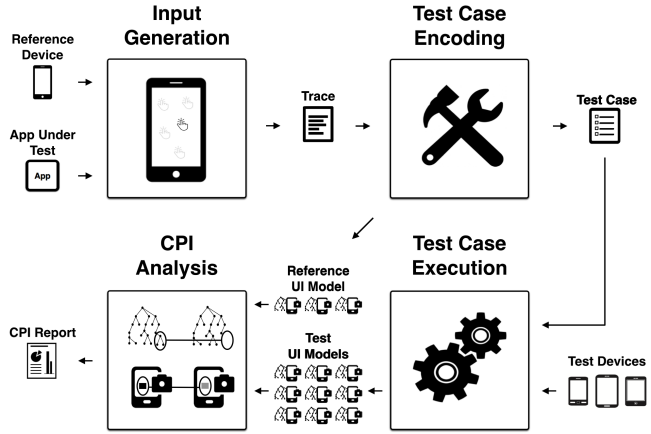


News

Jul 2016

"Due to the wide array of available Android devices, we are targeting our support to a select number of Android devices to continue improving our overall Salesforce1 for Android user experience"

DiffDroid Overview



Empirical Evaluation

Research Questions:

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DiffDroid?

RQ3: Are there similarities among devices exhibit- ing CPIs?

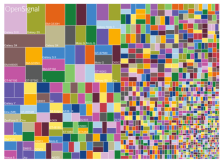
Future Work

- Additional user study
- API based differential analysis
- Multi-class classifier approach
- CPIs repair technique

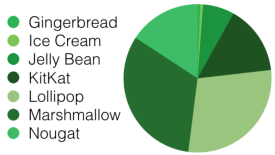
Summary

Fragmentation

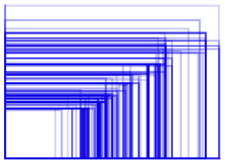
Device



System



Screen

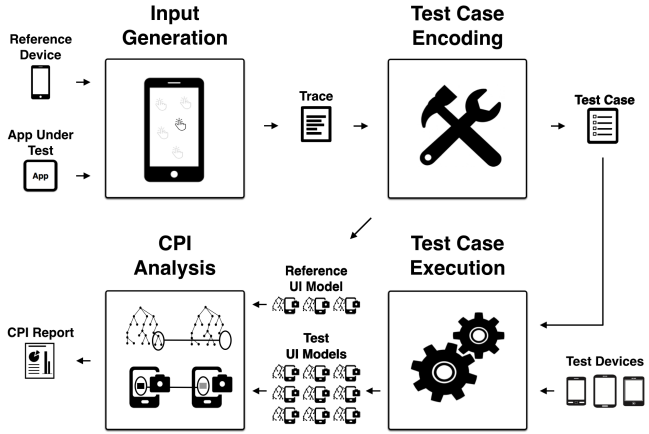


News

Jul 2016

"Due to the wide array of available Android devices, we are targeting our support to a select number of Android devices to continue improving our overall Salesforce1 for Android user experience"

DiffDroid Overview



Empirical Evaluation

Research Questions:

RQ1: Can DiffDroid detect cross-platform inconsistencies in mobile applications while reporting a limited number of false positives?

RQ2: What is the cost of running DiffDroid?

RQ3: Are there similarities among devices exhibit- ing CPIs?

Future Work

- Additional user study
- API based differential analysis
- Multi-class classifier approach
- CPIs repair technique