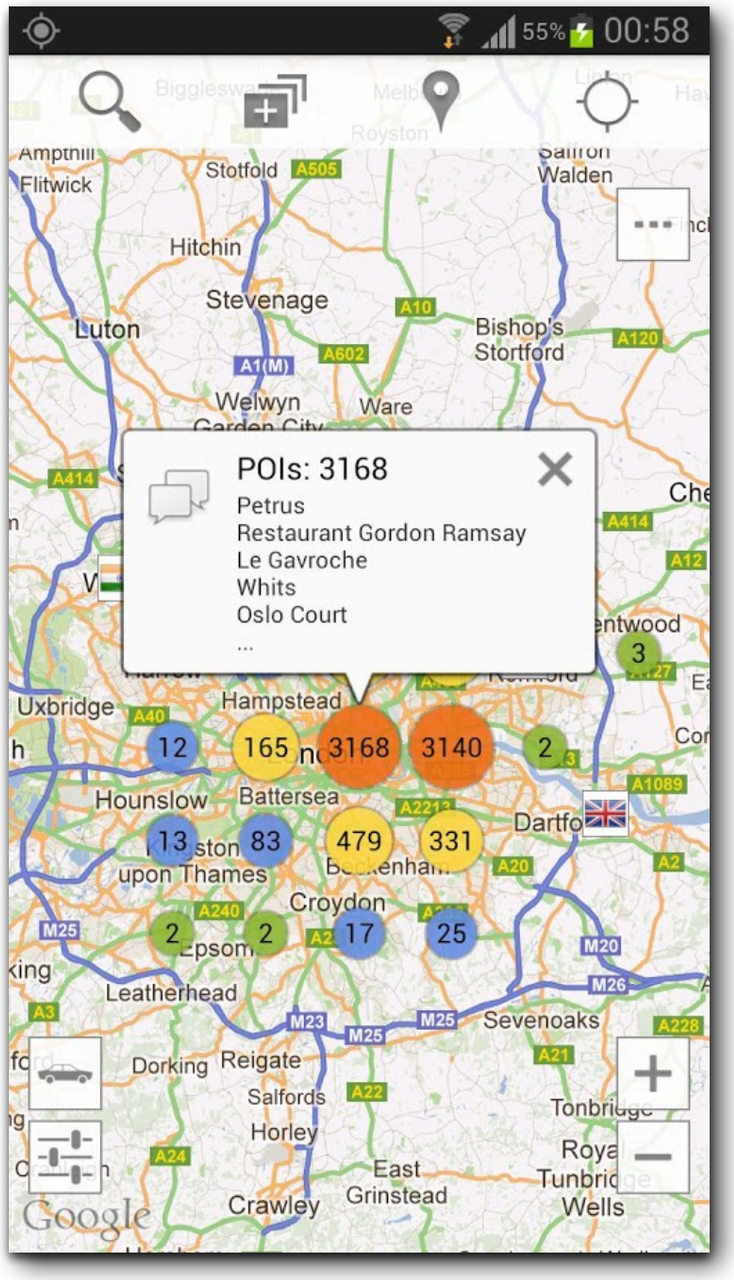
CHABADA:
Checking App Behavior Against App Descriptions

Alessandra Gorla
Saarland University, Germany

joint work with Konstantin Kuznetsov, Ilaria Tavecchia, Florian Gross and Andreas Zeller


S

London Restaurants

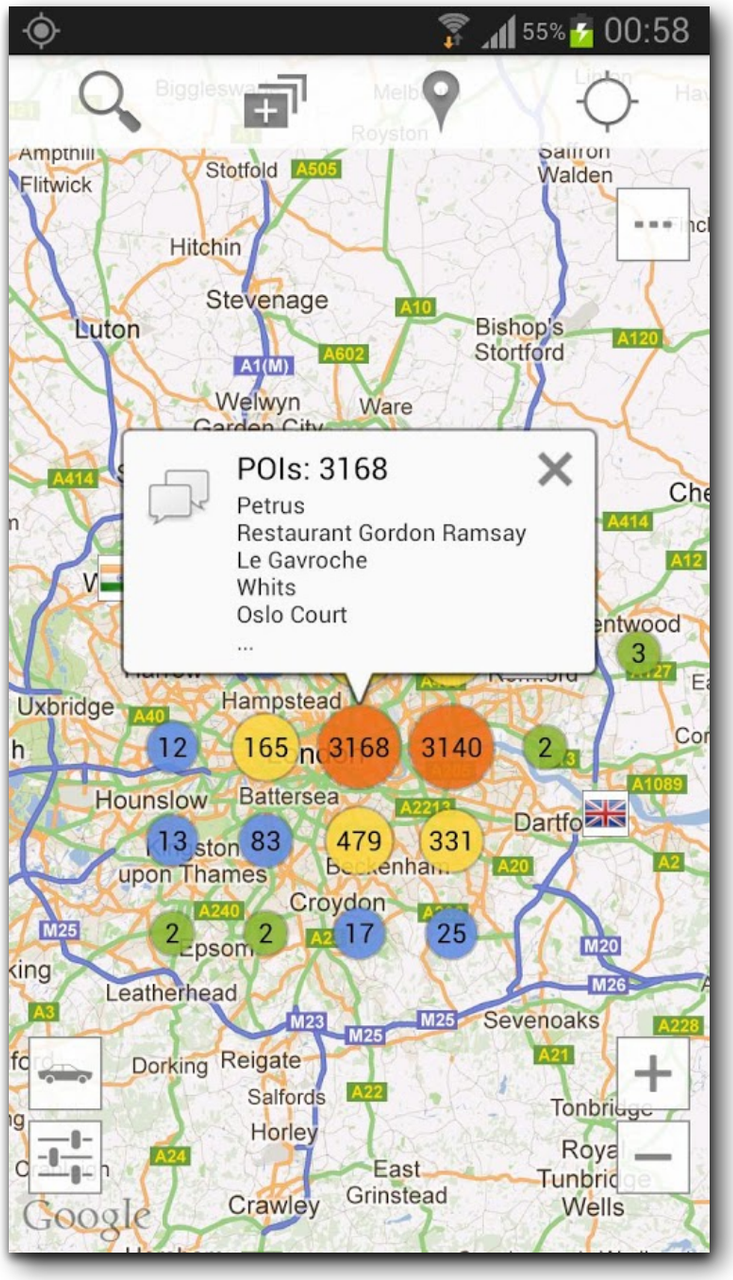


Looking for a restaurant, a bar, a pub or just to have fun in London? Search no more! This application has all the information you need:

• You can search for every type of food you want: french, british, chinese, indian etc.
• You can use it if you are in a car, on a bicycle or walking
• You can view all objectives on the map
• You can search objectives
• You can view objectives near you
• You can view directions (visual route, distance and duration)
• You can use it with Street View
• You can use it with Navigation

Keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

London Restaurants



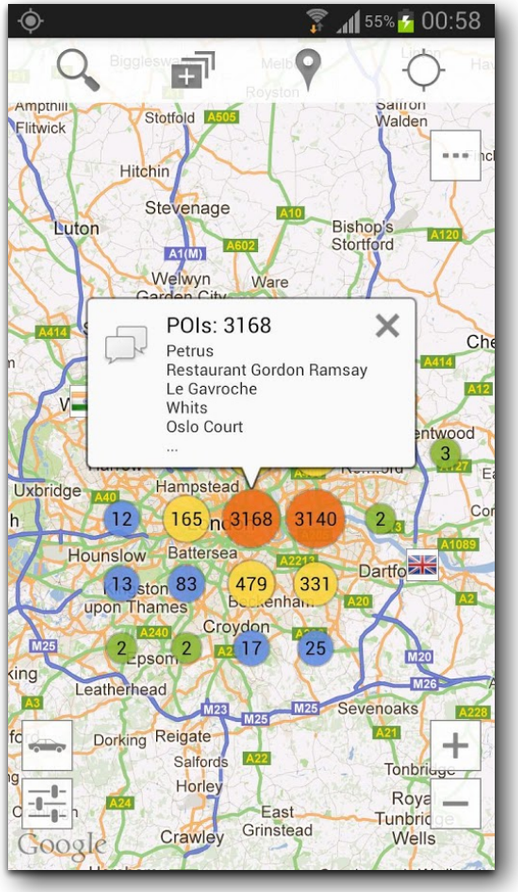
Looking for a restaurant, a bar, a pub or just to have fun in London? Search no more! This application has all the information you need:

• You can search for every type of food you want: french, british, chinese, indian etc.
• You can use it if you are in a car, on a bicycle or walking
• You can view all objectives on the map
• You can search objectives
• You can view objectives near you
• You can view directions (visual route, distance and duration)
• You can use it with Street View
• You can use it with Navigation

Keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

Also sends out *account info*
Also sends out *mobile phone number*
Also sends out *your device ID*

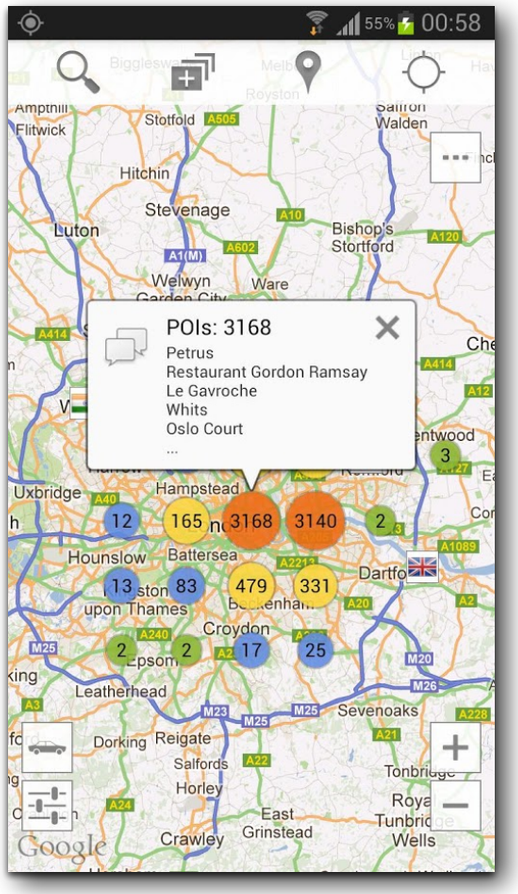
What is malicious?



London Restaurants

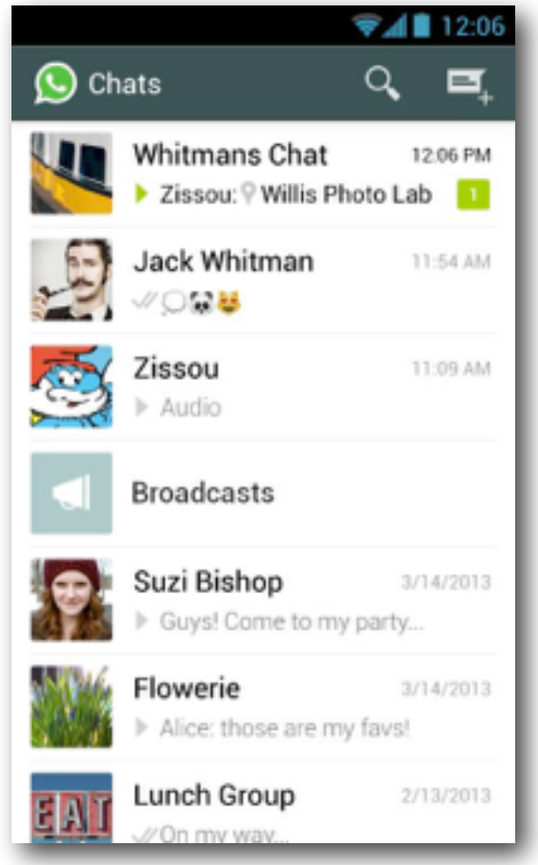
Also sends out *account info*
Also sends out *mobile phone number*
Also sends out *your device ID*

What is malicious?



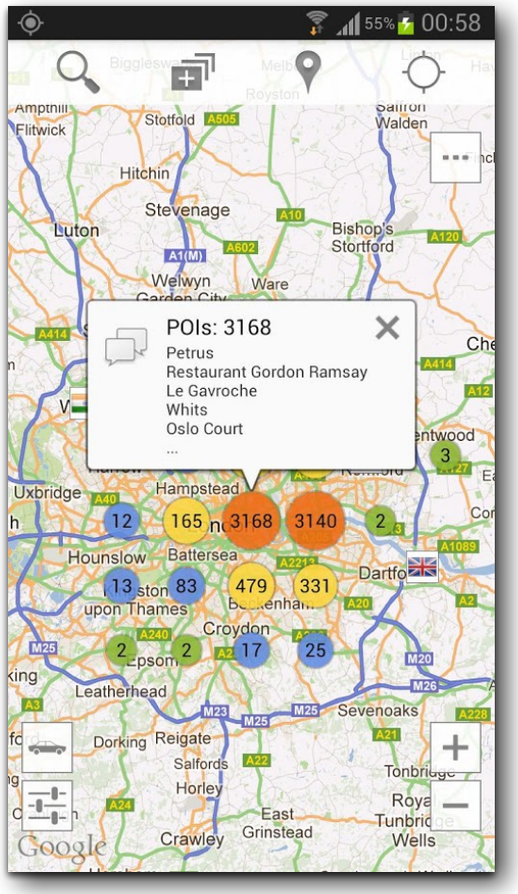
London Restaurants

Also sends out *account info*
Also sends out *mobile phone number*
Also sends out *your device ID*



WhatsApp messenger

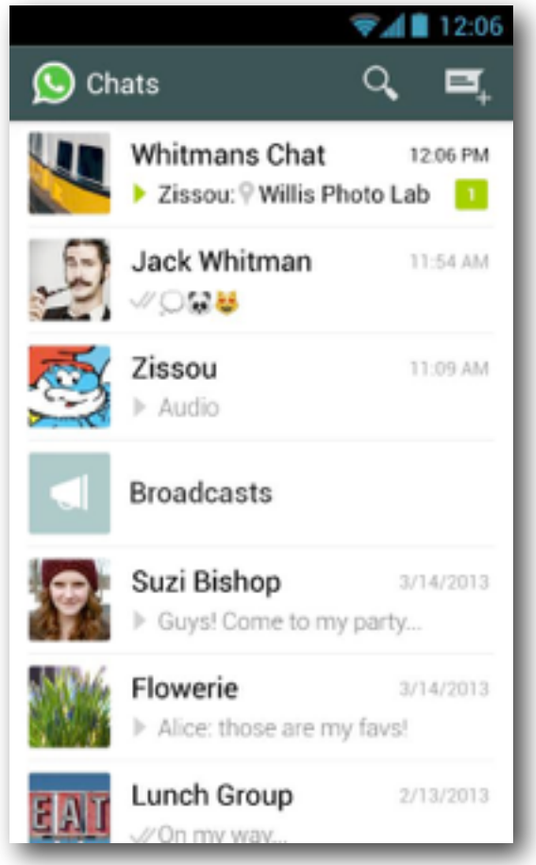
What is malicious?



London Restaurants

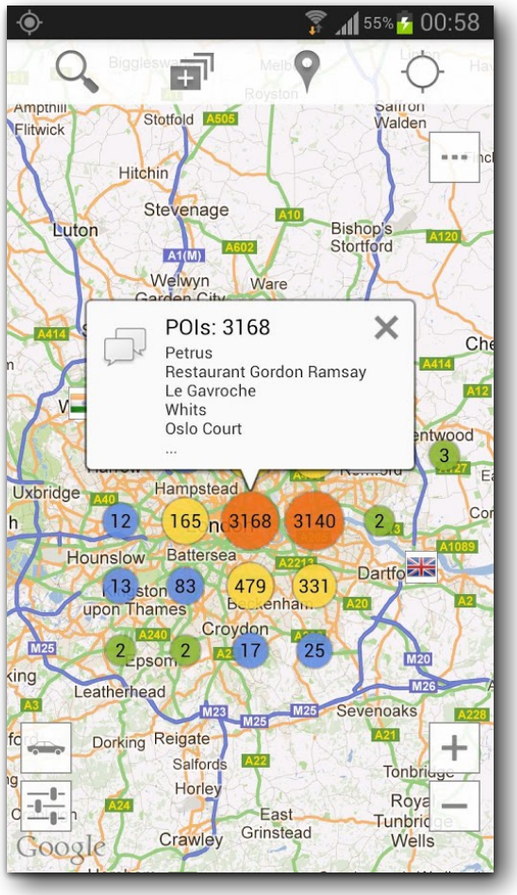
Also sends out *account info*
Also sends out *mobile phone number*
Also sends out *your device ID*

Also sends out account info
Also sends out mobile phone number
Also sends out your device ID



WhatsApp messenger

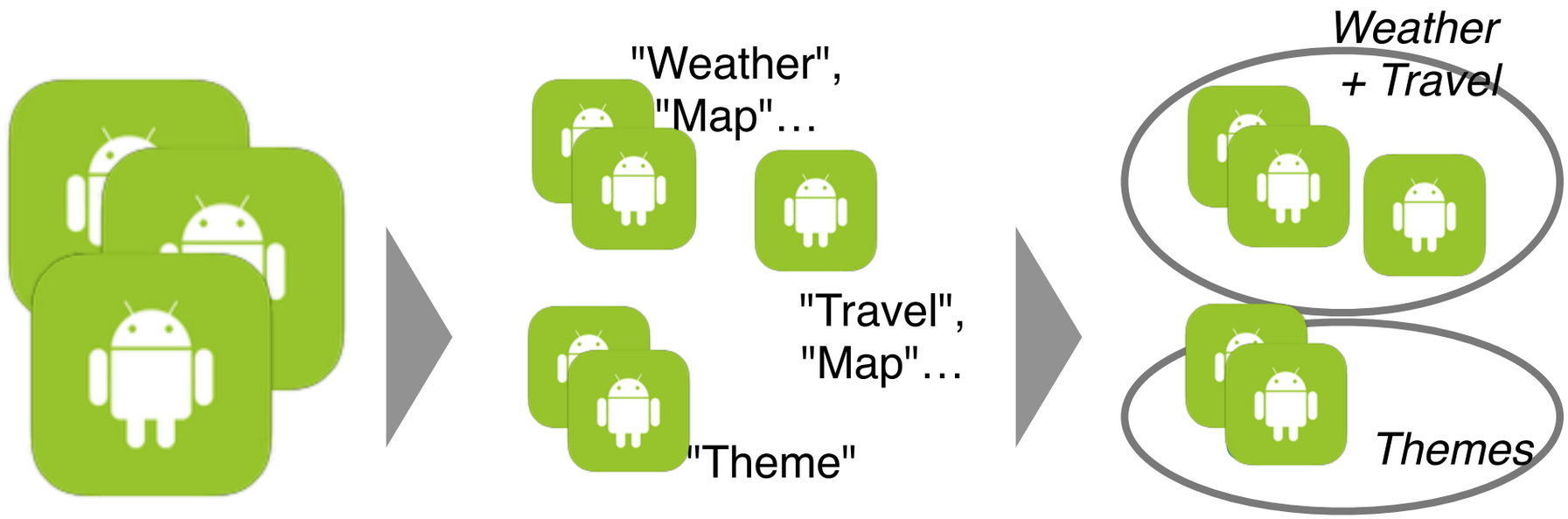
What is normal?



London Restaurants

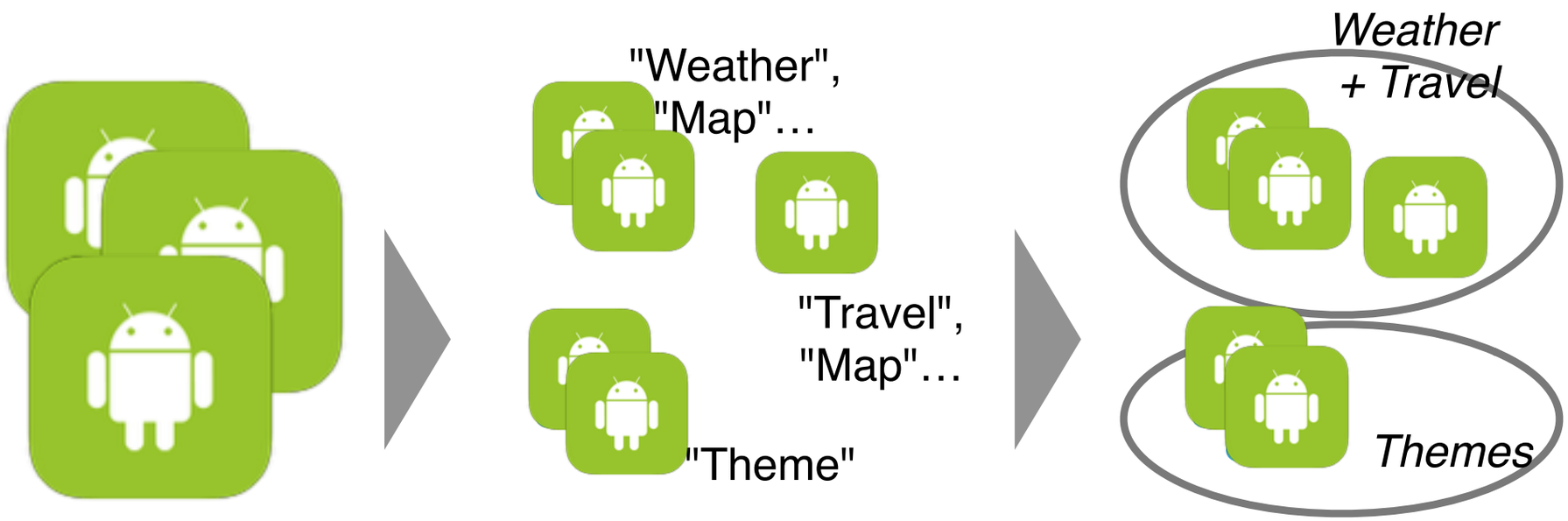
- "London Restaurants" is a "travel" app
- For "travel" apps, sending account infos is *abnormal*
- For "messaging" apps, this is far *more likely*

CHABADA

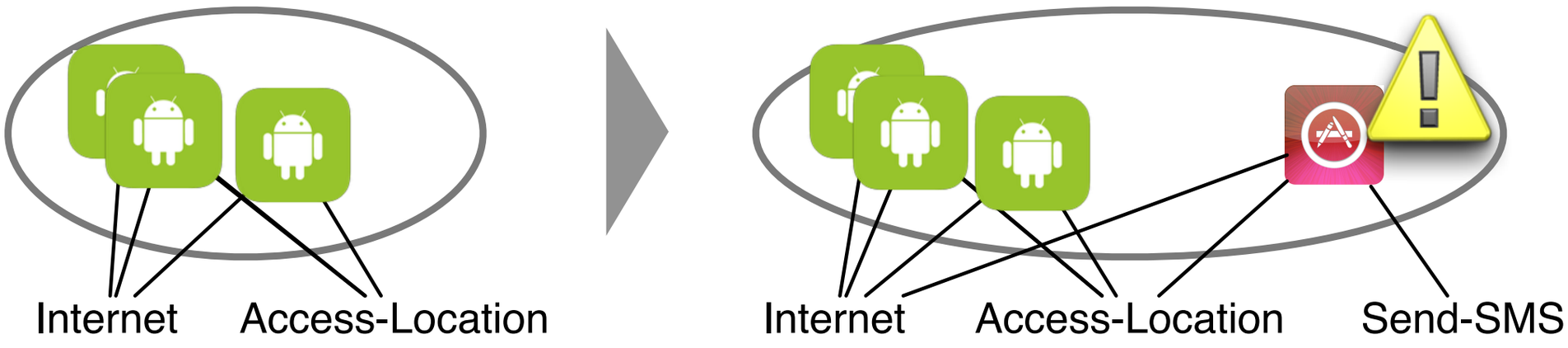


1. App collection **2. Topics** **3. Clusters**

CHABADA

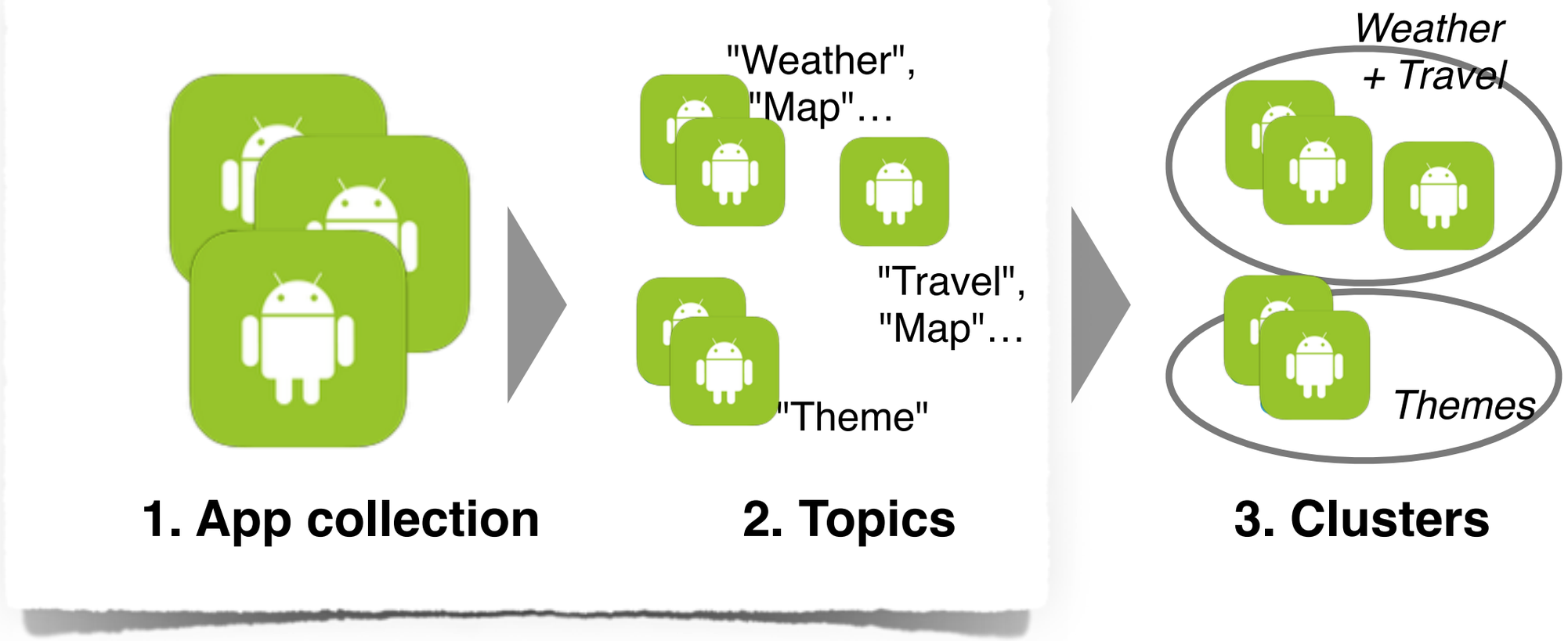


1. App collection **2. Topics** **3. Clusters**

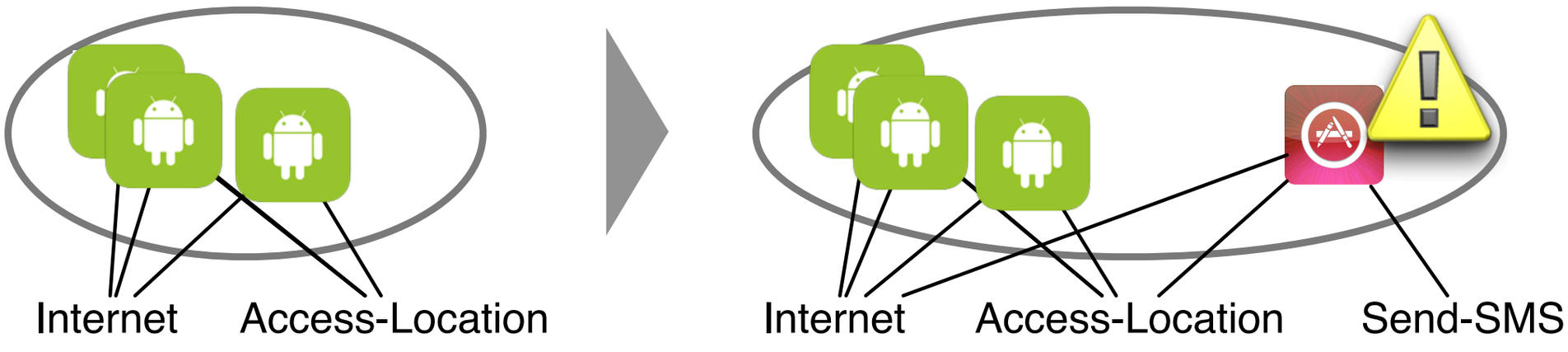


4. APIs **5. Outliers**

CHABADA

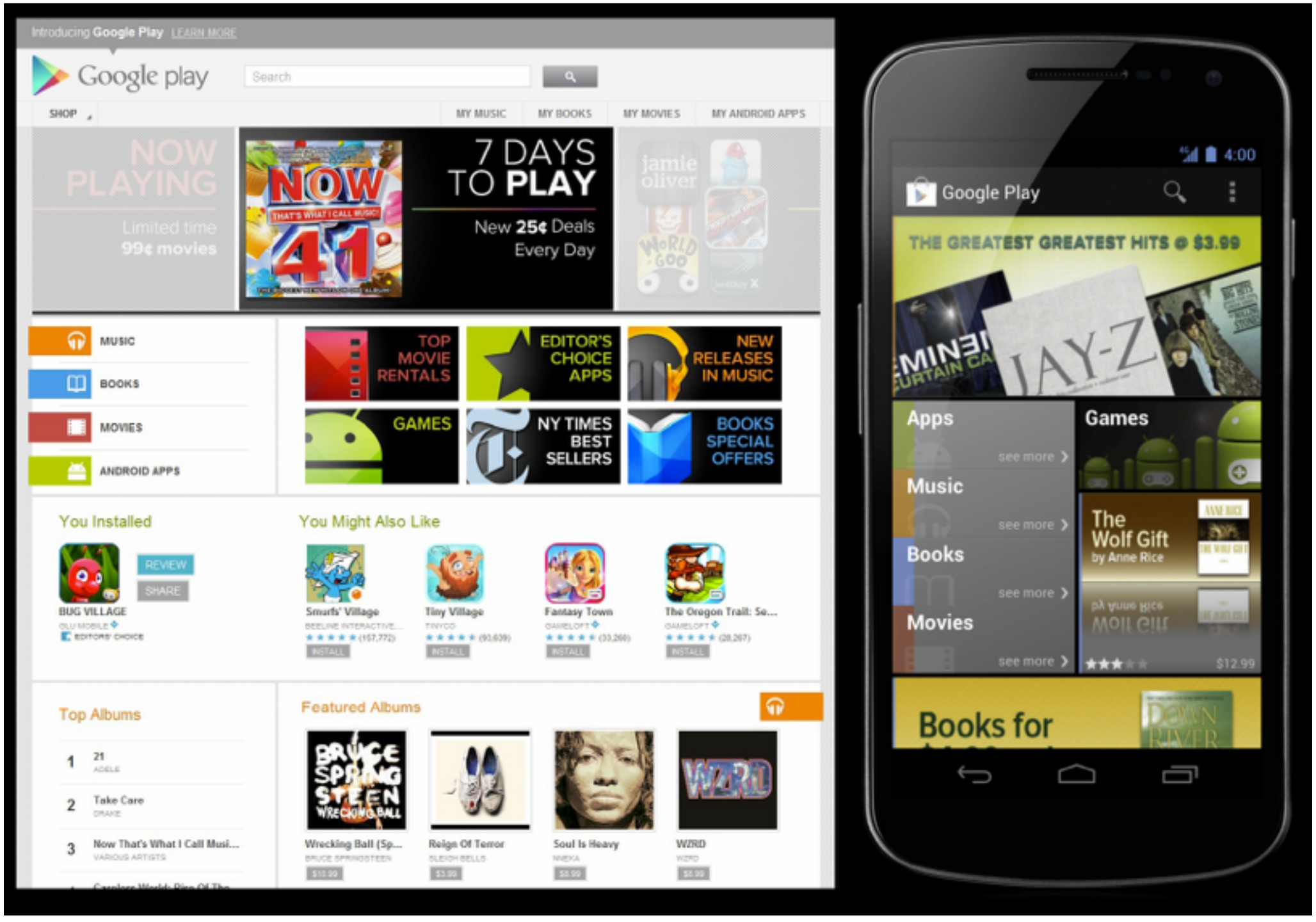


1. App collection **2. Topics** **3. Clusters**

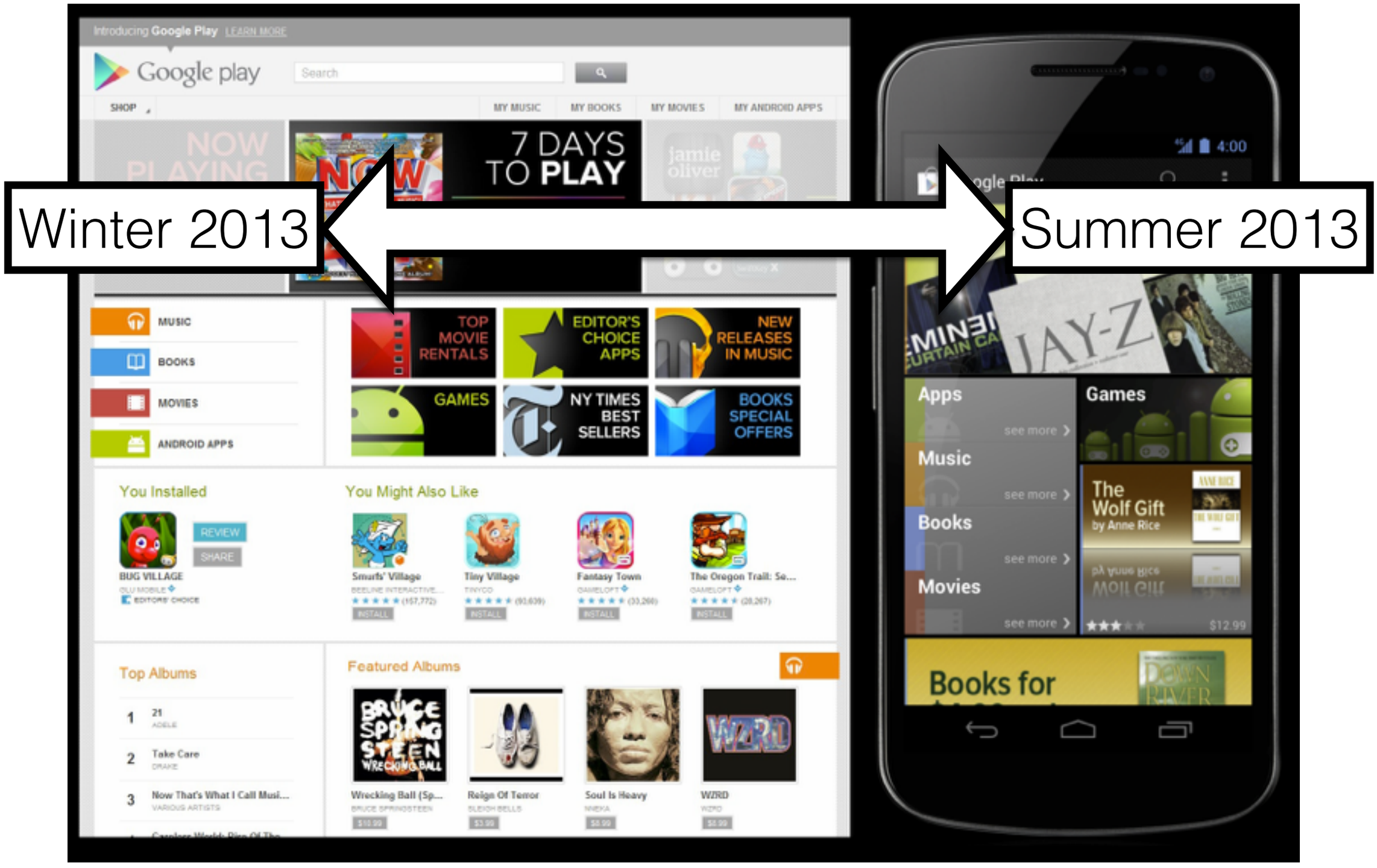


4. APIs **5. Outliers**

Apps collection

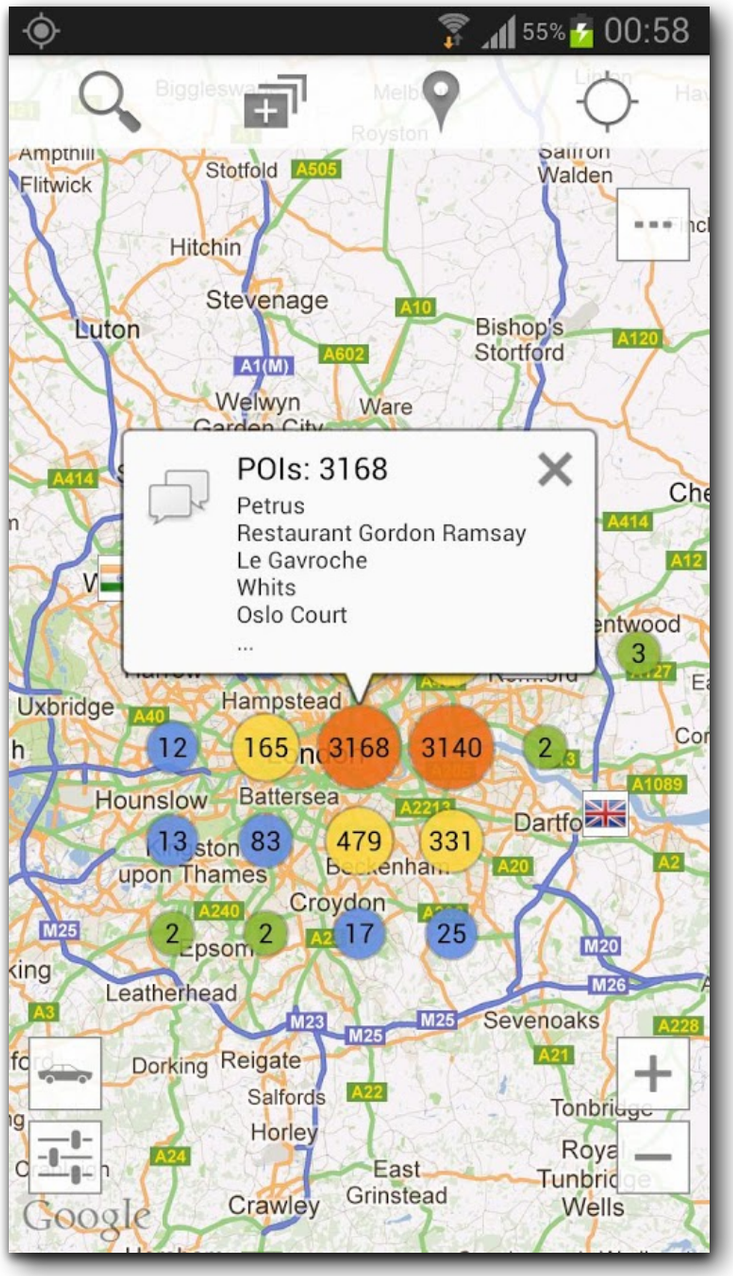


Apps collection



Apps collection

Stemming

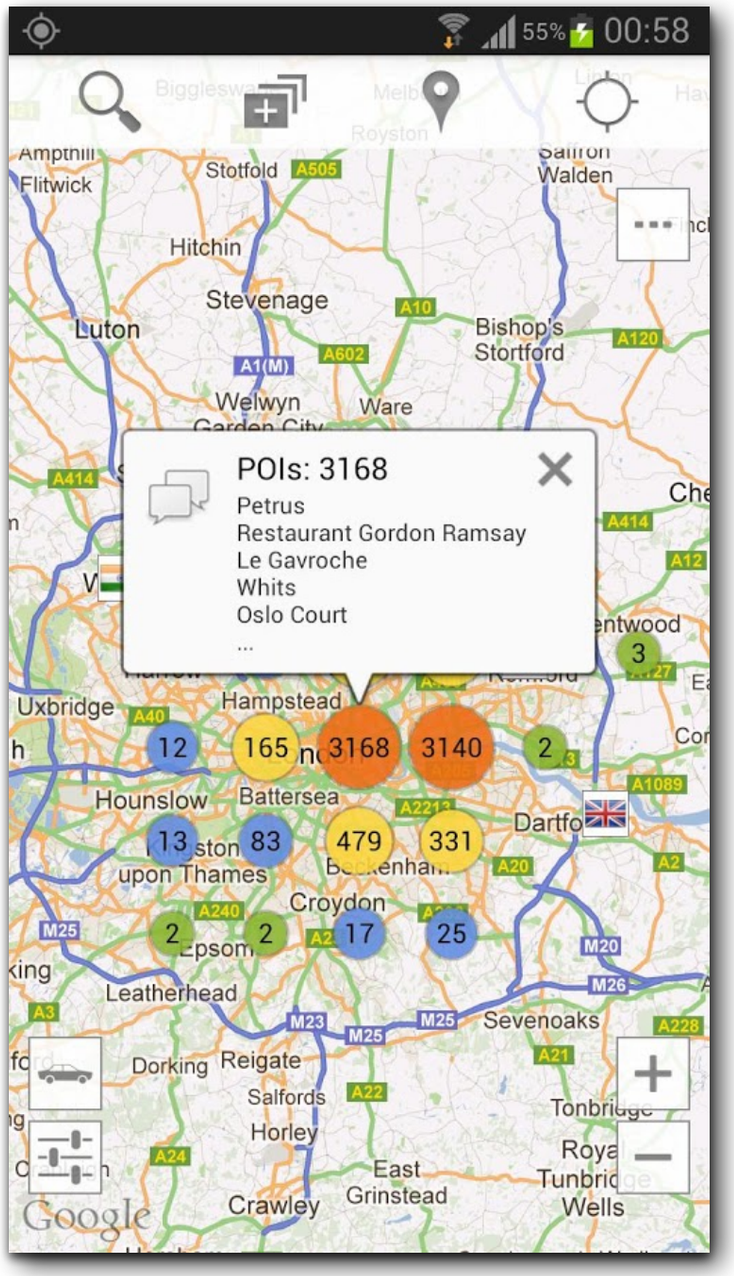


looking for a restaurant, a bar, a pub or just to have fun in london? search no more! this application has all the information you need:

- you can search for every type of food you want: french, british, chinese, indian etc.
- you can use it if you are in a car, on a bicycle or walking
- you can view all objectives on the map
- you can search objectives
- you can view objectives near you
- you can view directions (visual route, distance and duration)
- you can use it with street view
- you can use it with navigation

keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

Stemming



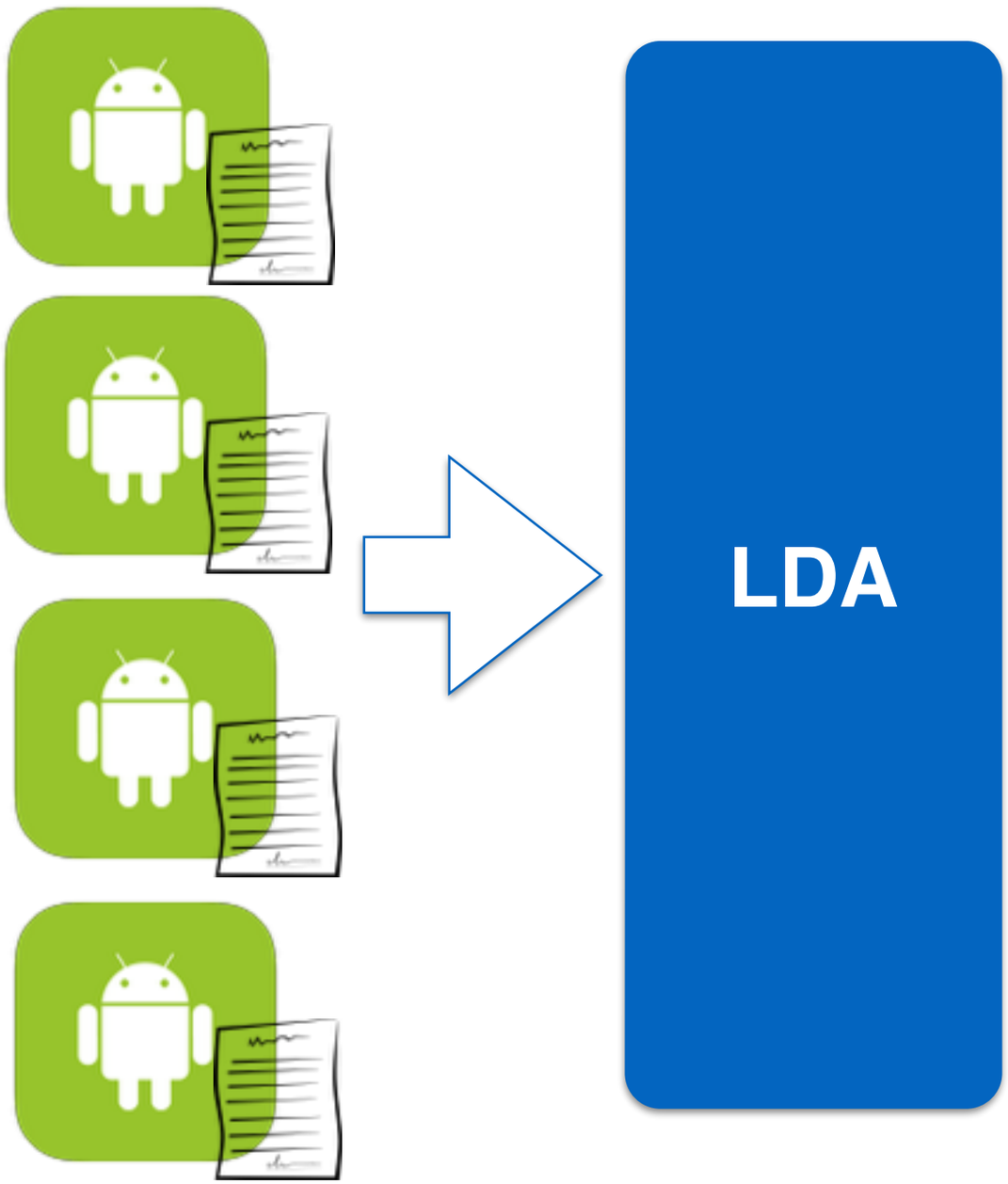
looking for a restaurant, a bar, a pub or just to have fun in london? search no more! this application has all the information you need:

- you can search for everi type of food you want: french, british, chinese, indian etc.
- you can use it if you are in a car, on a bicycle or walking
- you can view all objectives on the map
- you can search objectives
- you can view objectives near you
- you can view directions (visual route, distance and duration)
- you can use it with street view
- you can use it with navigation

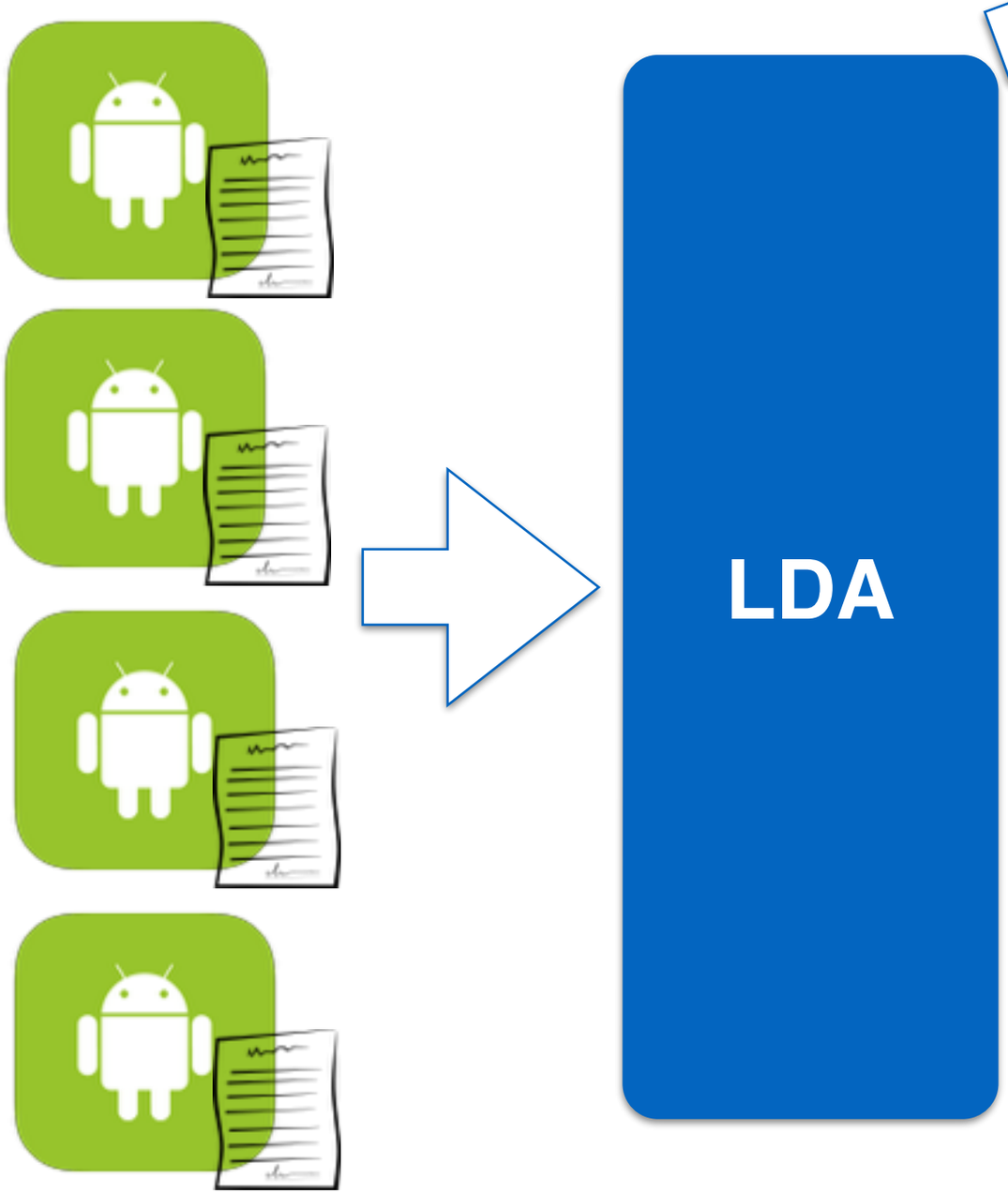
keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

Topic Analysis

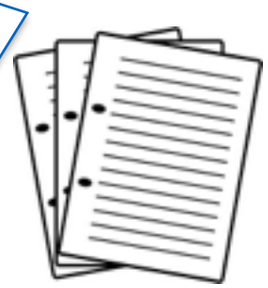
Topic Analysis



Topic Analysis



T1: [map, navigation, street,tour, …]

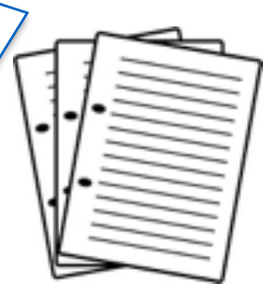


T2: [weight, body, exercise,run …]

Topic Analysis

LDA

T1: [map, navigation, street,tour, …]



T2: [weight, body, exercise,run …]

T1: 80%, T2: 20%

T2:90%, T1: 10%

T1: 70%, T2: 30%

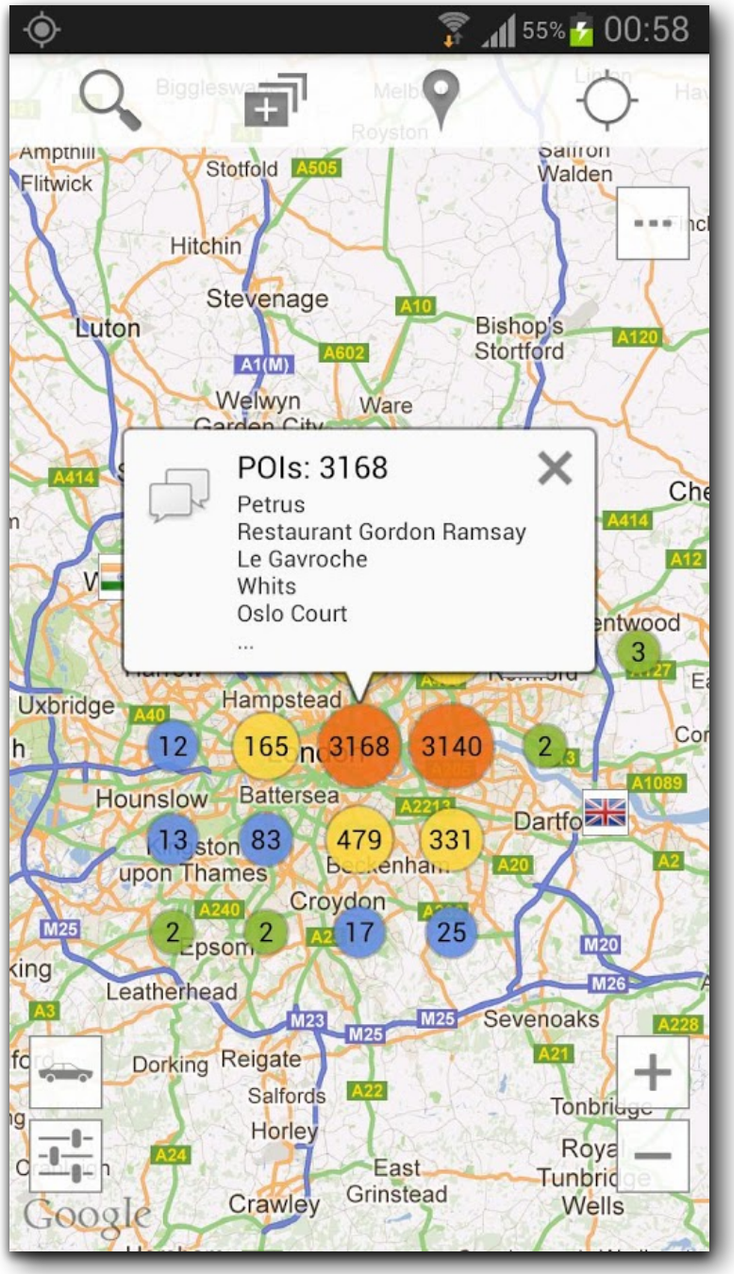
T2:80%, T1: 20%

Topics

Id	Assigned Name	Most Representative Words (stemmed)
0	“personalize”	galaxi, nexu, device, screen, effect, instal, customis
1	“game and cheat sheets”	game, video, page, cheat, link, tip, trick
2	“money”	slot, machine, money, poker, currenc, market, trade, stock, casino coin, finance
3	“tv”	tv, channel, countri, live, watch, germani, nation, bbc, newspap
4	“music”	music, song, radio, play, player, listen
5	“holidays” and religion	christmas, halloween, santa, year, holiday, islam, god
6	“navigation and travel”	map, inform, track, gps, navig, travel
7	“language”	language, word, english, learn, german, translat
8	“share”	email, ad, support, facebook, share, twitter, rate, suggest
9	“weather and stars”	weather, forecast, locate, temperatur, map, city, light
10	“files and video”	file, download, video, media, support, manage, share, view, search

13	“design and art”	life, peopl, natur, form, feel, learn, art, design, uniqu, effect, modern
14	“food and recipes”	recip, cake, chicken, cook, food
15	“personalize”	theme, launcher, download, install, icon, menu
16	“health”	weight, bodi, exercise, diet, workout, medic
17	“travel”	citi, guid, map, travel, flag, countri, attract
18	“kids and bodies”	kid, anim, color, girl, babi, pictur, fun, draw, design, learn
19	“ringtones and sound”	sound, rington, alarm, notif, music
20	“game”	game, plai, graphic, fun, jump, level, ball, 3d, score
21	“search and browse”	search, icon, delet, bookmark, link, homepag, shortcut, browser
22	“battle games”	story, game, monster, zombi, war, battle
23	“settings and utils”	screen, set, widget, phone, batteri
24	“sports”	team, football, leagu, player, sport, basketbal
25	“wallpapers”	wallpap, live, home, screen, background, menu
26	“connection”	device, connect, network, wifi, blootooth, internet, remot, server
27	“policies and ads”	live, ad, home, applovin, notif, data, polici, privacy, share, airpush, advertis
28	“popular media”	seri, video, film, album, movi, music, award, star, fan, show, gangnam, top, bieber
29	“puzzle and card games”	game, plai, level, puzzl, player, score, challeng, card

London Restaurant Topics



look london restaur search bar pub just applic fun

inform can search need everi type food want french

british chines indian etc car bicycl walk

can us can view object map visual rout

can search object search can view distanc

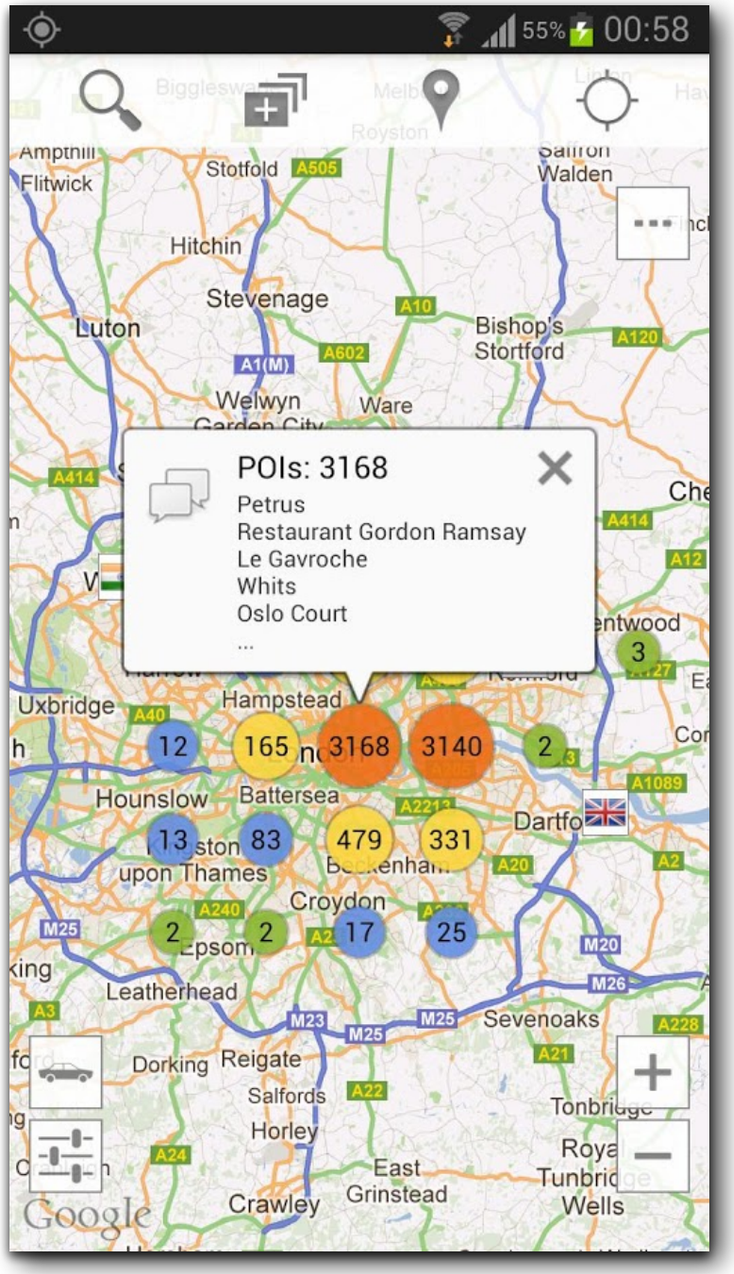
durat can view direct object near

can us street view can us navig

keyword london restaur bar pub food view

breakfast lunch dinner meal eat supper street navig

London Restaurant Topics



look london restaur search bar pub just applic fun

inform can search need everi type food want french

british chines indian etc car bicycl walk

can us can view object map visual rout

can search object search can view distanc

durat can view direct object near

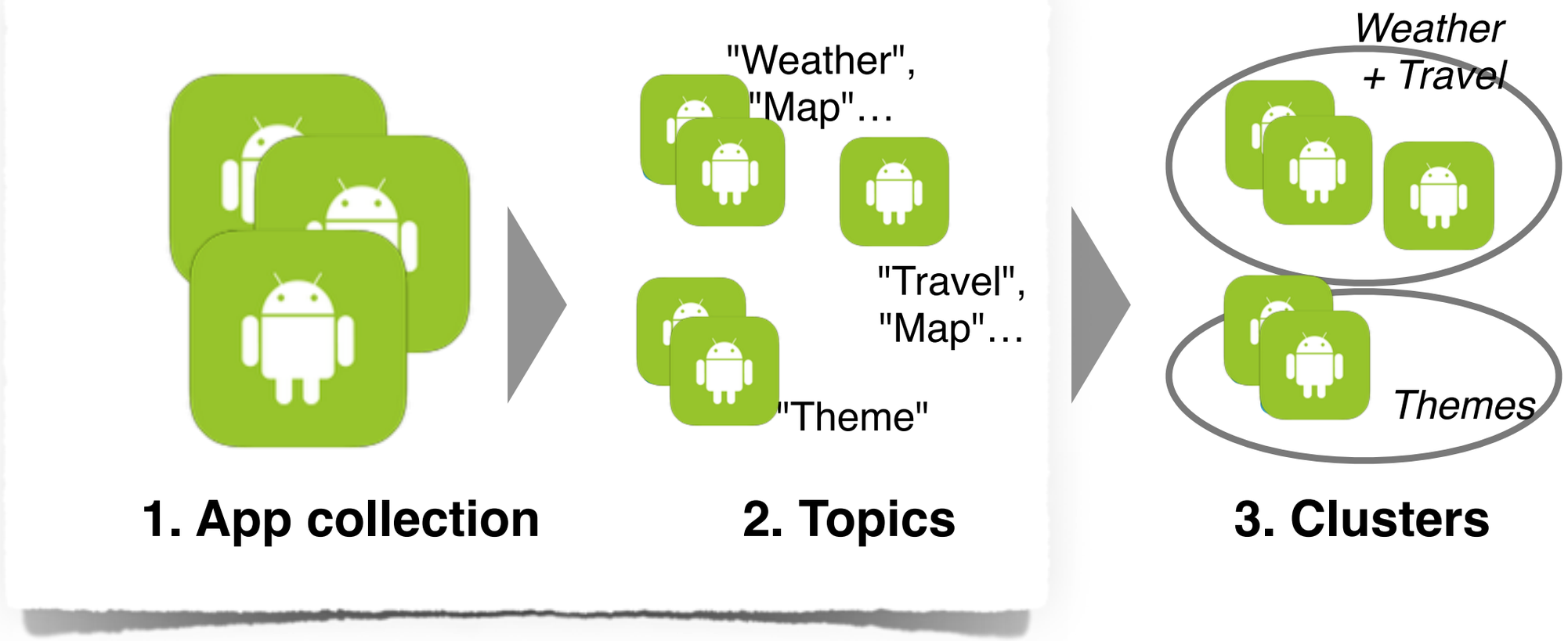
can us street view can us navig

keyword london restaur bar pub food view

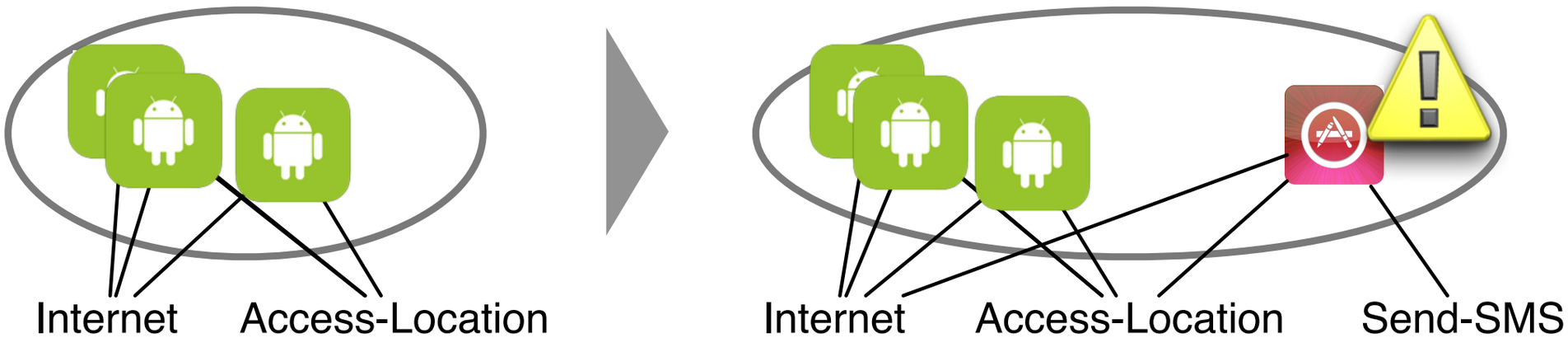
breakfast lunch dinner meal eat supper street navig

“navigation and travel” (59.8%)
“food and recipes” (19.9%)
“travel” (14.0%)

CHABADA



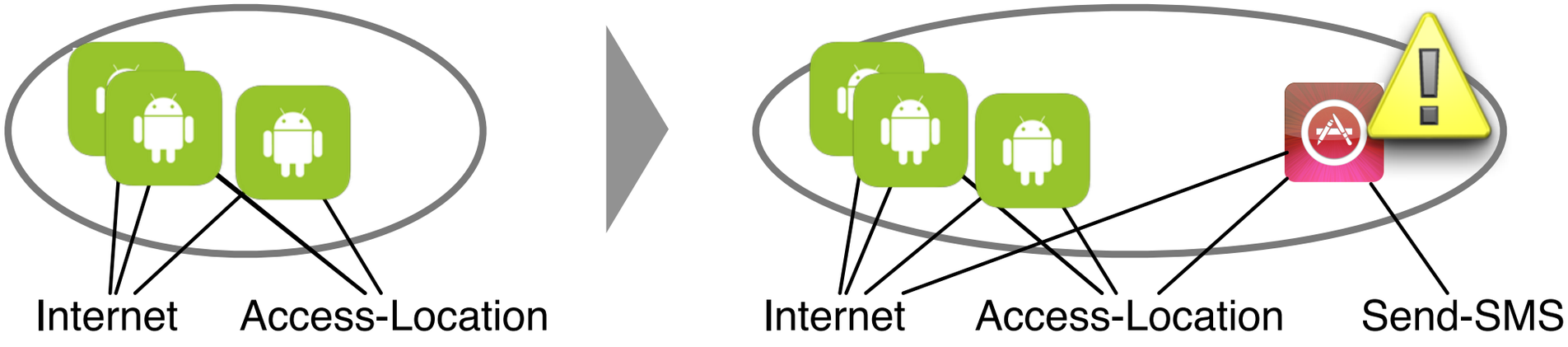
1. App collection **2. Topics** **3. Clusters**



4. APIs **5. Outliers**

CHABADA

1. App collection **2. Topics** **3. Clusters**



4. APIs **5. Outliers**

Clustering

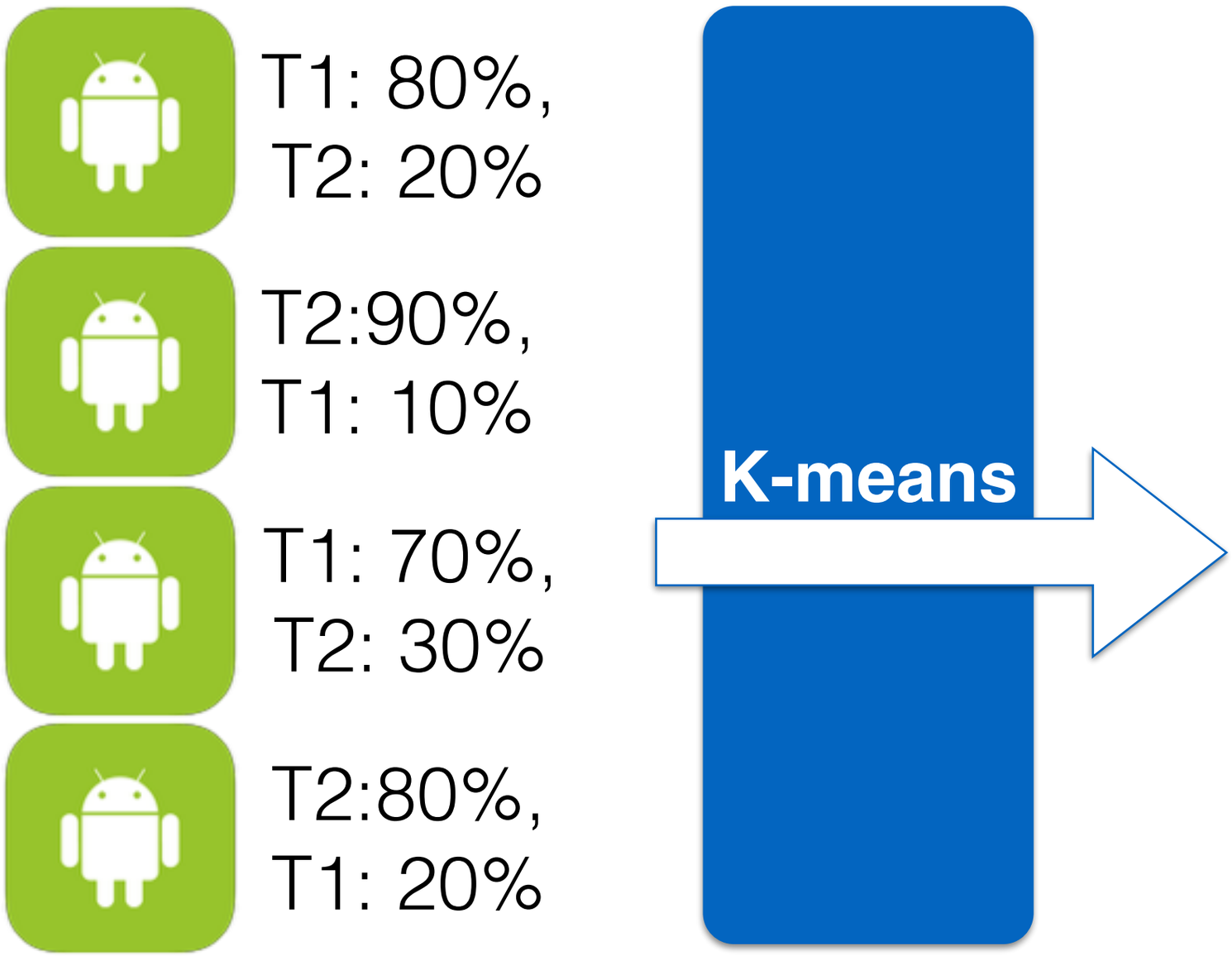
T1: 80%,
T2: 20%

T2:90%,
T1: 10%

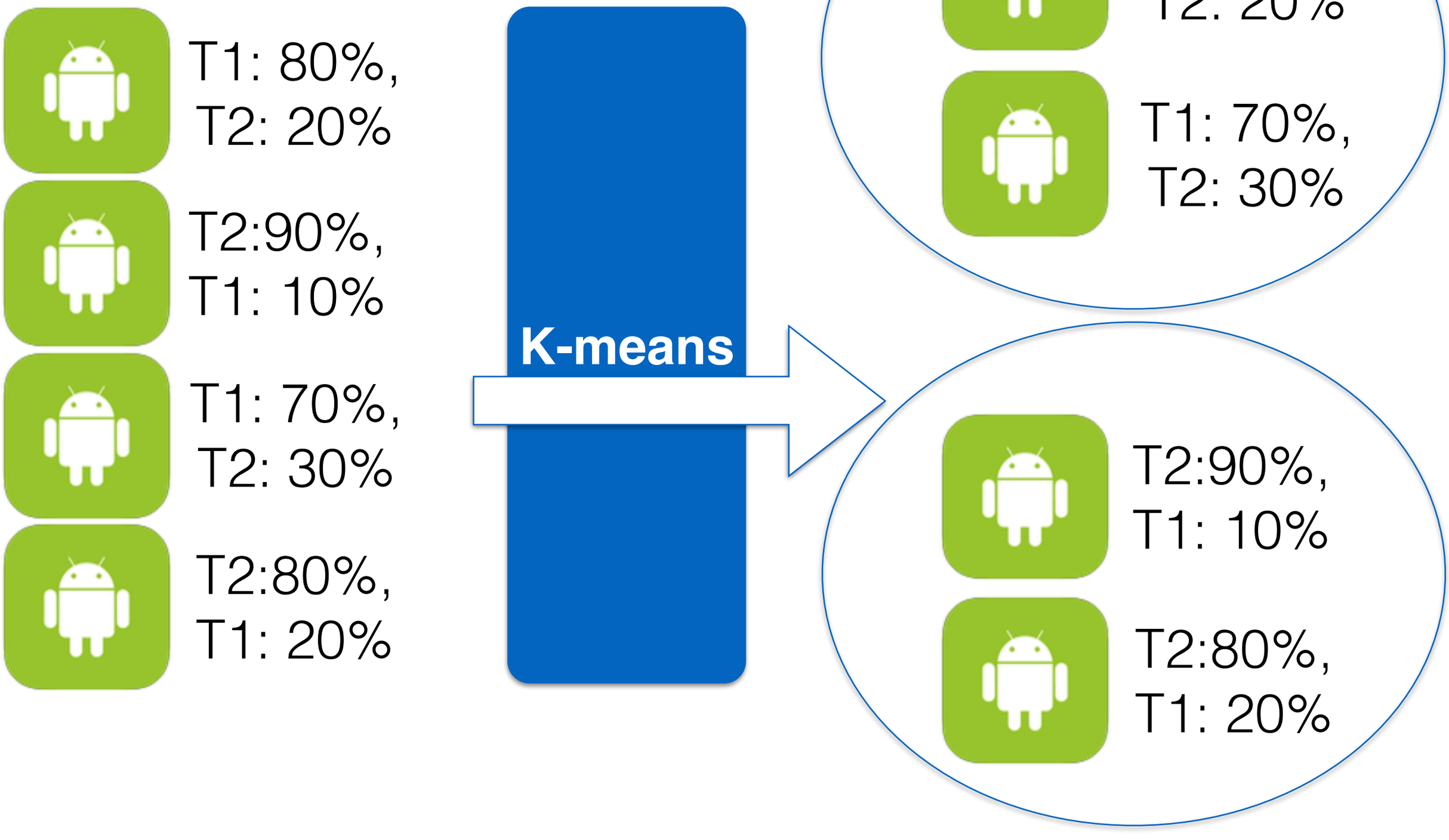
T1: 70%,
T2: 30%

T2:80%,
T1: 20%

Clustering



Clustering


T1: 80%,
T2: 20%
T2:90%,
T1: 10%
T1: 70%,
T2: 30%
T2:80%,
T1: 20%
K-means
T1: 80%,
T2: 20%
T1: 70%,
T2: 30%
T2:90%,
T1: 10%
T2:80%,
T1: 20%

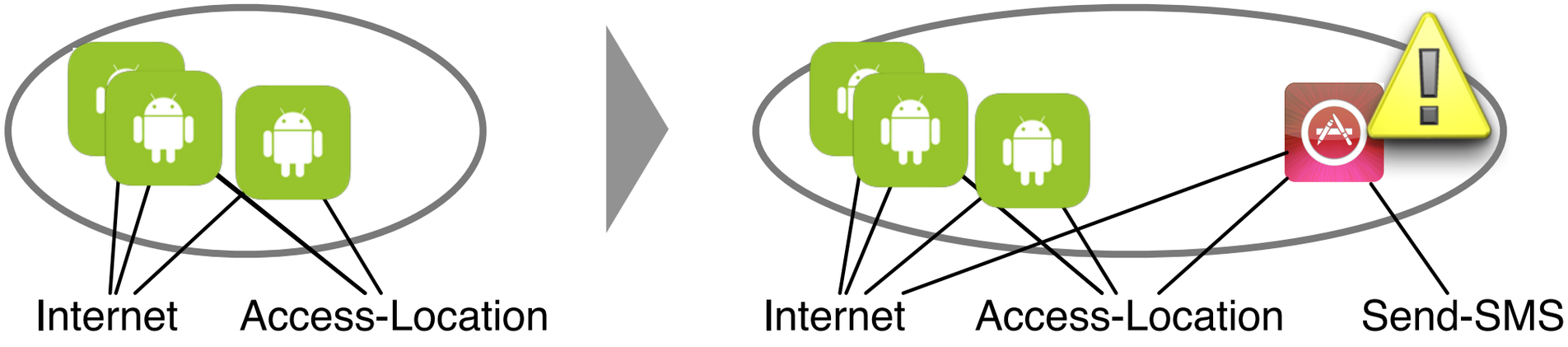
Clusters

Id	Assigned Name	Size	Most Important Topics
1	“sharing”	1,453	**share** (53%), settings and utils, navigation and travel
2	“puzzle and card games”	953	**puzzle and card games** (78%), share, game
3	“memory puzzles”	1,069	**puzzle and card games** (40%), game (12%), share
4	“music”	714	**music** (58%), share, settings and utils
5	“music videos”	773	**popular media** (44%), **holidays and religion** (20%), share
6	“religious wallpapers”	367	**holidays and religion** (56%), design and art, wallpapers
7	“language”	602	**language** (67%), share, settings and utils
8	“cheat sheets”	785	**game and cheat sheets** (76%), share, popular media
9	“utils”	1,300	**settings and utils** (62%), share, connection
10	“sports game”	1,306	**game** (63%), battle games, puzzle and card games
11	“battle games”	953	**battle games** (60%), **game**

19	“sports”	580	**sports** (62%), share, popular media
20	“files and videos”	679	**files and videos** (63%), share, settings and utils
21	“search and browse”	363	**search and browse** (64%), game, puzzle and card games
22	“advertisements”	380	**policies and ads** (97%)
23	“design and art”	978	**design and art** (48%), share, game
24	“car games”	449	**cars** (51%), game, puzzle and card games
25	“tv live”	500	**tv** (57%), share, navigation and travel
26	“adult photo”	828	**photo and social** (59%), share, settings and utils
27	“adult wallpapers”	543	**wallpapers** (51%), share, kids and bodies
28	“ad wallpapers”	180	**policies and ads** (46%), wallpapers, settings and utils
29	“ringtones and sound”	662	**ringtones and sound** (68%), share, settings and utils
30	“theme wallpapers”	593	**wallpapers** (90%), holidays and religion, share
31	“personalize”	402	**personalize** (86%), share, settings and utils
32	“settings and wallpapers”	251	**settings and utils** (37%), **wallpapers** (37%), personalize

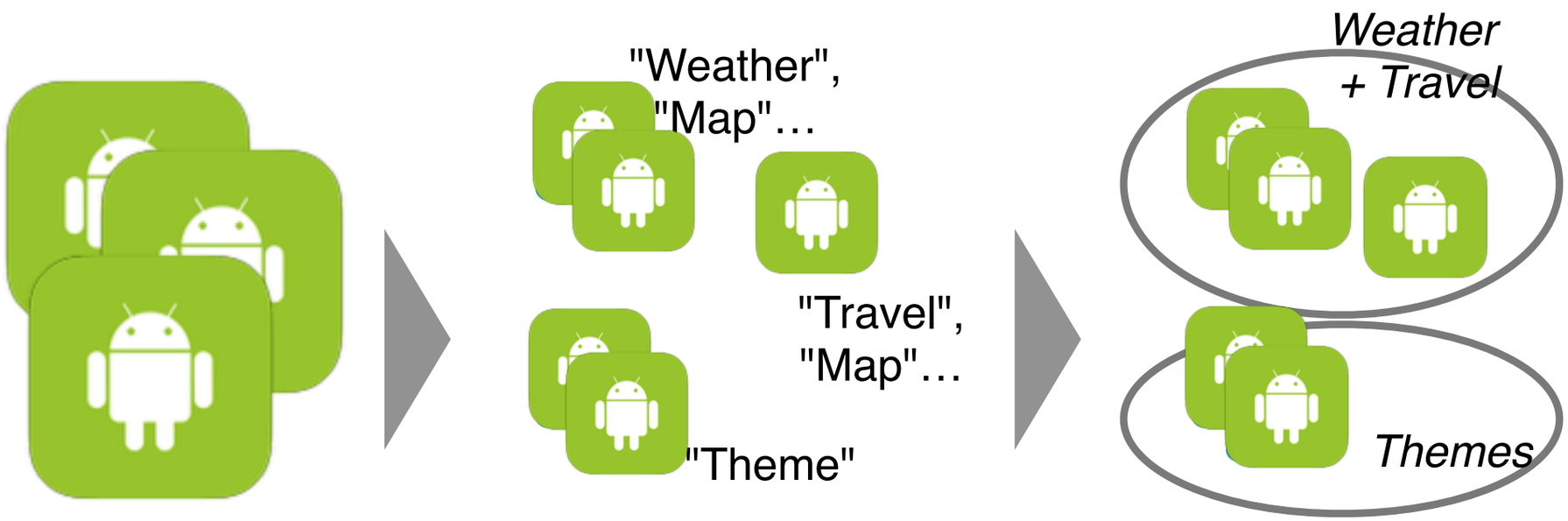
CHABADA

1. App collection **2. Topics** **3. Clusters**

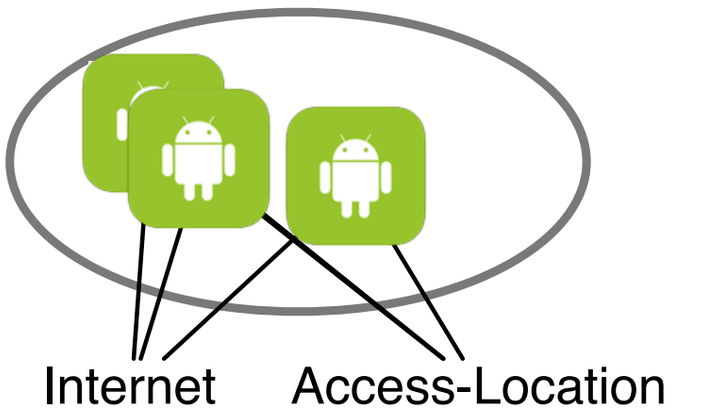


4. APIs **5. Outliers**

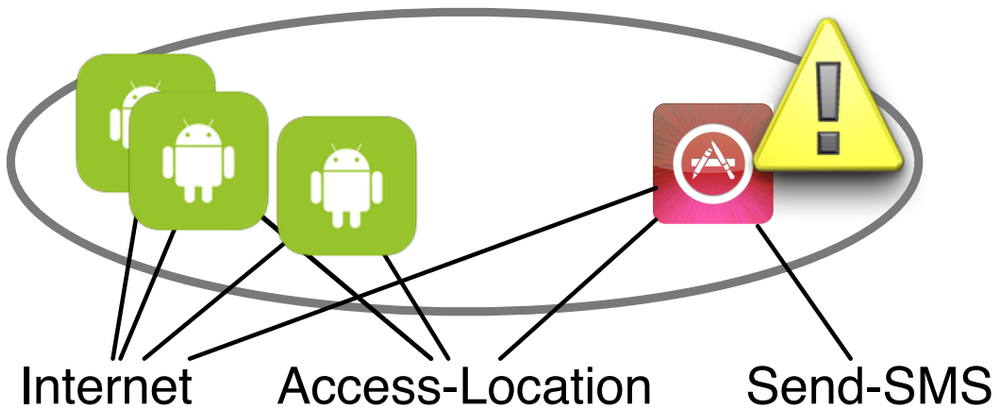
CHABADA



1. App collection **2. Topics** **3. Clusters**



4. APIs



5. Outliers

API Analysis

Android APIs API level: 19

android
android.accessibilityservice
android.accounts
android.animation
android.app
android.app.admin
android.app.backup
android.appwidget
android.bluetooth
android.content
android.content.pm
android.content.res
android.database

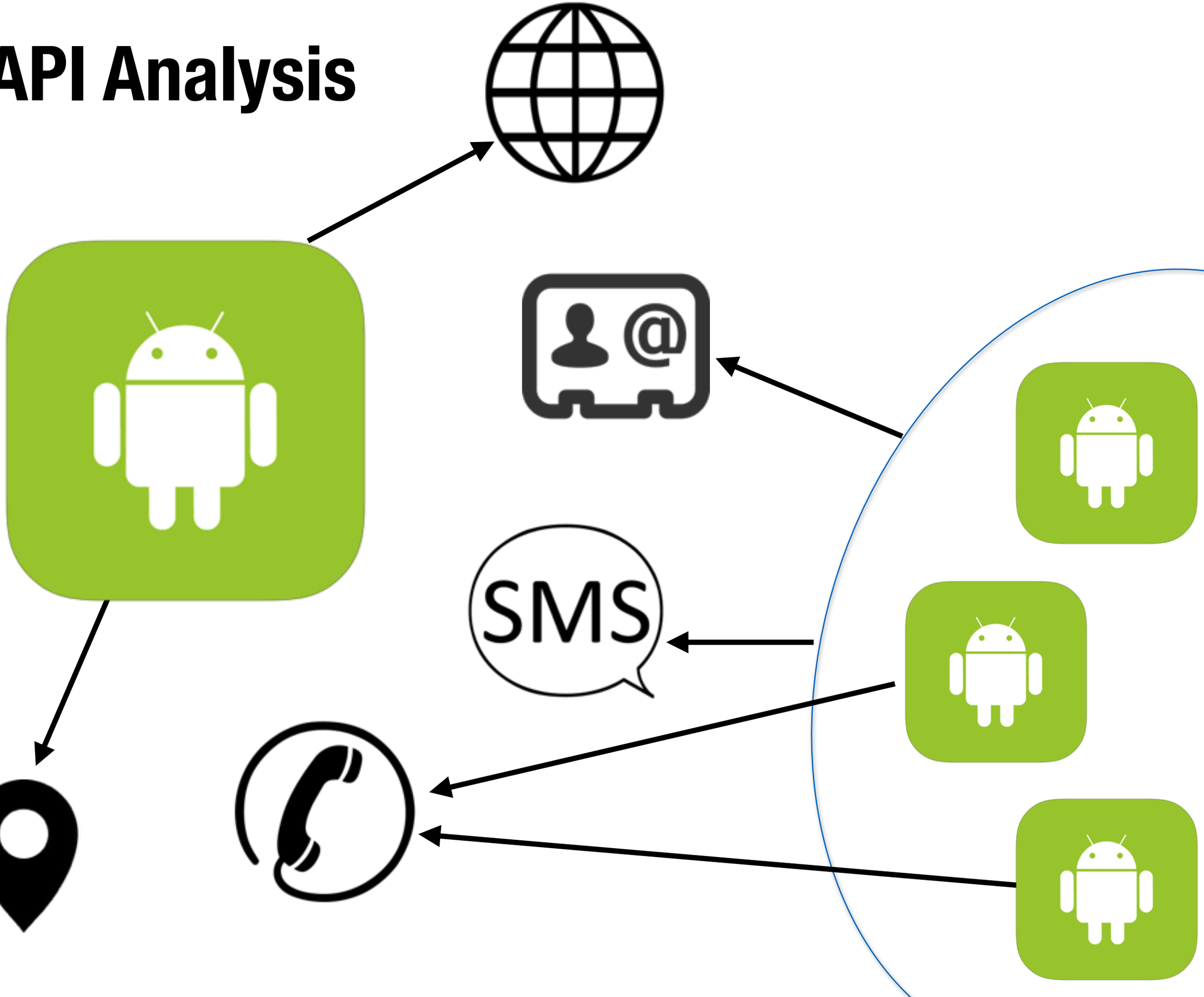
API Analysis

API Analysis

API Analysis

API Analysis

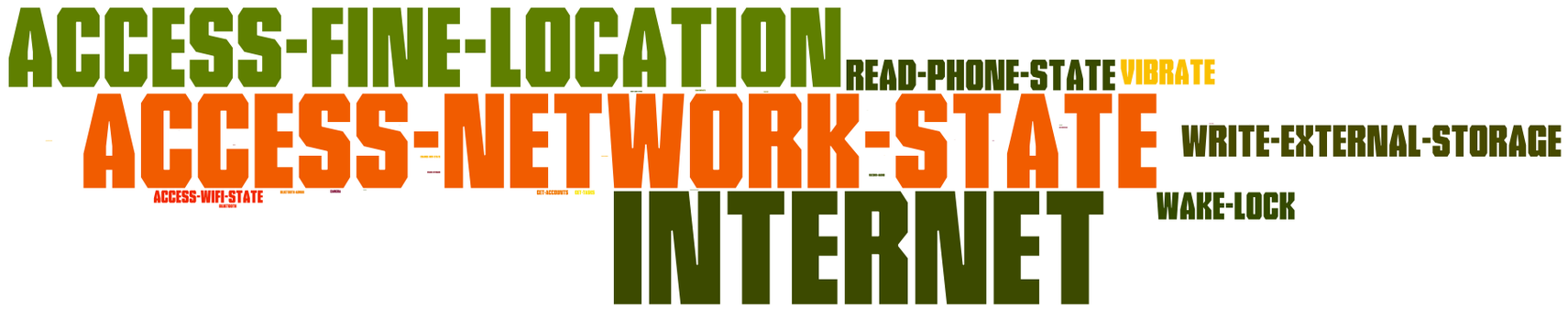
API Analysis



“Travel” cluster

Description

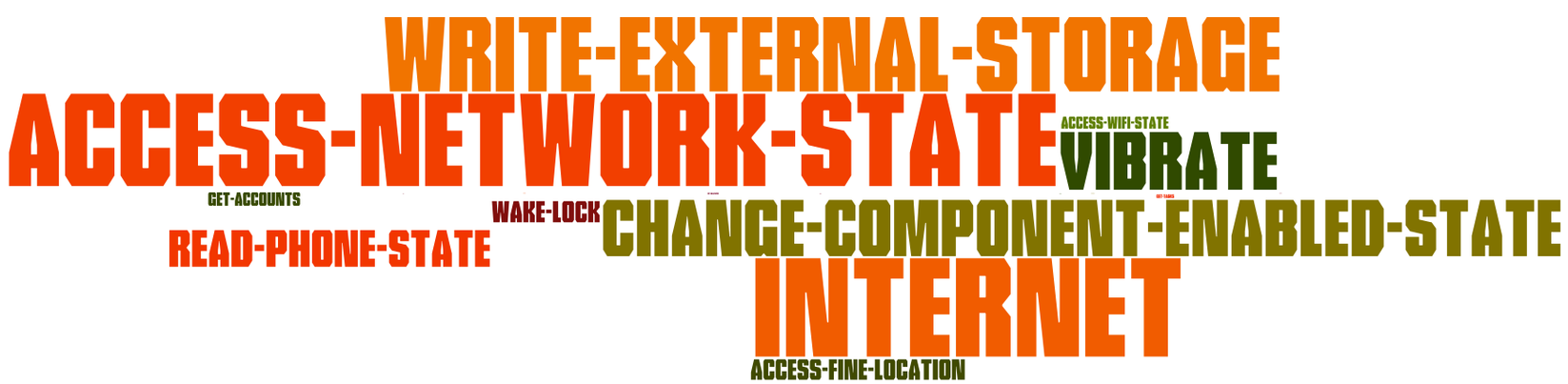
Permissions of APIs used



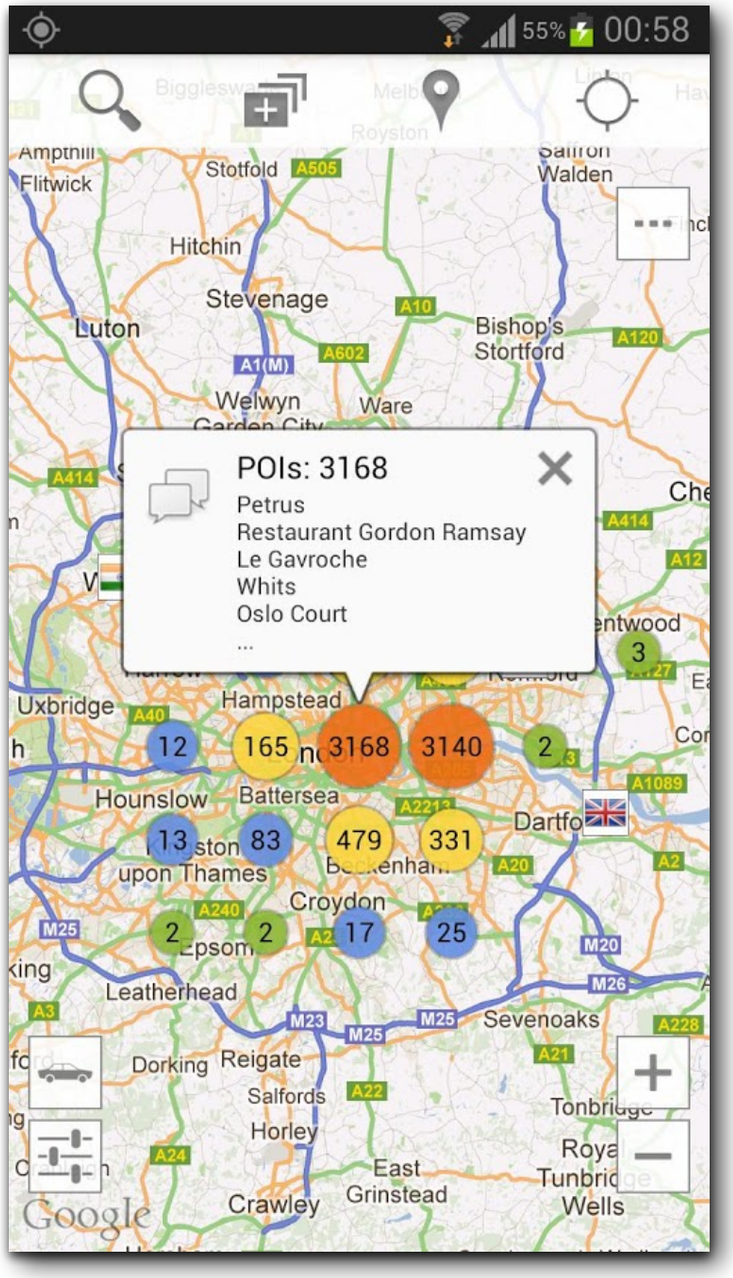
“Personalize” cluster

Description

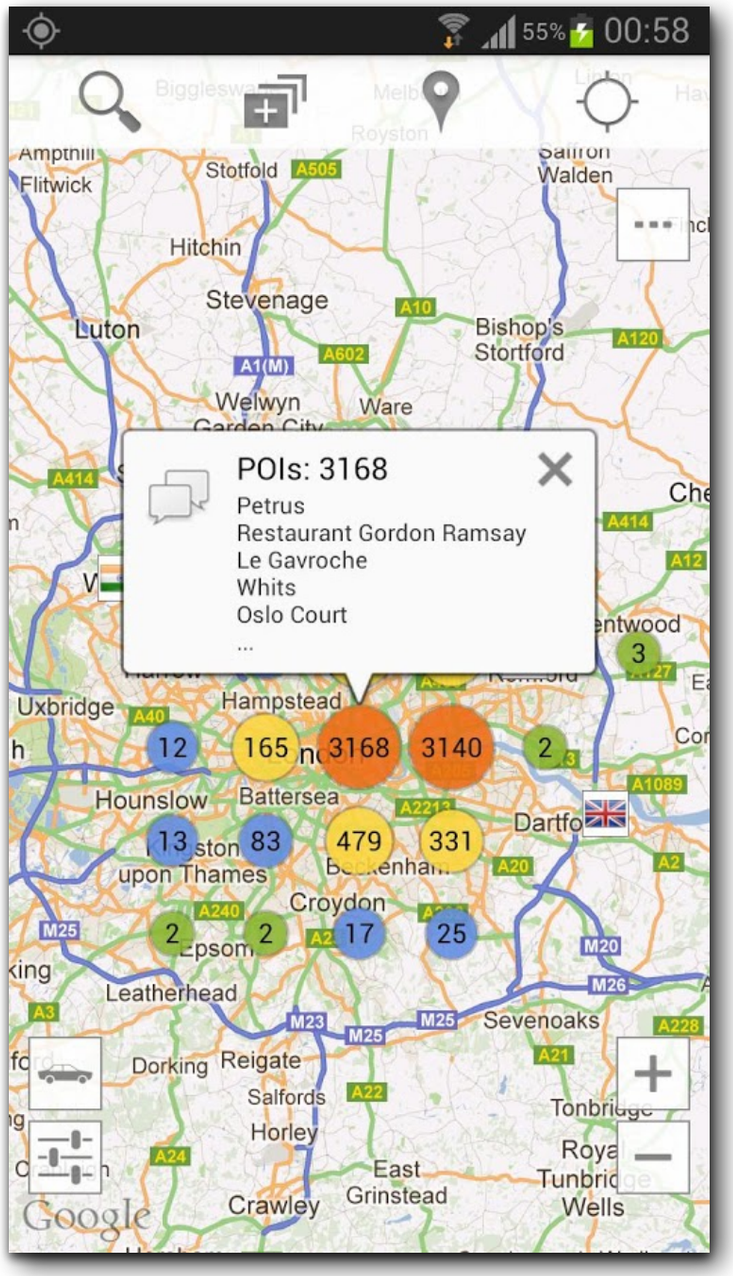
Permissions of APIs used



London Restaurants

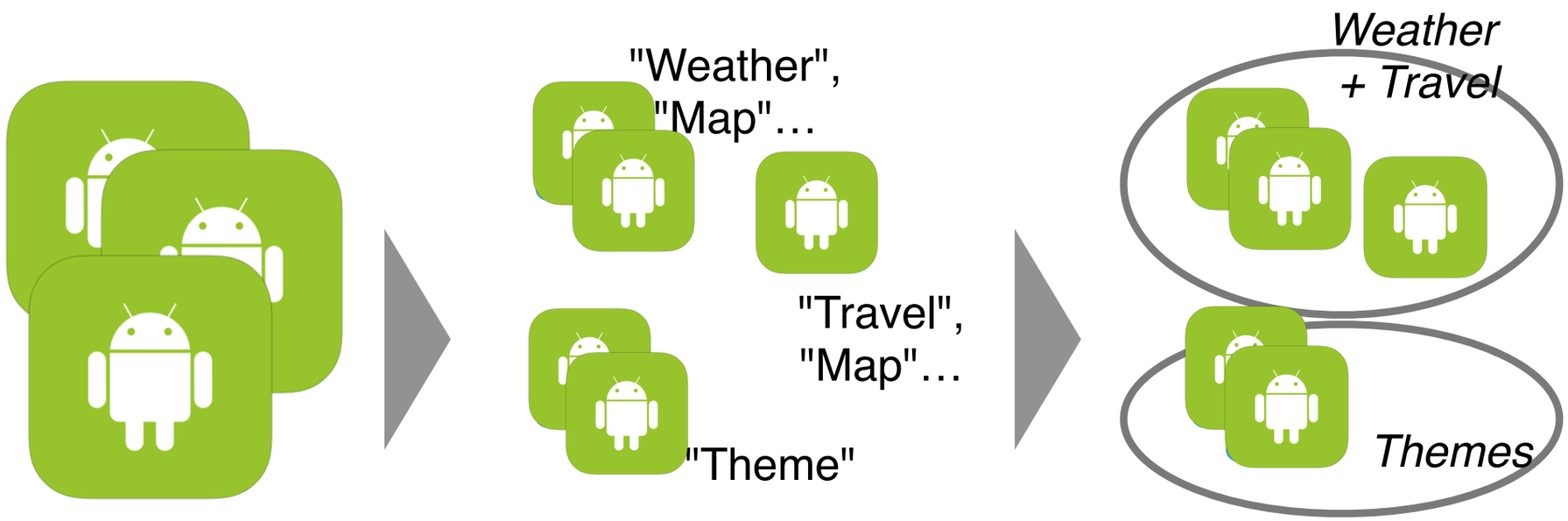


London Restaurants

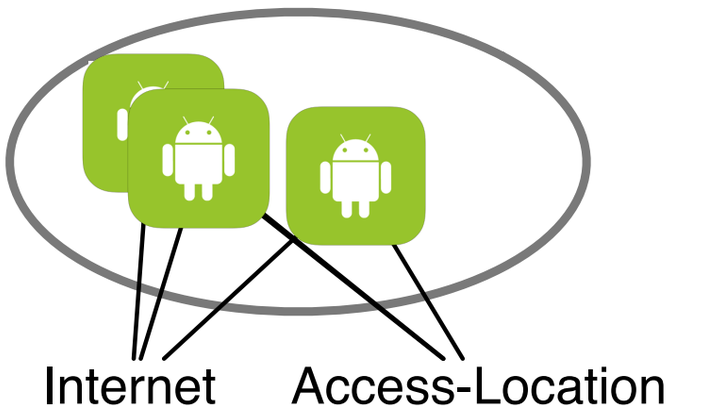


android.net.ConnectivityManager.getActiveNetworkInfo()
android.webkit.WebView()
java.net.HttpURLConnection.connect()
android.app.NotificationManager.notify()
java.net.URL.openConnection()
android.telephony.TelephonyManager.getDeviceId()
org.apache.http.impl.client.DefaultHttpClient()
org.apache.http.impl.client.DefaultHttpClient.execute()
android.location.LocationManager.getBestProvider()
android.telephony.TelephonyManager.getLine1Number()
android.net.wifi.WifiManager.isWifiEnabled()
android.accounts.AccountManager.getAccountsByType()
android.net.wifi.WifiManager.getConnectionInfo()
android.location.LocationManager.getLastKnownLocation()
android.location.LocationManager.isProviderEnabled()
android.location.LocationManager.requestLocationUpdates()
android.net.NetworkInfo.isConnectedOrConnecting()
android.net.ConnectivityManager.getAllNetworkInfo()

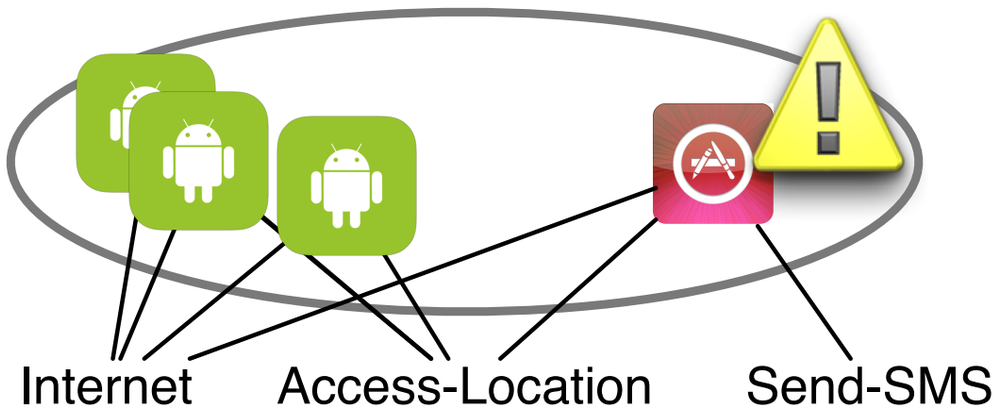
CHABADA



1. App collection **2. Topics** **3. Clusters**

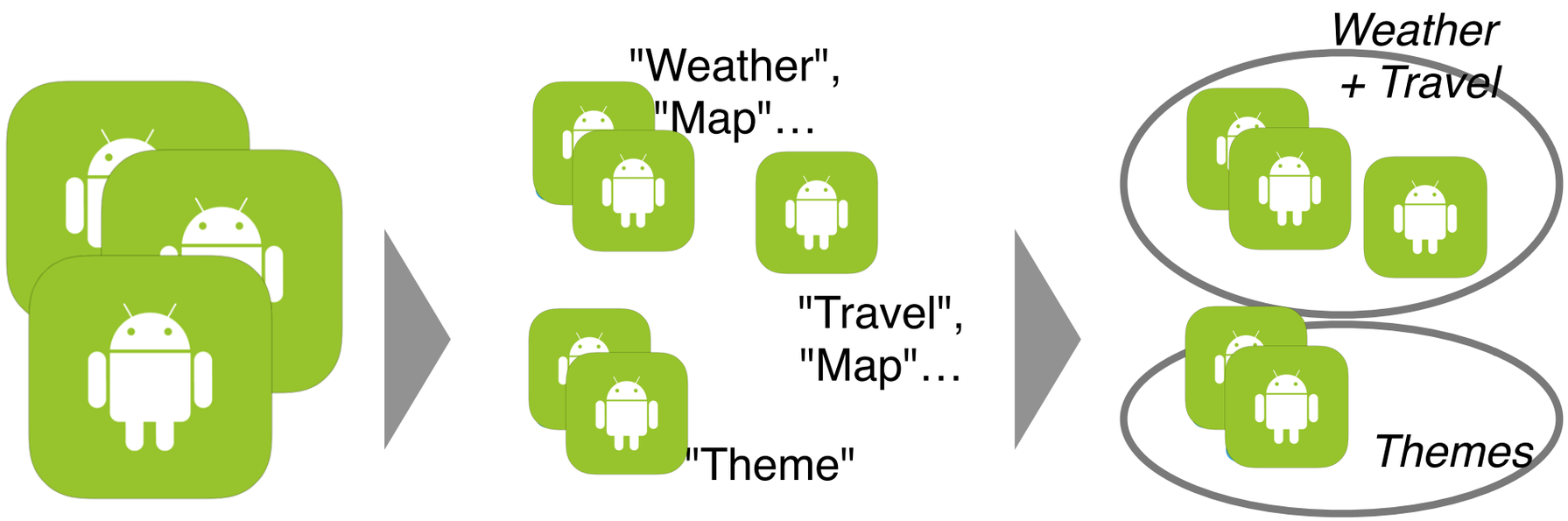


4. APIs

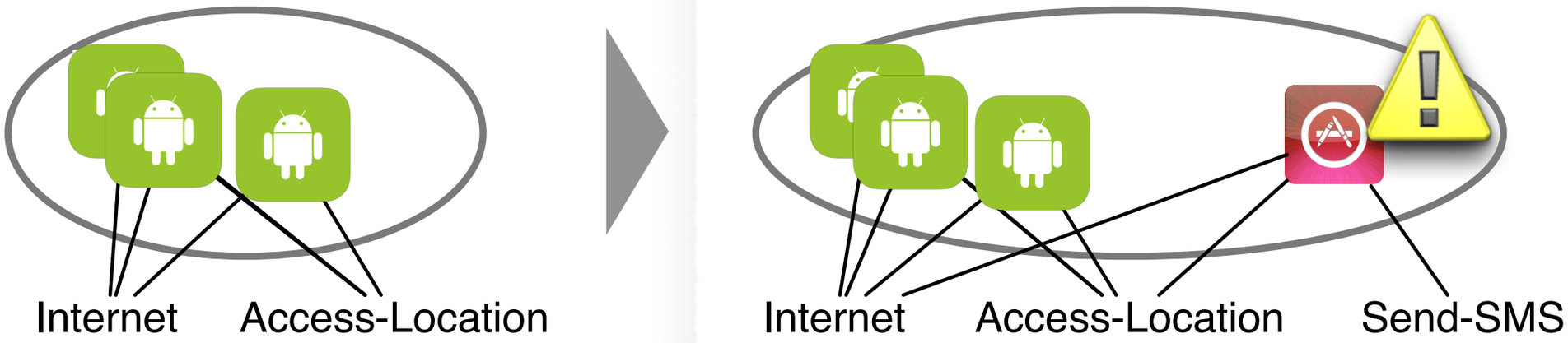


5. Outliers

CHABADA

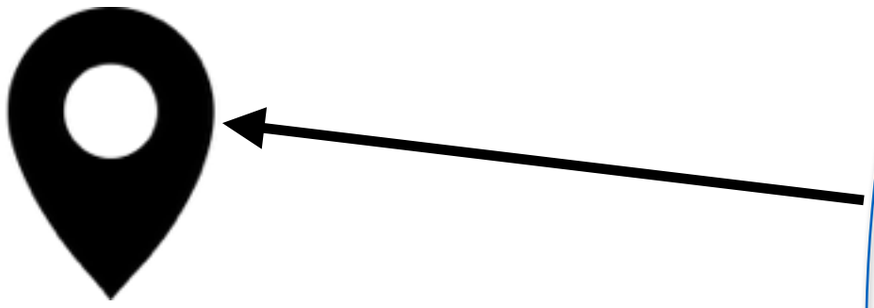


1. App collection **2. Topics** **3. Clusters**



4. APIs **5. Outliers**

Anomaly detection



Anomaly detection

SMS

Anomaly detection

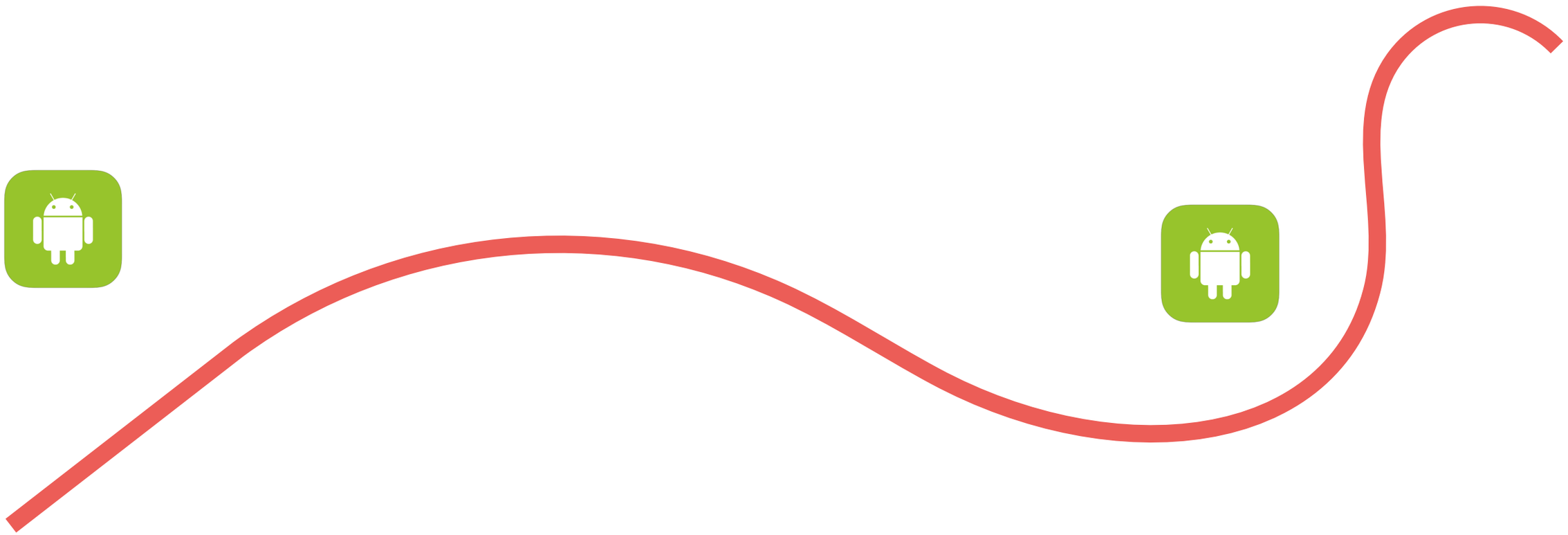
- In each cluster, identified anomalies through *one-class support vector machine* (OC-SVM)
- Features of each app: a vector of *(sensitive APIs, binary value)*

Anomaly detection

- In each cluster, identified anomalies through *one-class support vector machine* (OC-SVM)
- Features of each app: a vector of *(sensitive APIs, binary value)*

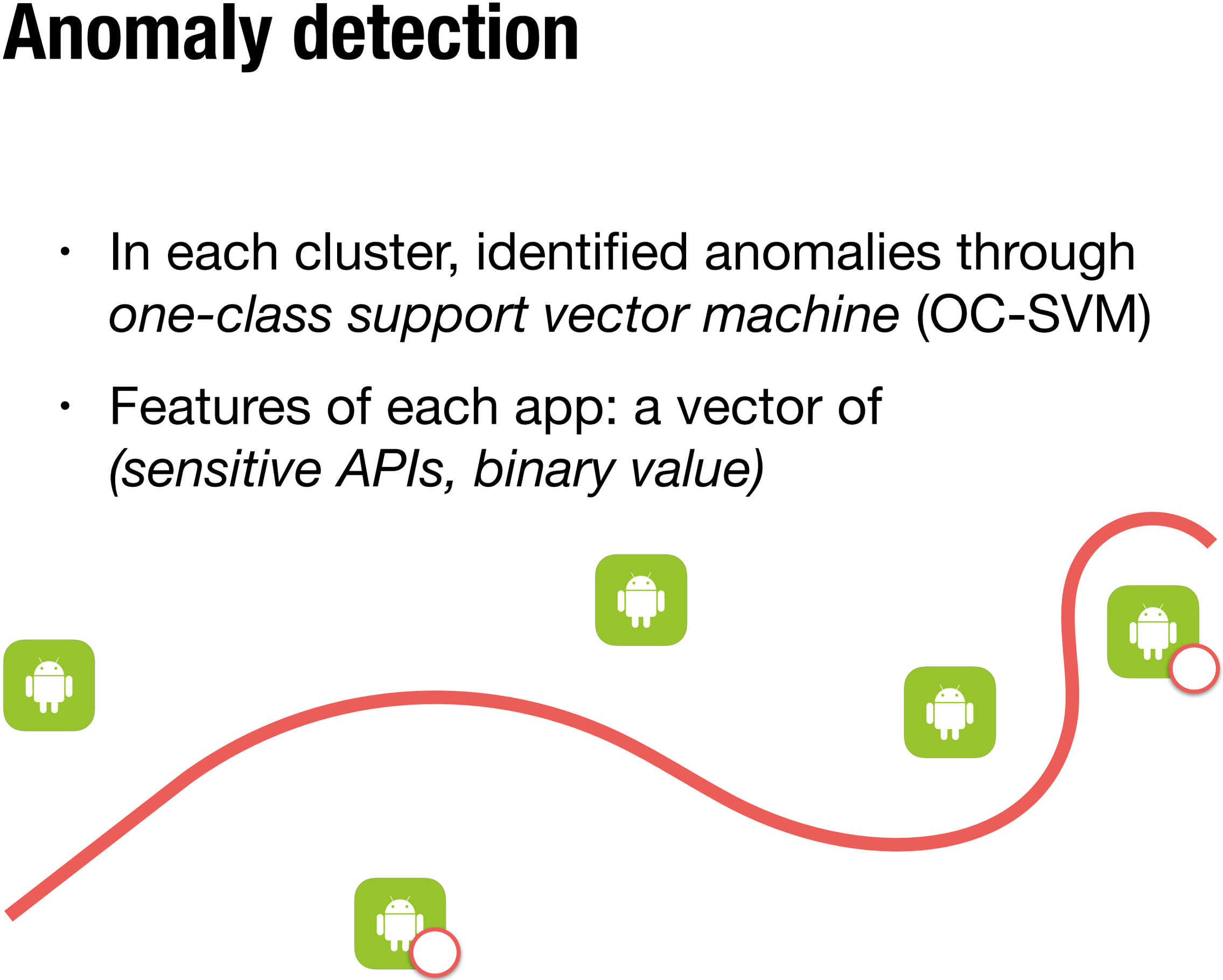
Anomaly detection

- In each cluster, identified anomalies through *one-class support vector machine* (OC-SVM)
- Features of each app: a vector of *(sensitive APIs, binary value)*



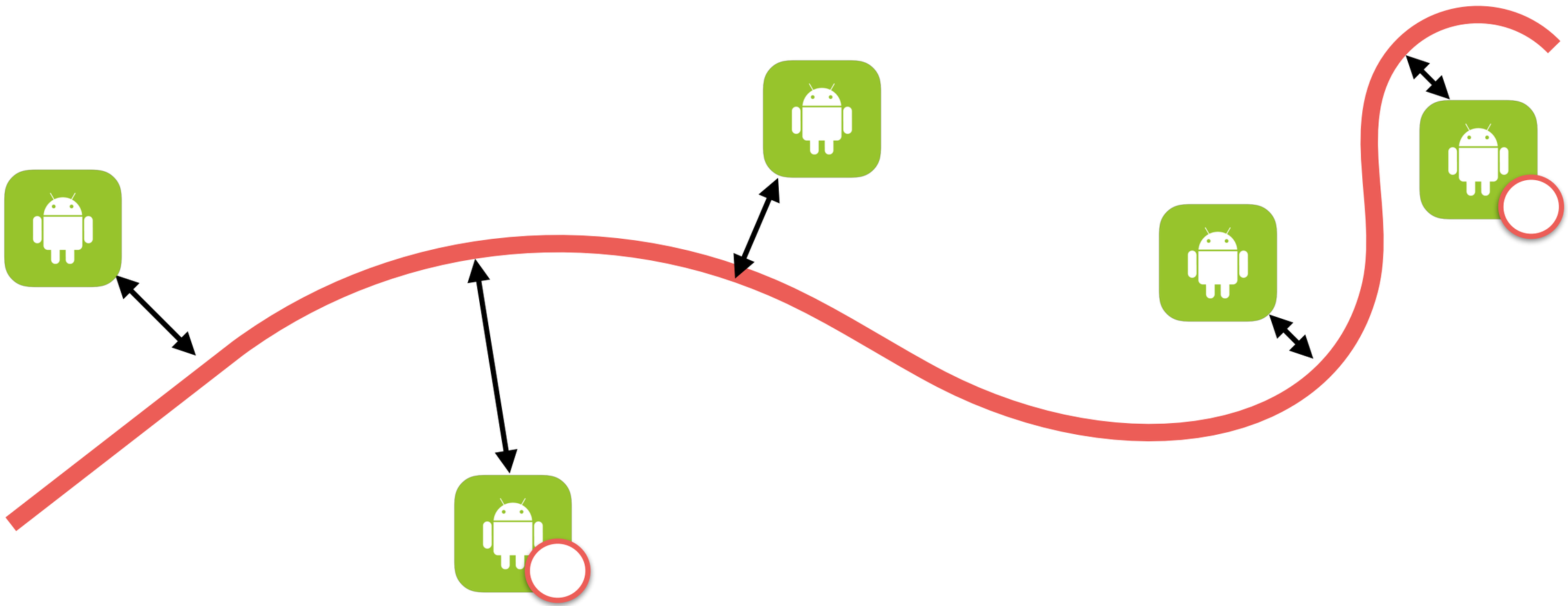
Anomaly detection

- In each cluster, identified anomalies through *one-class support vector machine* (OC-SVM)
- Features of each app: a vector of *(sensitive APIs, binary value)*



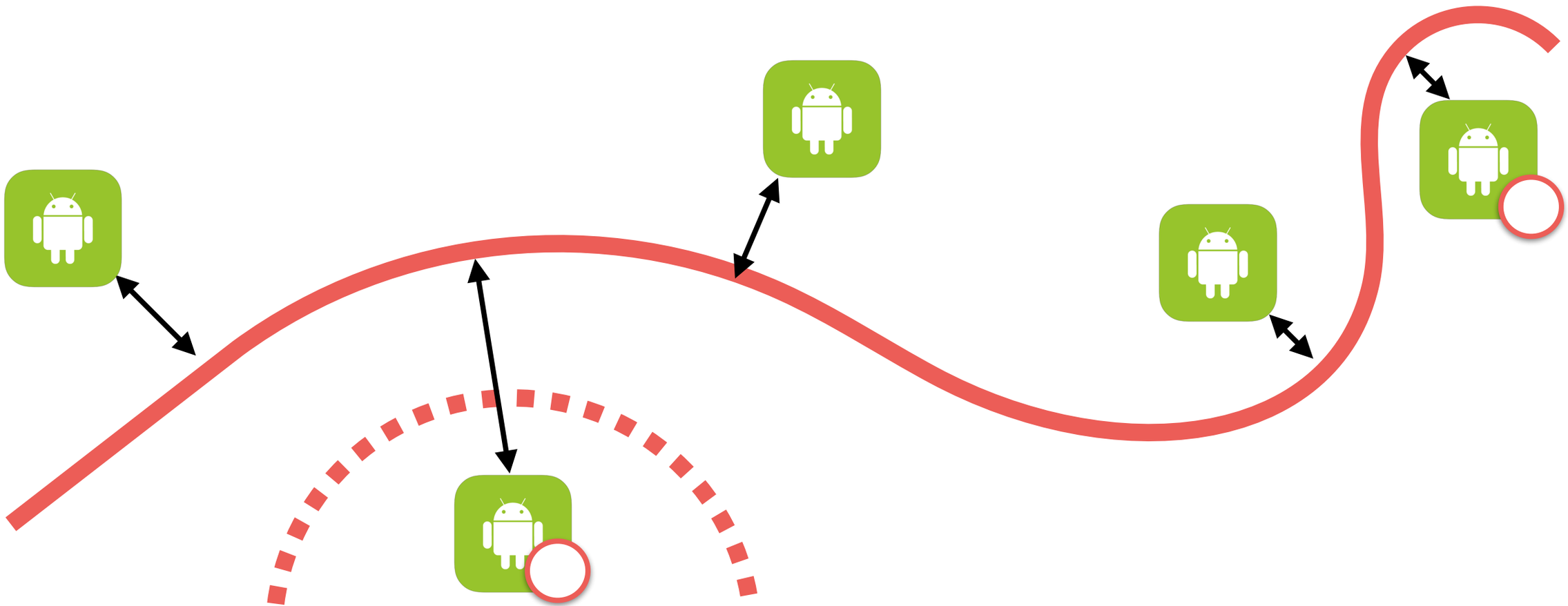
Anomaly detection

- In each cluster, identified anomalies through *one-class support vector machine* (OC-SVM)
- Features of each app: a vector of *(sensitive APIs, binary value)*

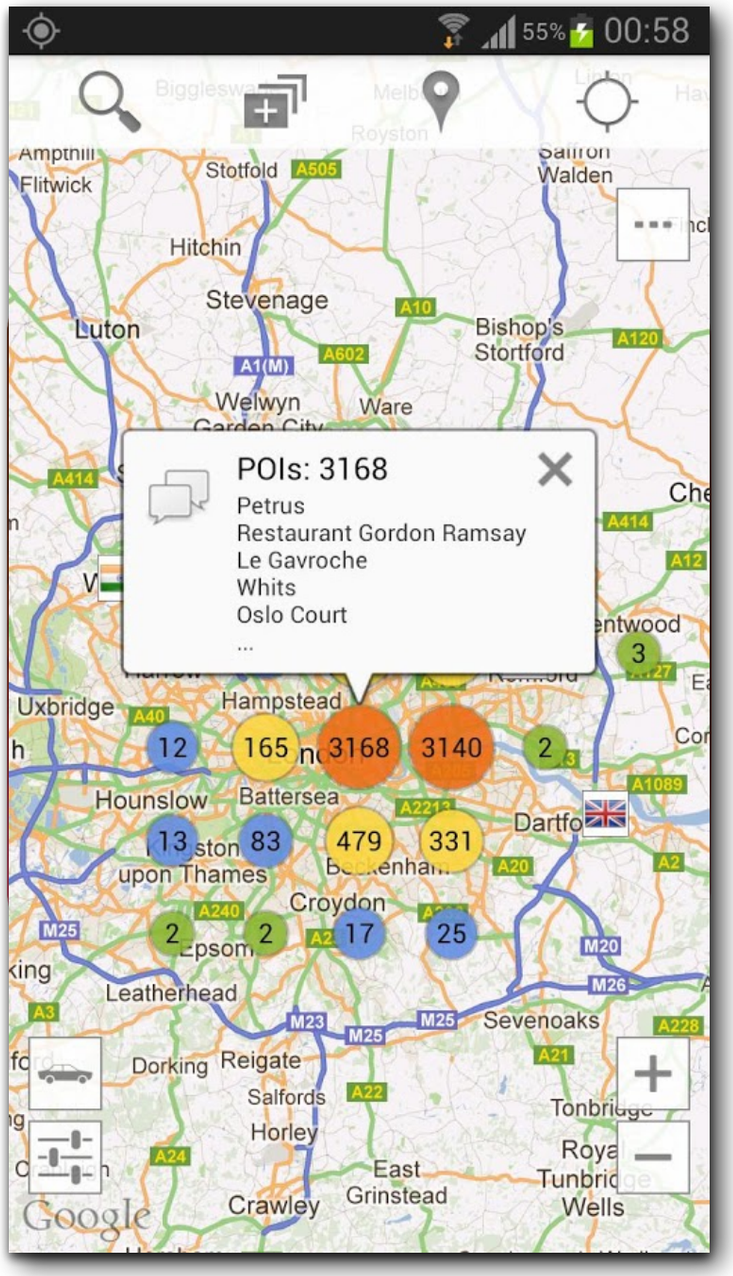


Anomaly detection

- In each cluster, identified anomalies through *one-class support vector machine* (OC-SVM)
- Features of each app: a vector of *(sensitive APIs, binary value)*

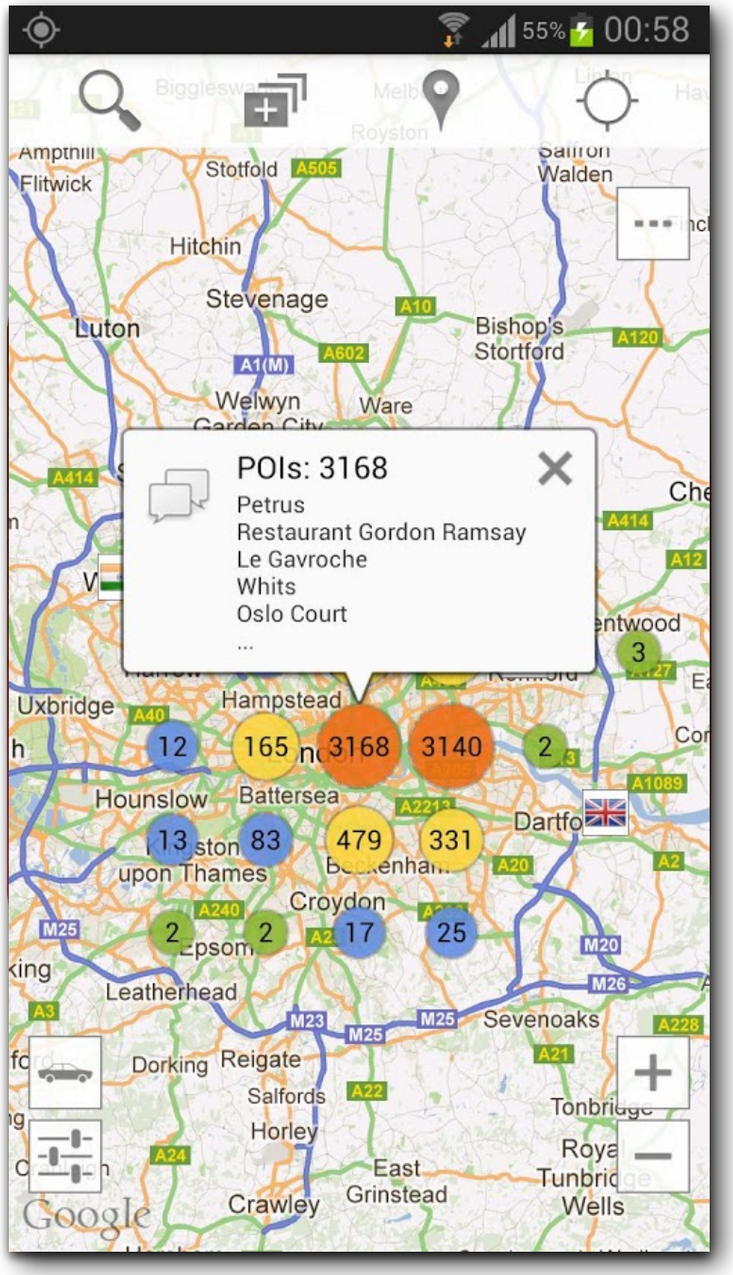


London Restaurants



android.net.ConnectivityManager.getActiveNetworkInfo()
android.webkit.WebView()
java.net.HttpURLConnection.connect()
android.app.NotificationManager.notify()
java.net.URL.openConnection()
android.telephony.TelephonyManager.getDeviceId()
org.apache.http.impl.client.DefaultHttpClient()
org.apache.http.impl.client.DefaultHttpClient.execute()
android.location.LocationManager.getBestProvider()
android.telephony.TelephonyManager.getLine1Number()
android.net.wifi.WifiManager.isWifiEnabled()
android.accounts.AccountManager.getAccountsByType()
android.net.wifi.WifiManager.getConnectionInfo()
android.location.LocationManager.getLastKnownLocation()
android.location.LocationManager.isProviderEnabled()
android.location.LocationManager.requestLocationUpdates()
android.net.NetworkInfo.isConnectedOrConnecting()
android.net.ConnectivityManager.getAllNetworkInfo()

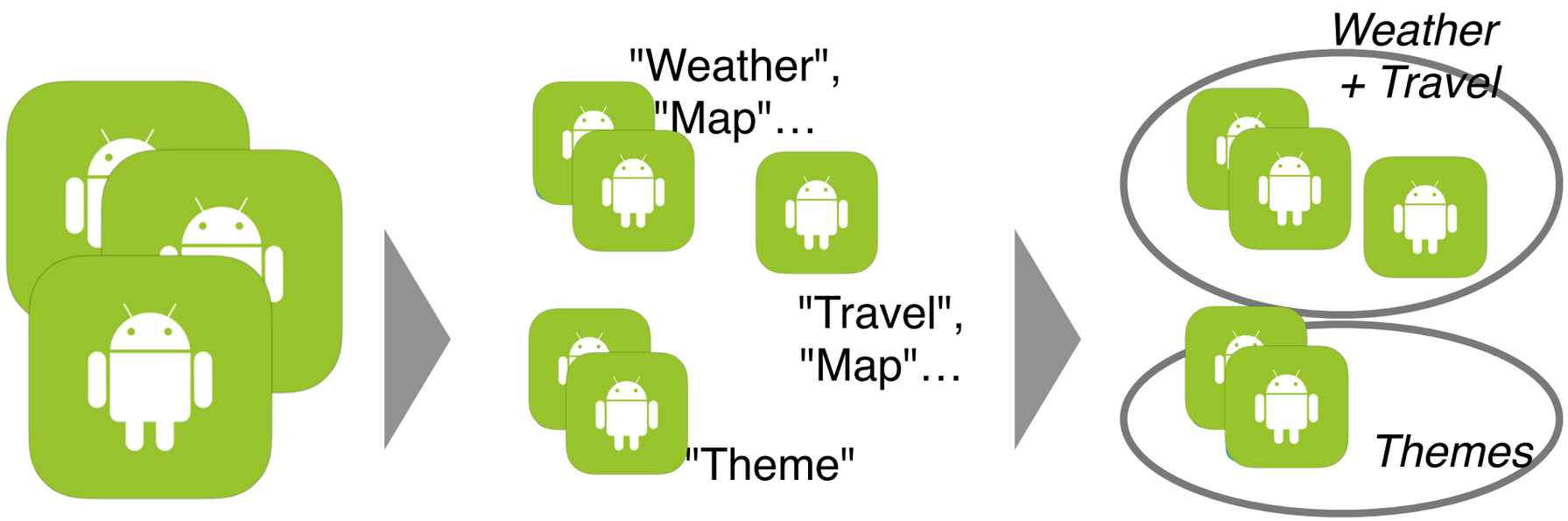
London Restaurants



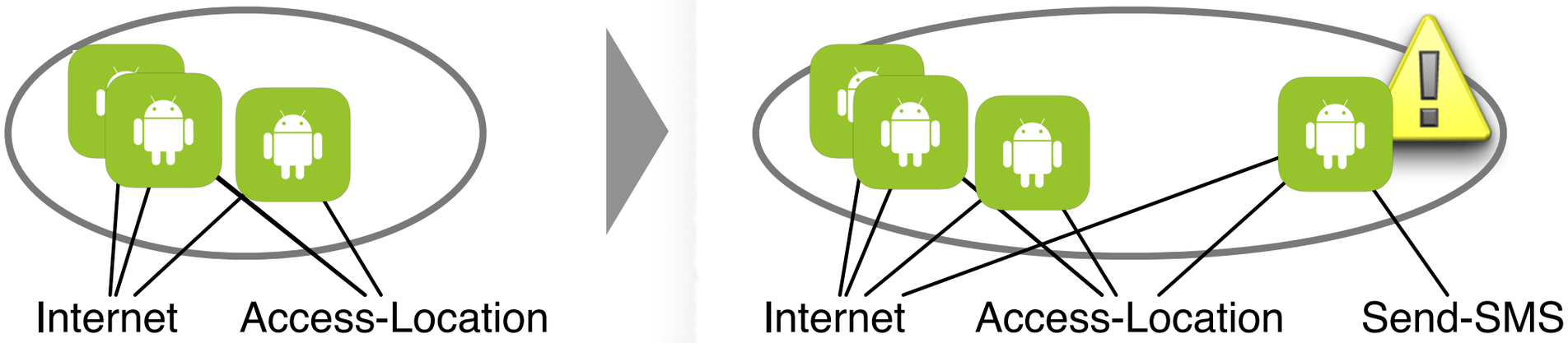
android.net.ConnectivityManager.getActiveNetworkInfo()
android.webkit.WebView()
java.net.HttpURLConnection.connect()
android.app.NotificationManager.notify()
java.net.URL.openConnection()
android.telephony.TelephonyManager.getDeviceId()
org.apache.http.impl.client.DefaultHttpClient()
org.apache.http.impl.client.DefaultHttpClient.execute()
android.location.LocationManager.getBestProvider()
android.telephony.TelephonyManager.getLine1Number()
android.net.wifi.WifiManager.isWifiEnabled()
android.accounts.AccountManager.getAccountsByType()
android.net.wifi.WifiManager.getConnectionInfo()
android.location.LocationManager.getLastKnownLocation()
android.location.LocationManager.isProviderEnabled()
android.location.LocationManager.requestLocationUpdates()
android.net.NetworkInfo.isConnectedOrConnecting()
android.net.ConnectivityManager.getAllNetworkInfo()

→ Identified as Anomaly

CHABADA



1. App collection **2. Topics** **3. Clusters**



4. APIs **5. Outliers**

Evaluation: Anomalies

Can CHABADA effectively identify *anomalous*(*) Android apps?

Evaluation: Anomalies

Can CHABADA effectively identify *anomalous*(*) Android apps?

(*) i.e., mismatches between description and behavior

Evaluation: Anomalies

Can CHABADA effectively identify *anomalous*(*) Android apps?

160 apps

(*) i.e., mismatches between description and behavior

Evaluation: Anomalies

Can CHABADA effectively identify *anomalous*(*) Android apps?

160 apps

26%
covert behavior

(*) i.e., mismatches between description and behavior

What makes an anomaly?

What makes an anomaly?

apploving

airpush

What makes an anomaly?

apploving

airpush

dubious behaviour

What makes an anomaly?

apploving

airpush

uncommon behaviour

dubious behaviour

What makes an anomaly?

apploving

airpush

uncommon behaviour

dubious behaviour

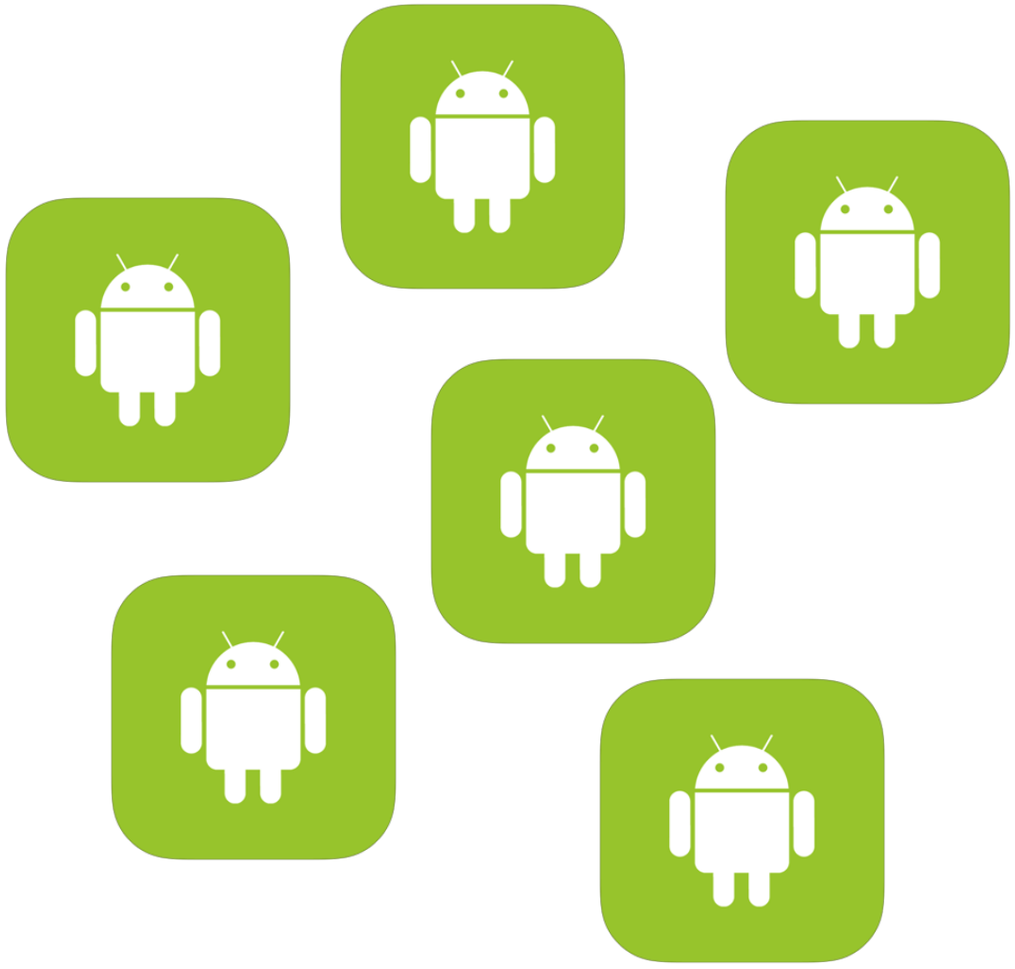
benign outliers

Evaluation: Malware

Can our technique be used to identify *malicious* Android applications?

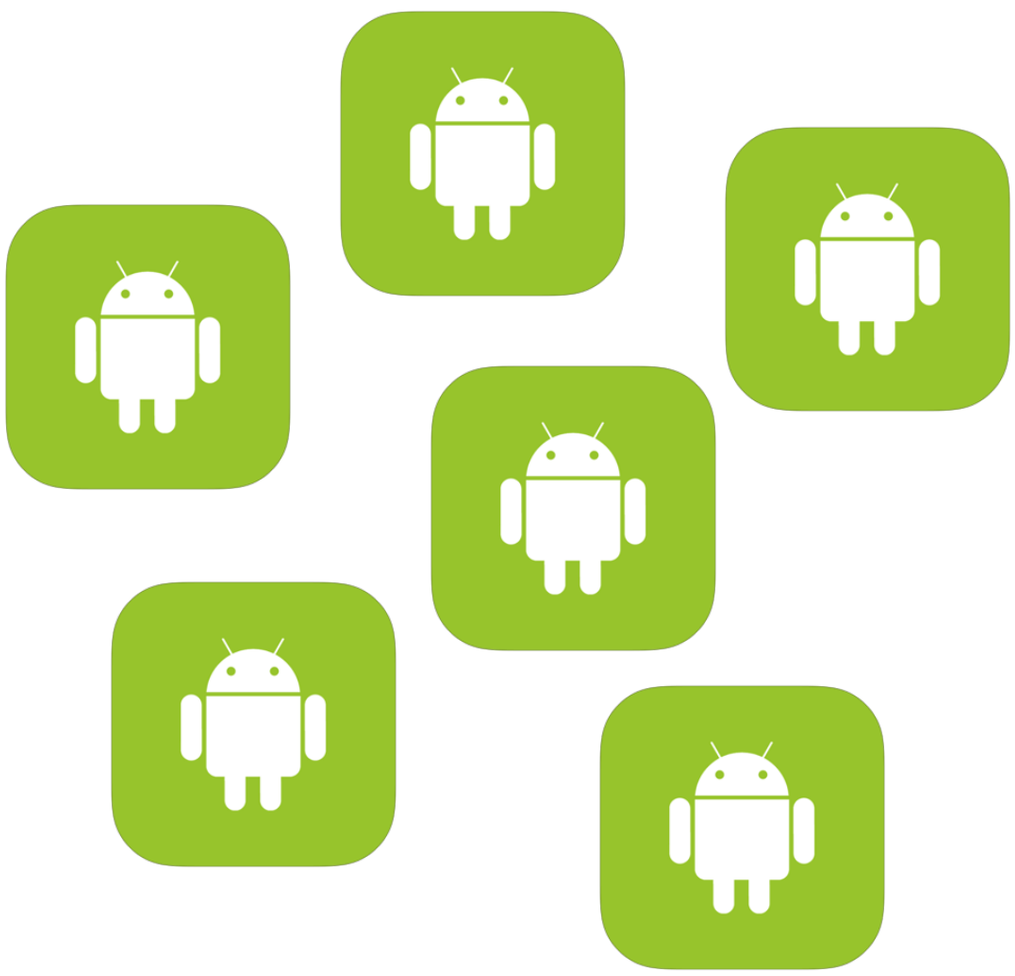
Evaluation: Malware

Can our technique be used to identify *malicious* Android applications?



Evaluation: Malware

Can our technique be used to identify *malicious* Android applications?

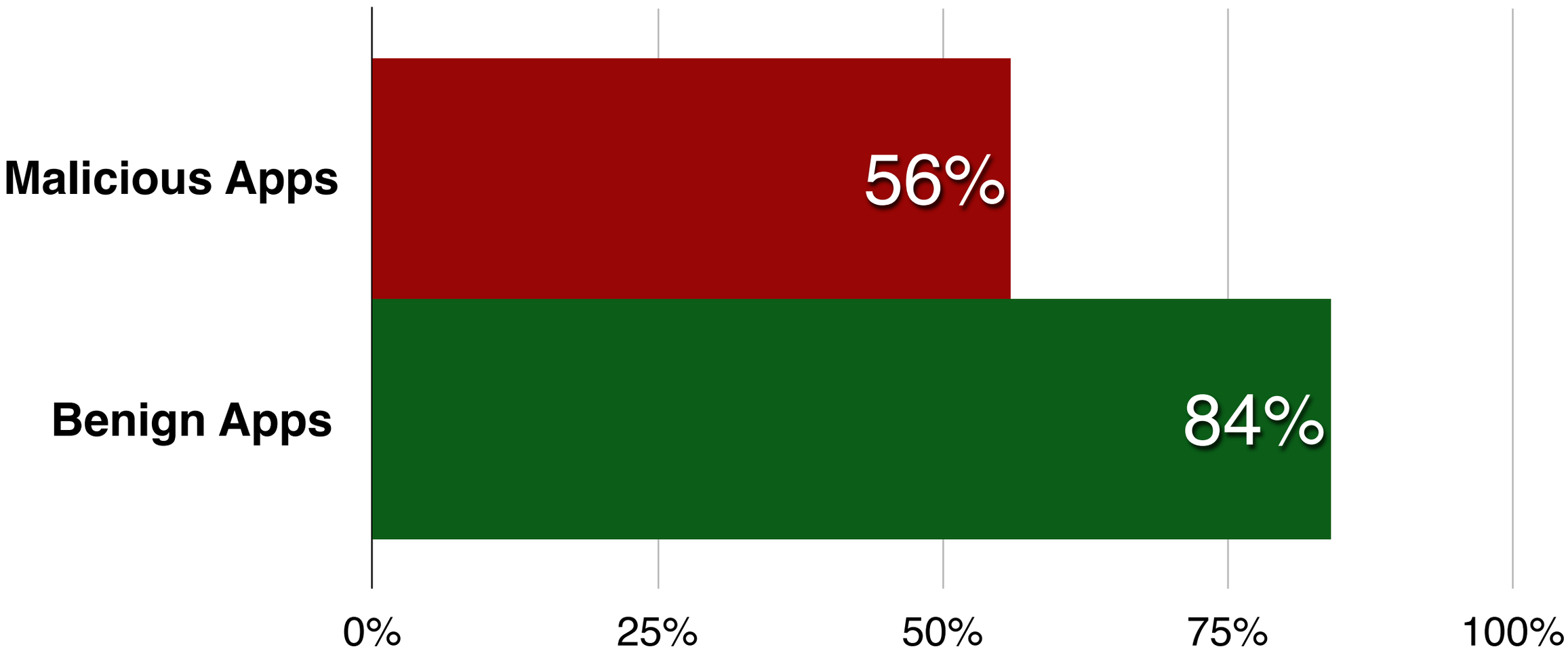


Classification with Clusters
(our approach)

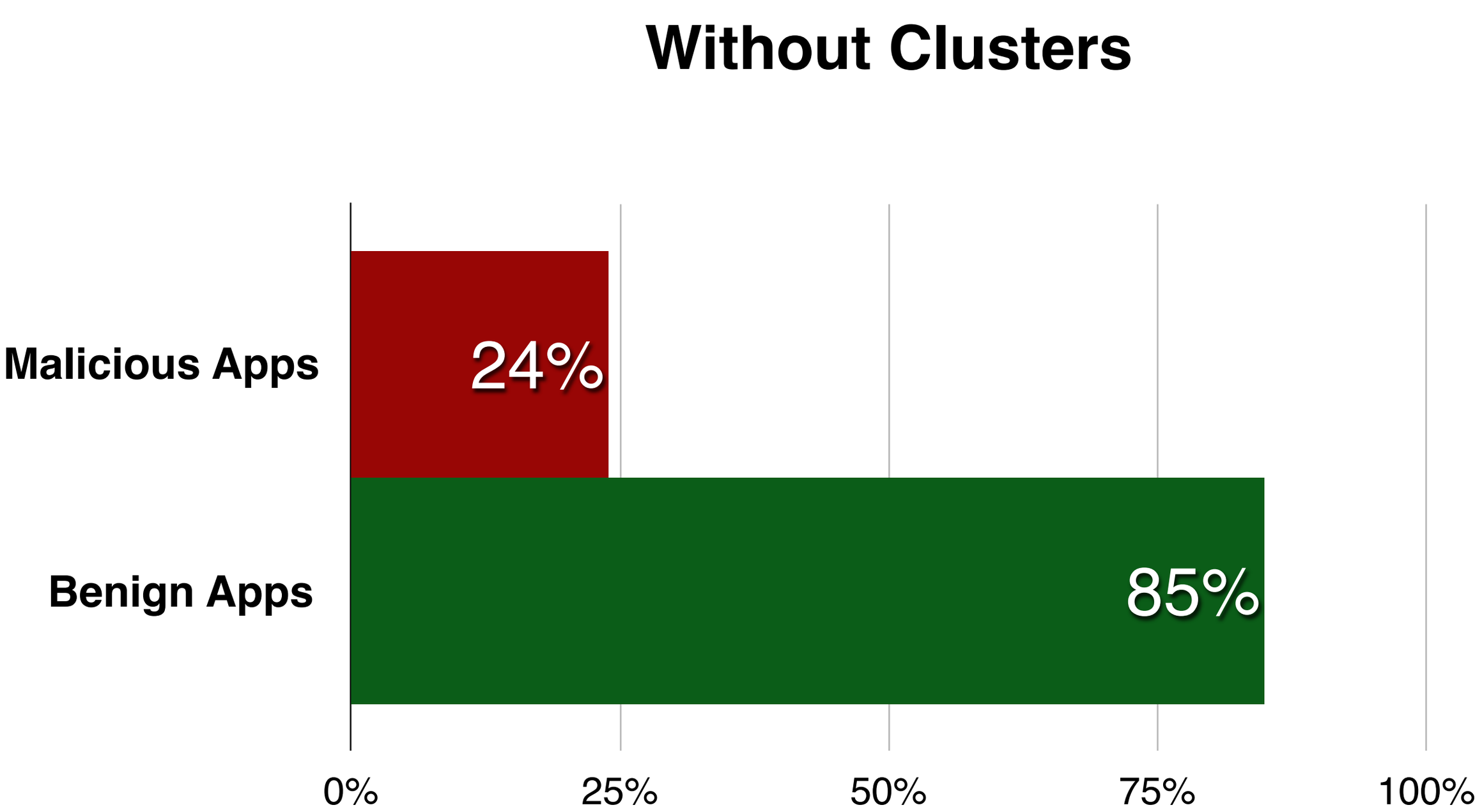
	Predicted as Malicious	Predicted as Benign
Malicious Apps	56%	44%
Benign Apps	16%	84%

Correct Classification

With Clusters (our approach)

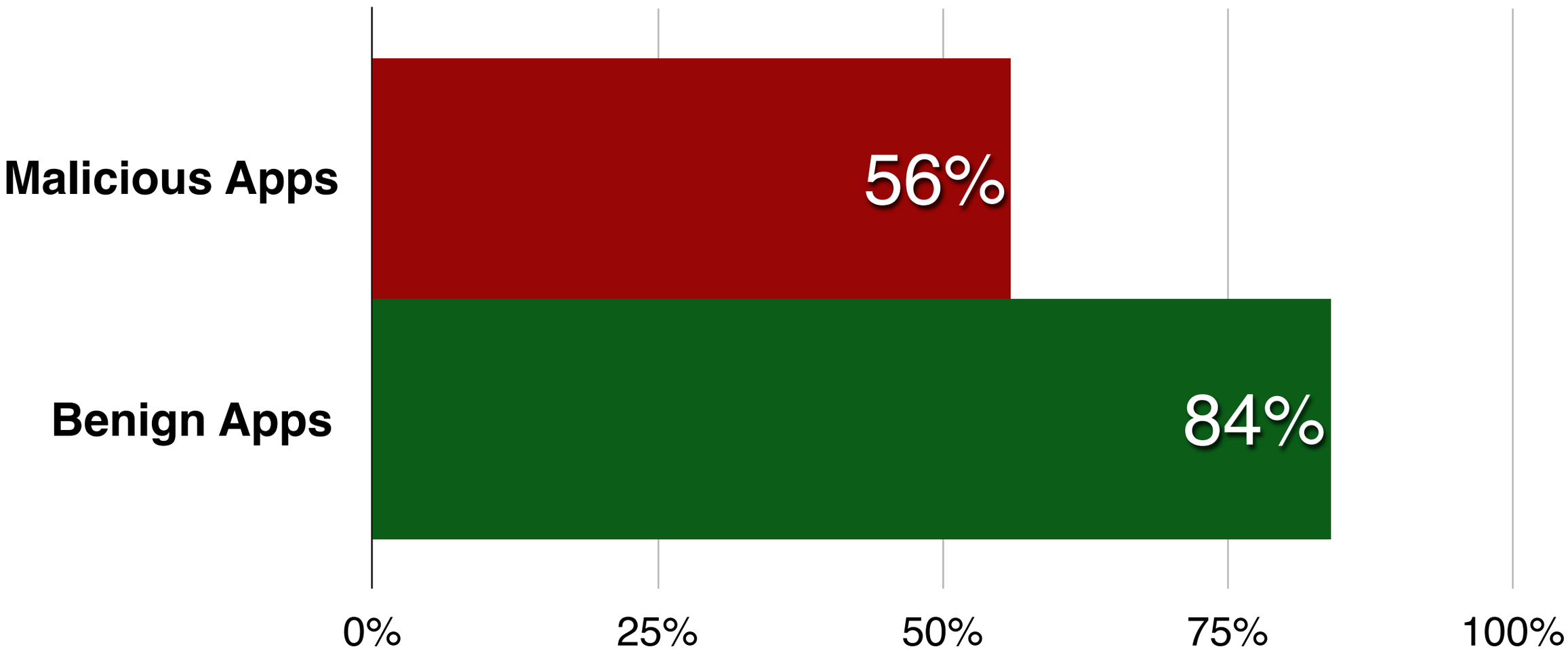


Correct Classification

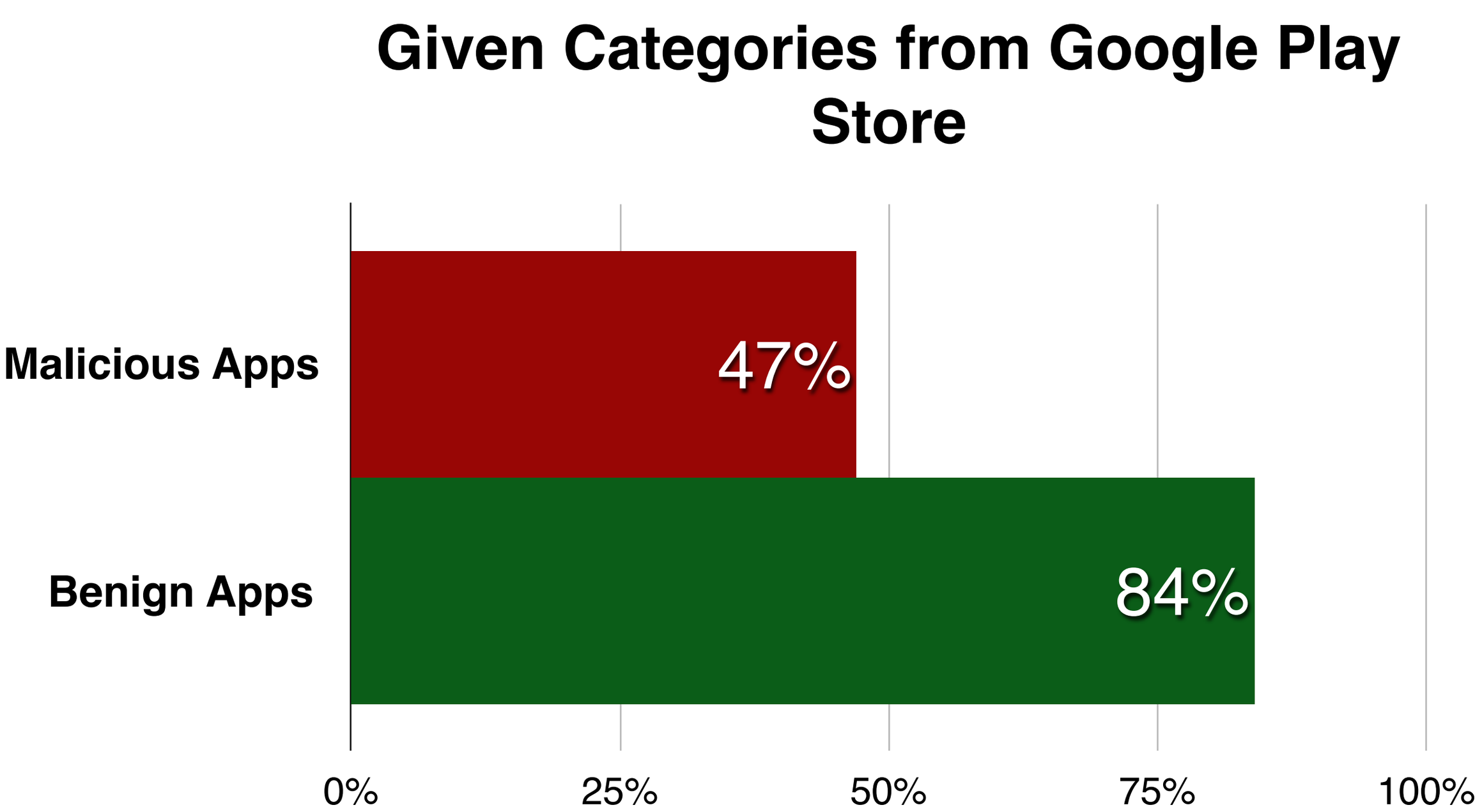


Correct Classification

With Clusters (our approach)

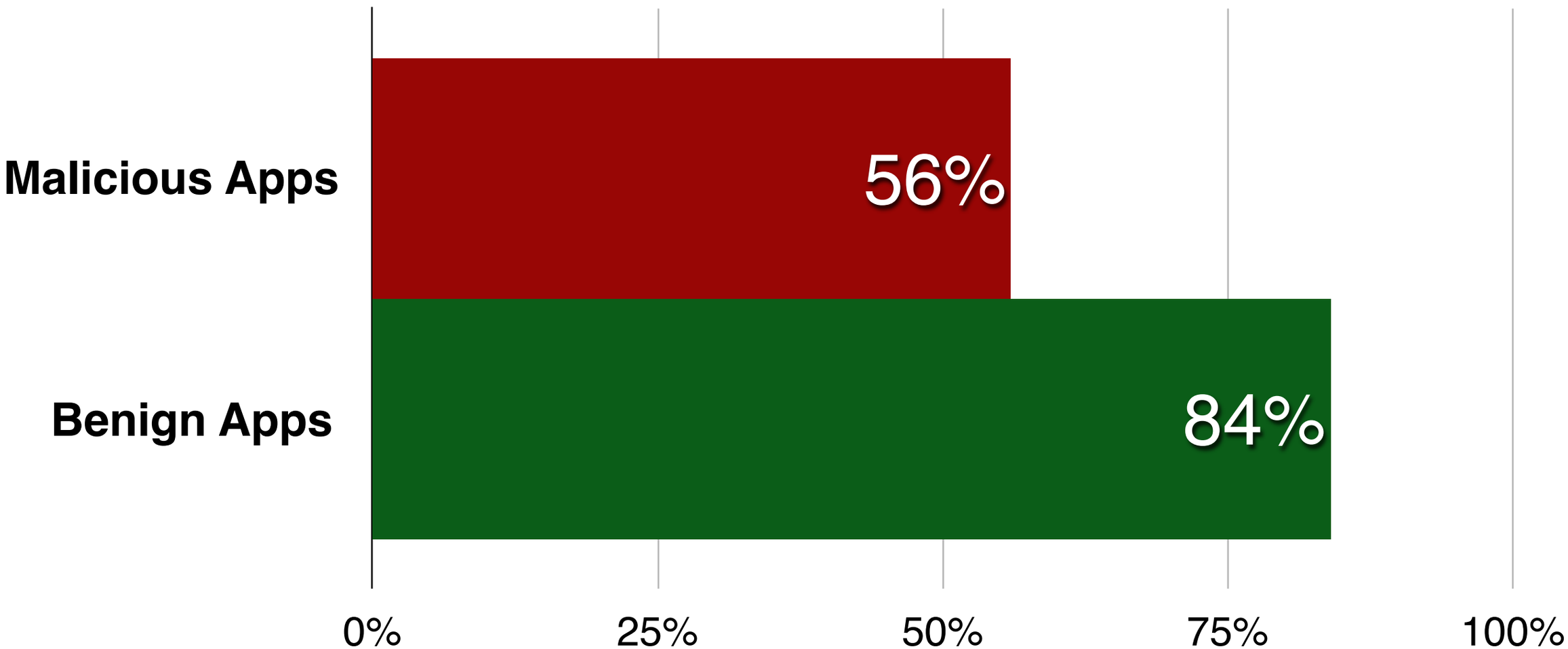


Correct Classification

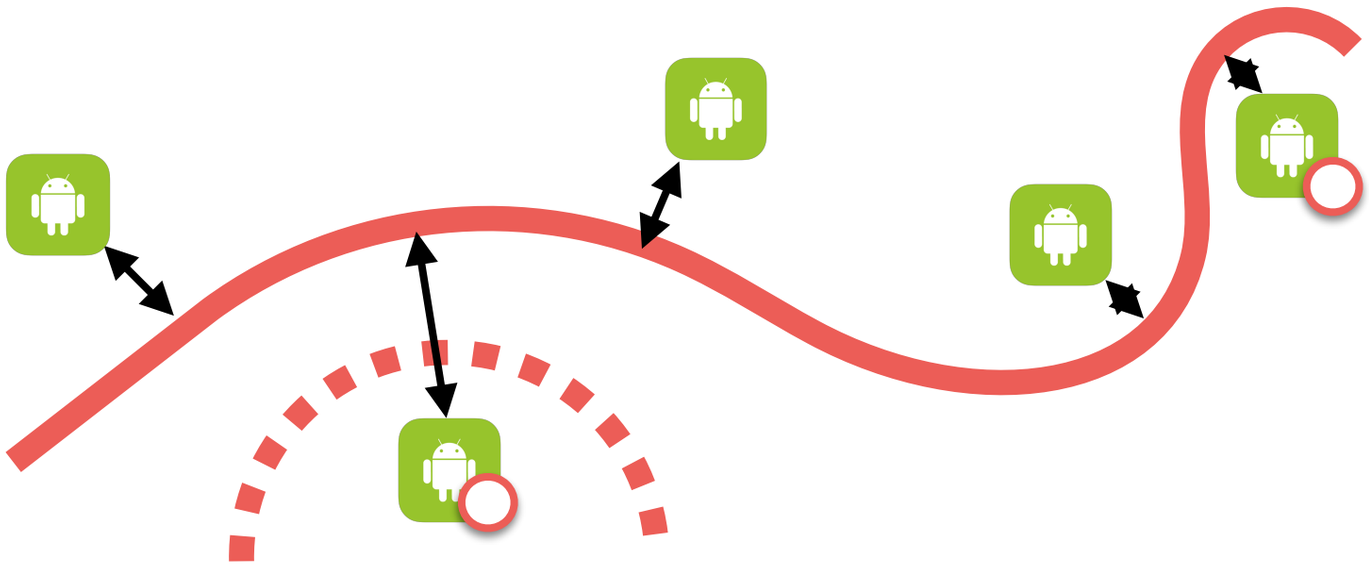


Correct Classification

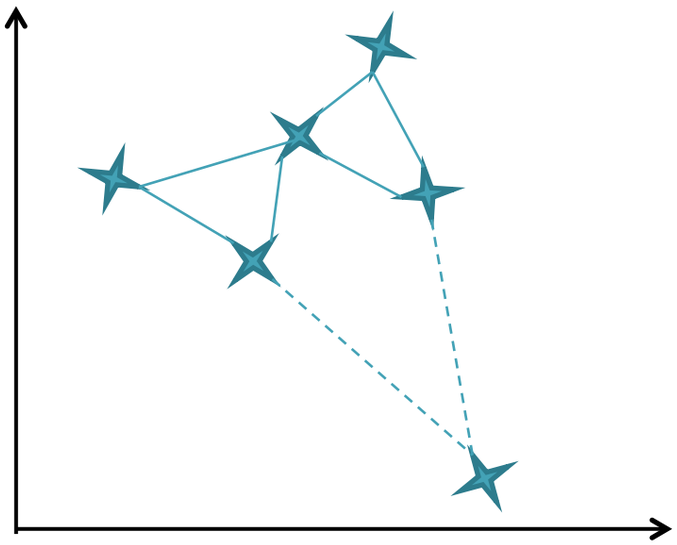
With Clusters (our approach)



Better anomaly detection



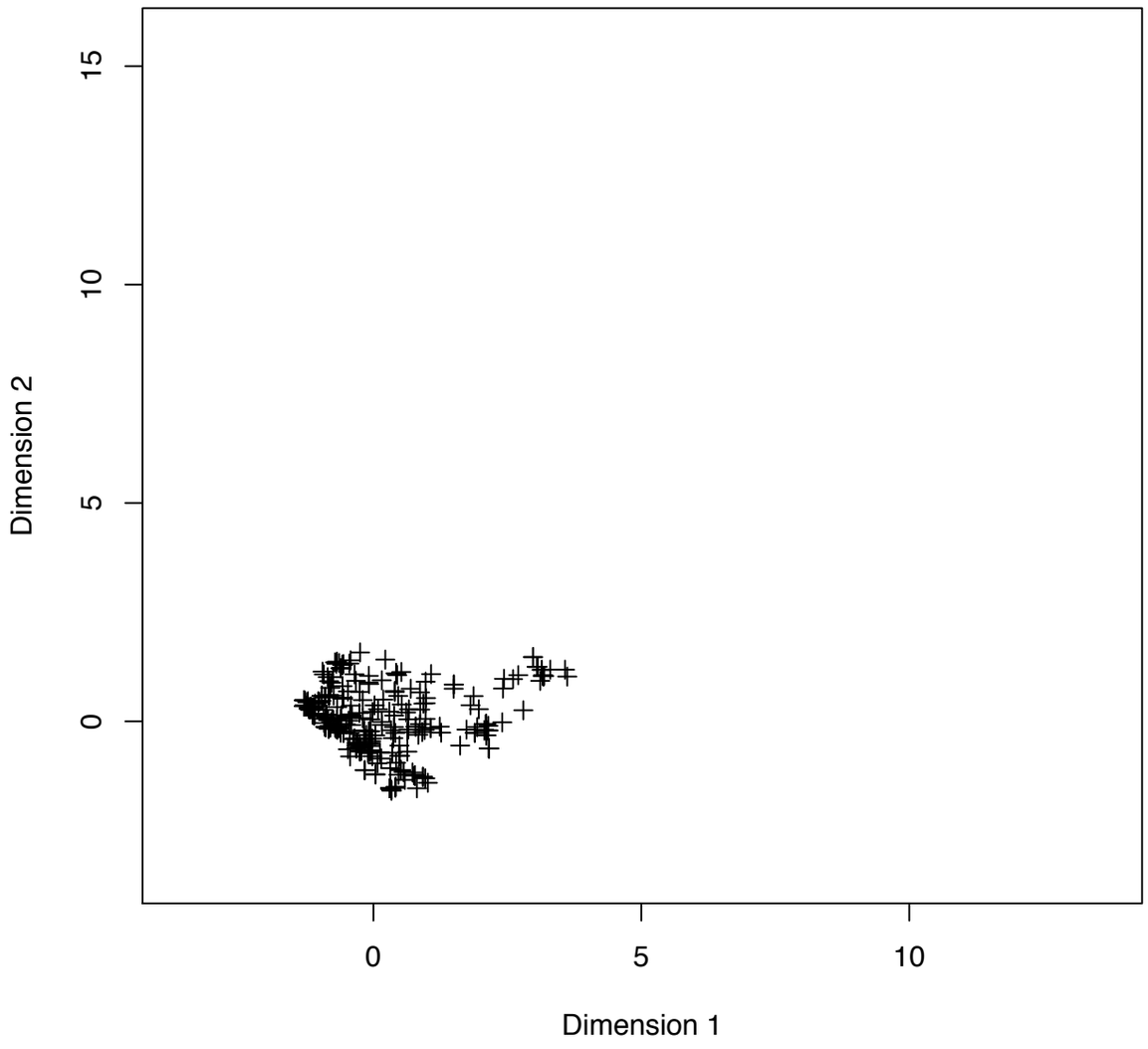
OC-SVM



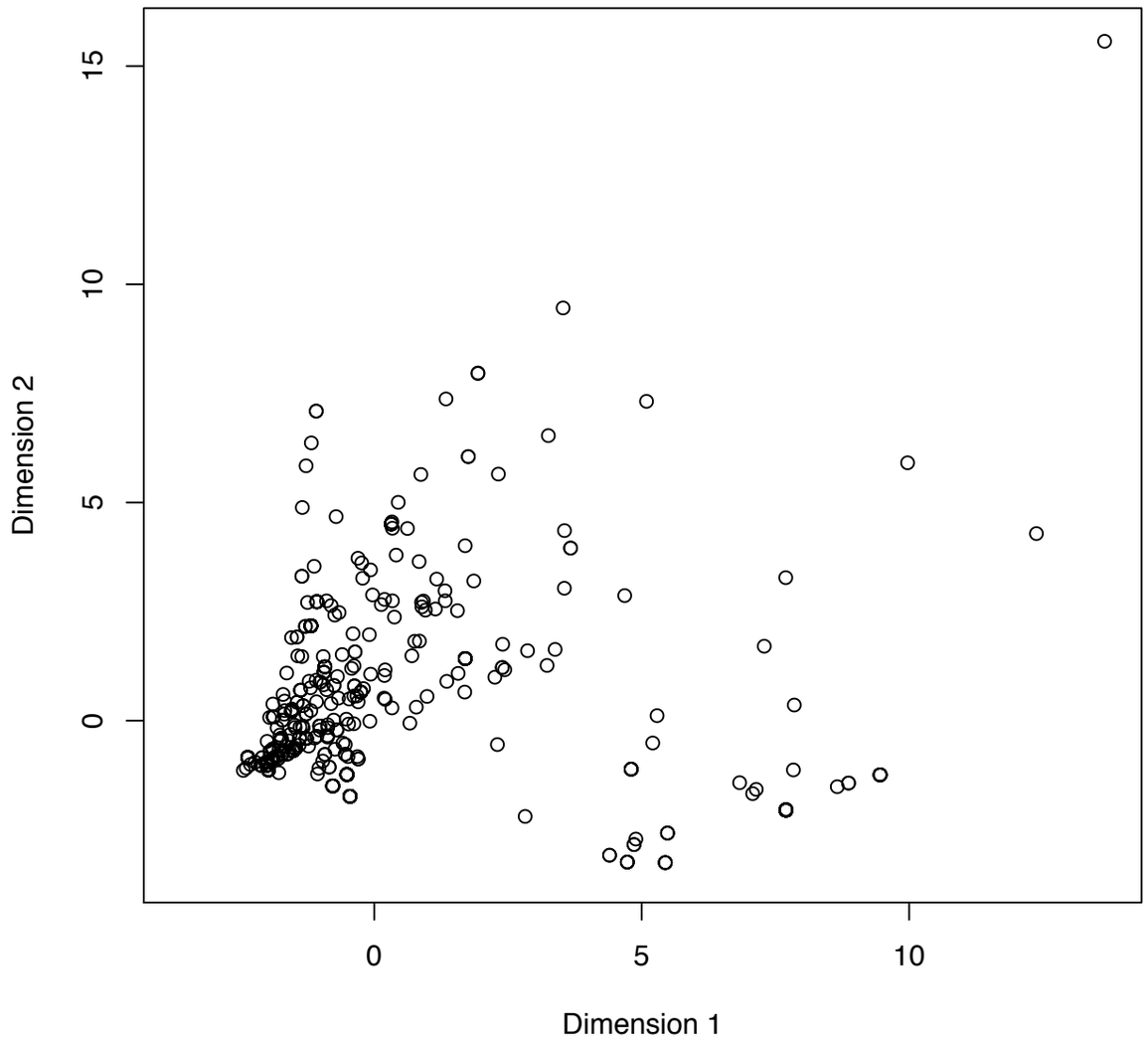
K-nearest neighbours

Better anomaly detection - API weight

cluster 29

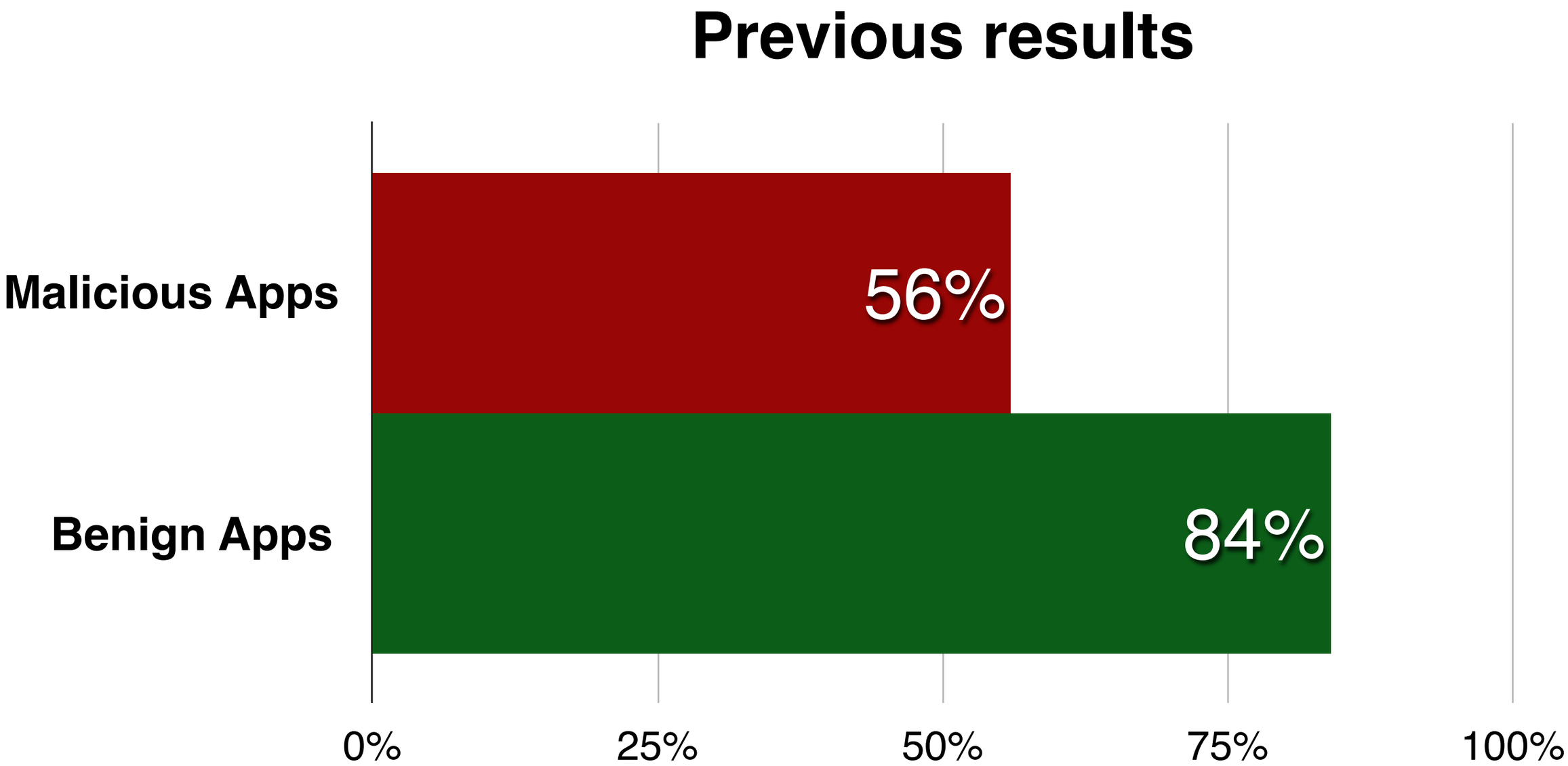


No weight

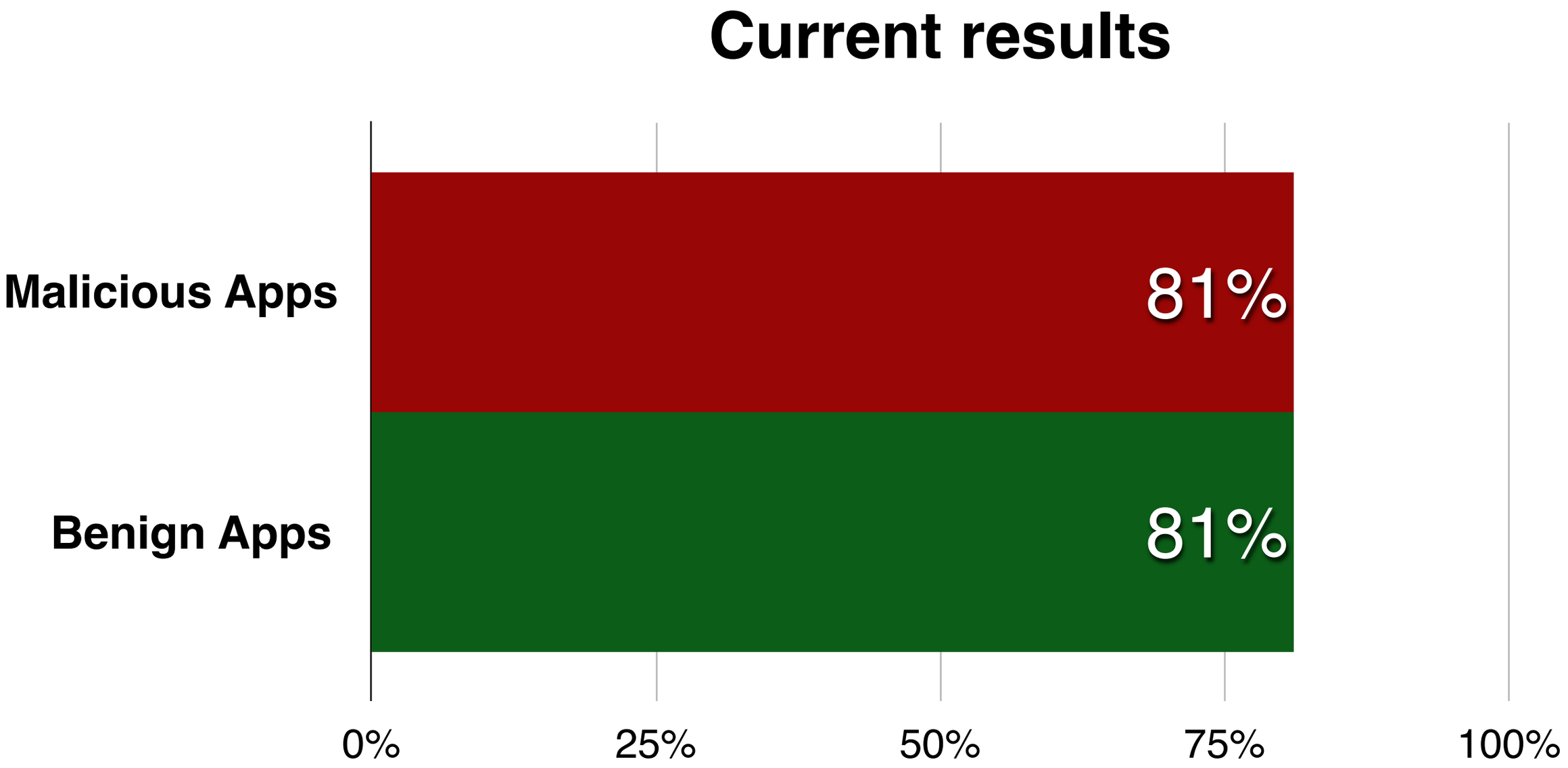


Weight with TF-IDF

Better anomaly detection



Better anomaly detection

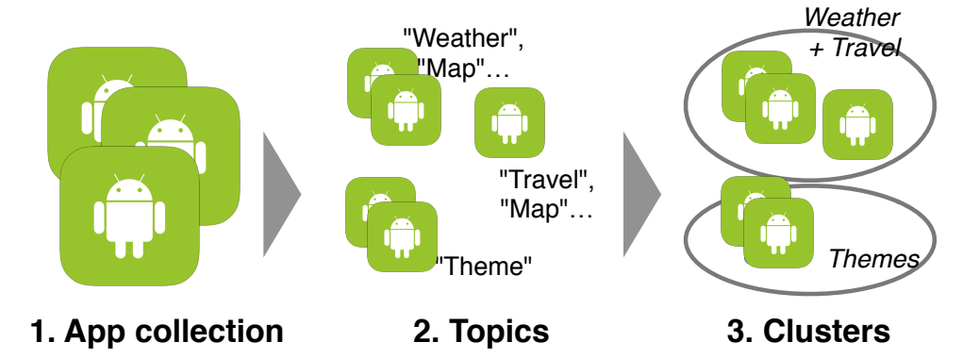


CHABADA:
Checking App Behavior Against App Descriptions

Alessandra Gorla
Saarland University, Germany

joint work with Konstantin Kuznetsov, Ilaria Tavecchia, Florian Gross and Andreas Zeller

CHABADA

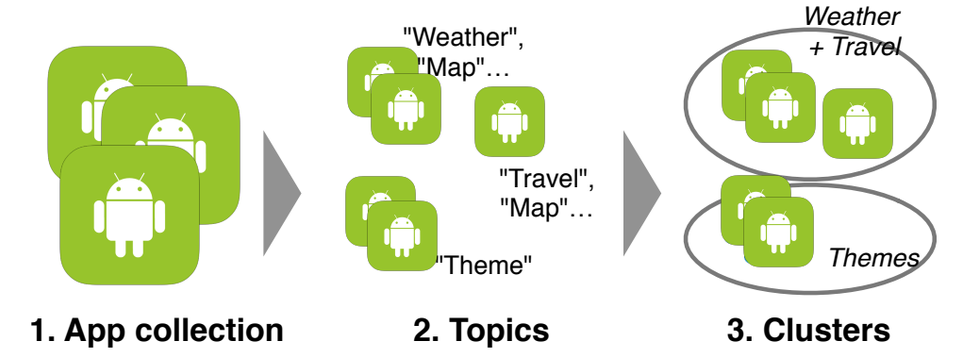


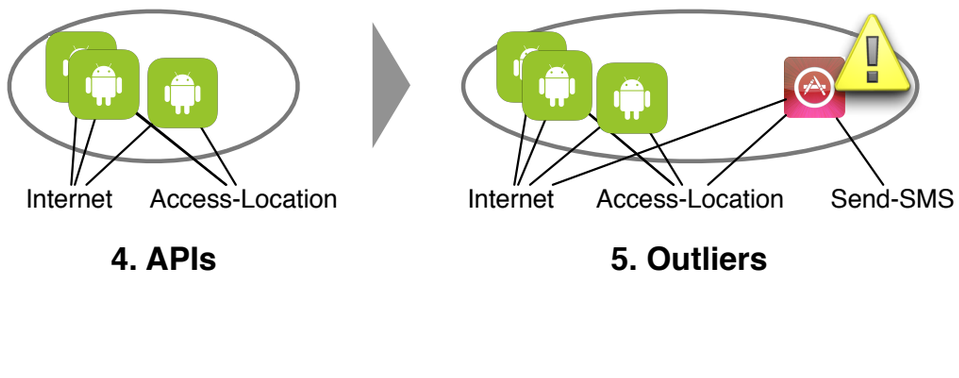
BADA:
pp Behavior
Descriptions

Alessandra Gorla
Saarland University, Germany

joint work with Konstantin Kuznetsov, Ilaria Tavecchia, Florian Gross and Andreas Zeller

CHABADA





“Travel” cluster

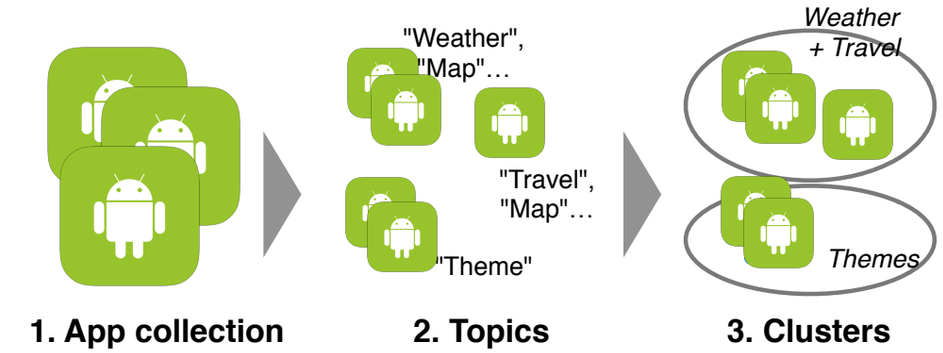
Description

Permissions of APIs used

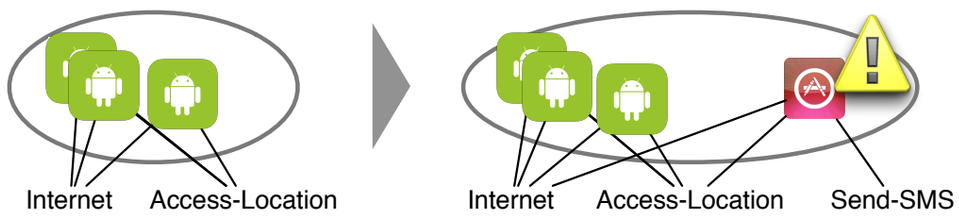
Alessandra Gorla
Saarland University, Germany

joint work with Konstantin Kuznetsov, Ilaria Tavecchia, Florian Gross and Andreas Zeller

CHABADA



1. App collection
2. Topics
3. Clusters



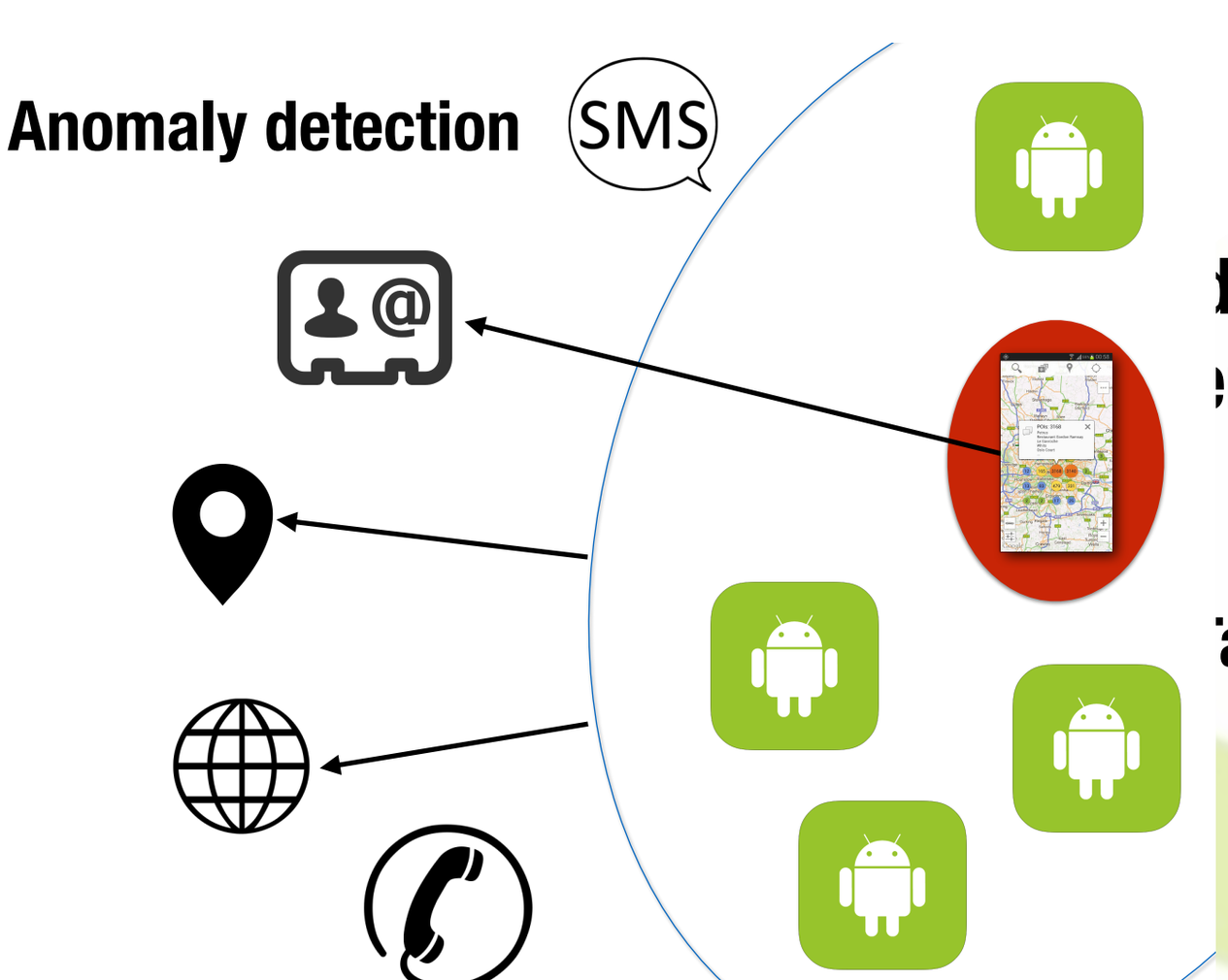
4. APIs
5. Outliers

“Travel” cluster

Description

Permissions of APIs used

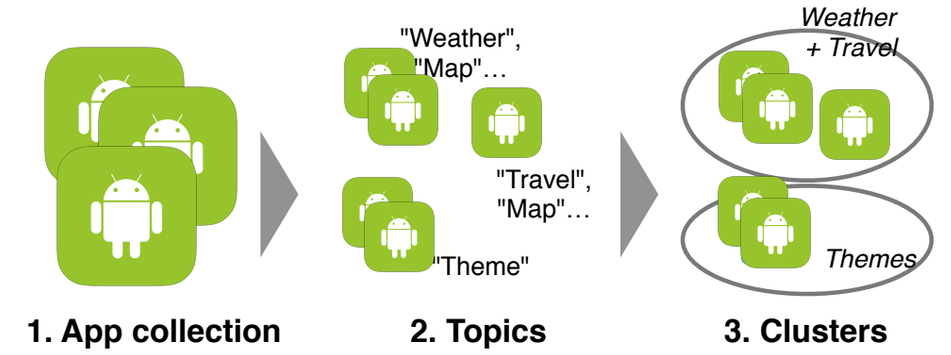
Anomaly detection



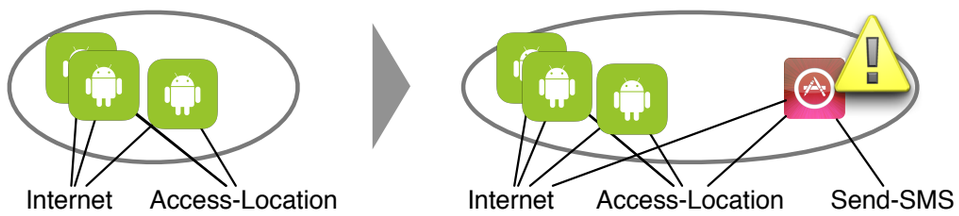
dra Gorla
ersity, Germany

avecchia, Florian Gross and Andreas Zeller

CHABADA



1. App collection
2. Topics
3. Clusters

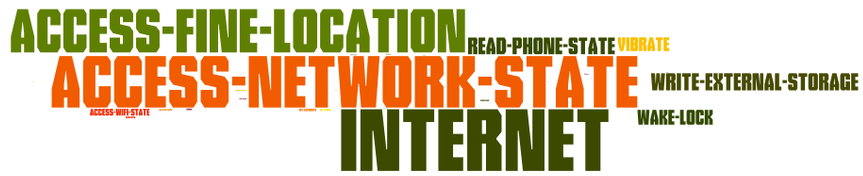


4. APIs
5. Outliers

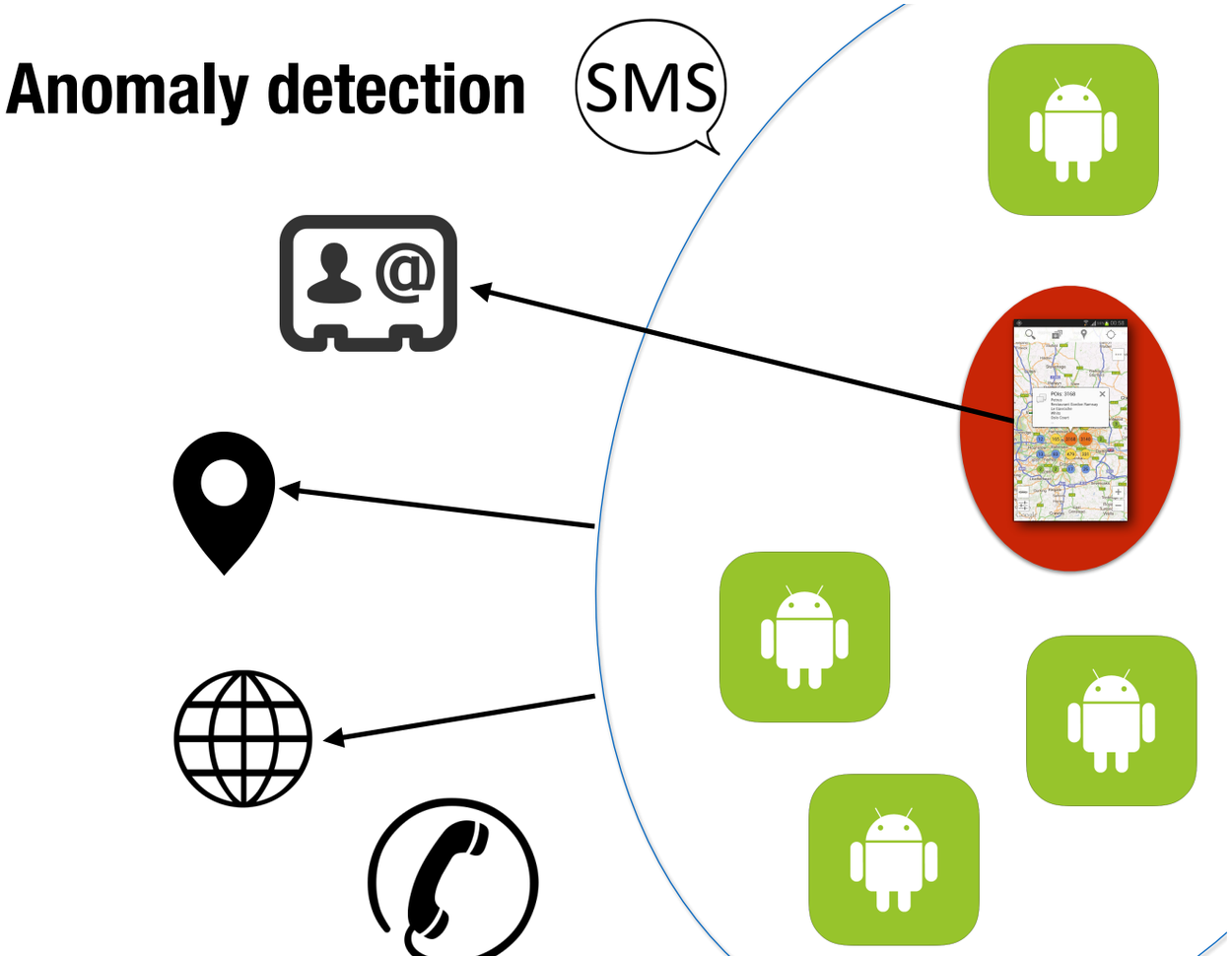
"Travel" cluster

Description

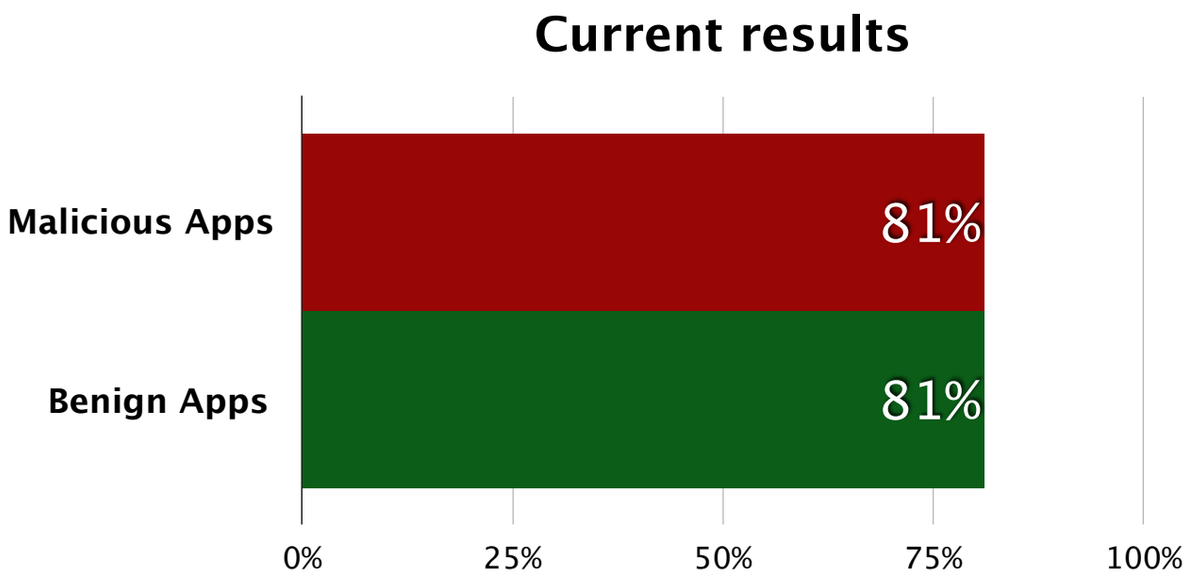
Permissions of APIs used



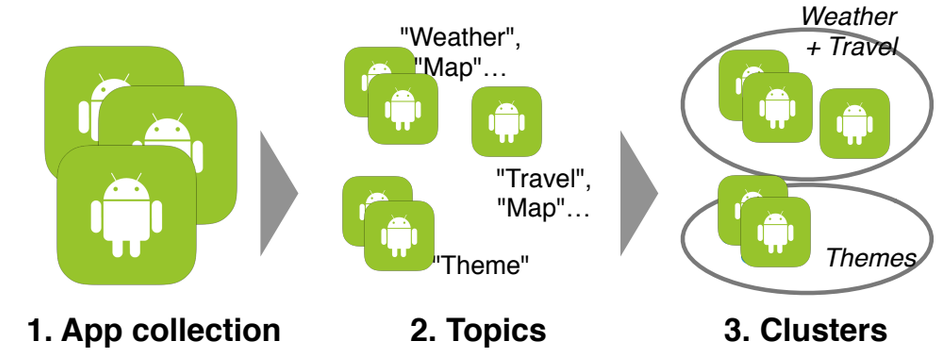
Anomaly detection

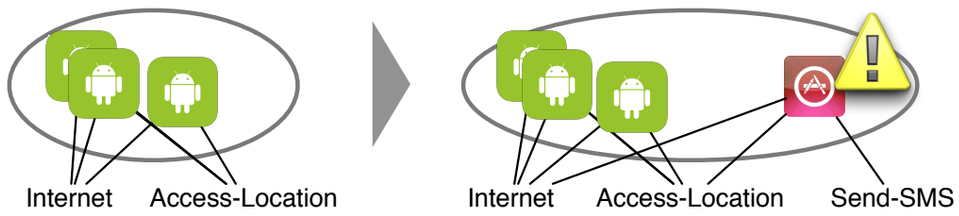


Better anomaly detection



CHABADA



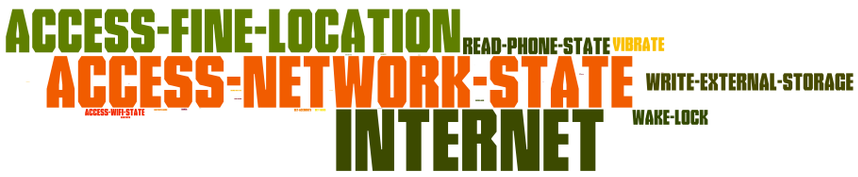


4. API

"Travel" cluster

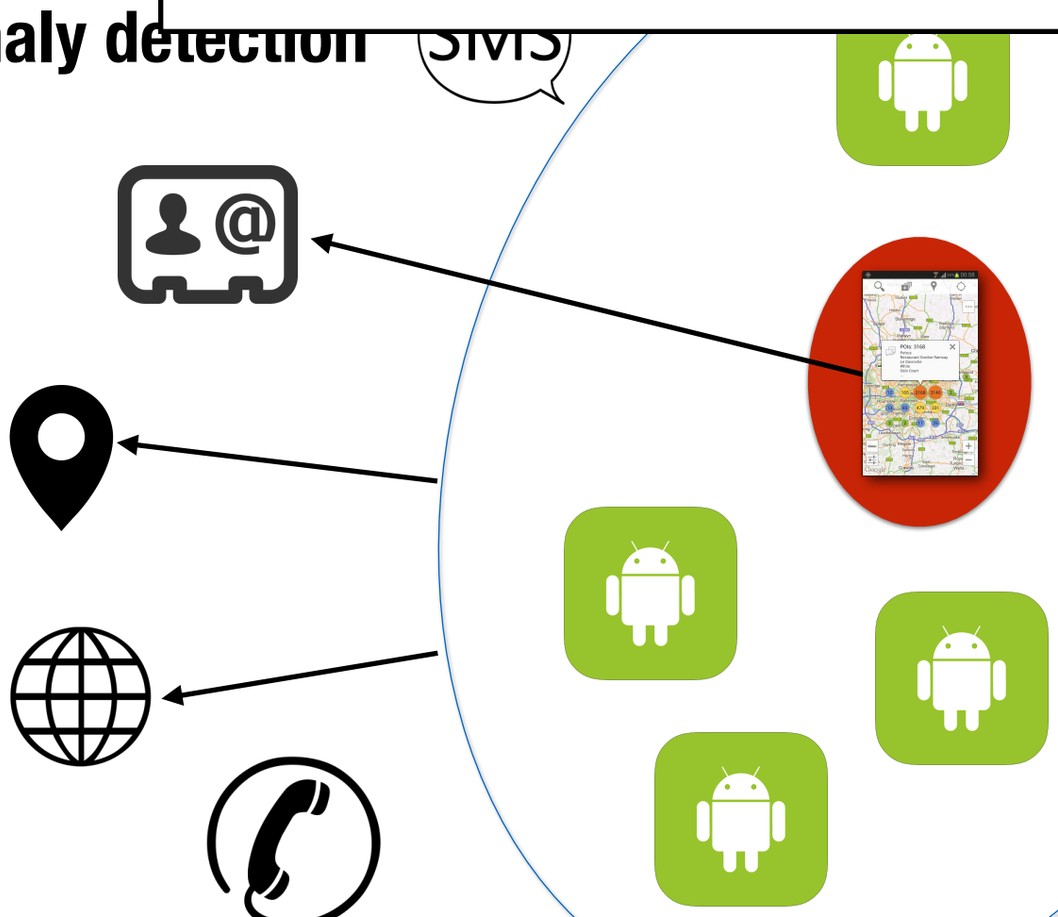
Description

Permissions of APIs used



www.st.cs.uni-saarland.de/chabada

Anomaly detection



Better anomaly detection

Current results