Android Compat Analysis Summary

Original Data Source:

A. Provided sets of installation and runtime logs.

1. Installation log – Benign 2018-2019, Malware 2018-2019 *(/media/SeagateBack/xqfu/InstallationReSults/ on the server 10.99.1.192)*
2. Runtime log – Benign 2018-2019, Malware 2010-2019 *(at /home/xqfu/runtimeResults on the server 10.99.1.192)*

B. Provided sets of source apks

1. /pool/home/zyzhang/myDownloads on 10.99.1.191
2. /media/SeagateBack/xqfu/myDownloads on 10.99.1.192
3. /storage_array/home/xqfu/andro-data/myDownloads on 10.99.1.190
4. /storage_array/home/xqfu/andro-data/AndroZoo on 10.99.1.190

C. Provided .csv file, each csv contains three apk identification and one app name that they belong to.

1. Benign 2018-2019 and Malware 2018-2019 (at /pool/home/zyzhang/myDownloads on 10.99.1.191)
   (NOTE that we should also need .csv 2010-2019 for both benign and malware)

NOTICE: $for\ all\ apk\ in\ A\ must\ \subseteq\ for\ all\ apk\ in\ B$, so to guarantee 100% coverage for SSPS of A, so far the coverage of SSPS of A << 100%.

Filtered Data Set:

D. ./InstallResult

- Contains a set of text files, each of which with the identity {apkYear, apiLevel}, that describe the total apks that are examined, the number of apks that are compatible, and the number of apks that are incompatible. Each compatible or incompatible apk is listed with its original identification number and further that each incompatible apk listed with its failure message.
- How do I obtain these text files? **NEED A.1 original source file**
  - bash ClassifierScript/runInstallTraces.sh **InstallationLogAddress**
  - will then saved a set of text files ./InstallResult (create folder if not existed)

E. ./RuntimeResult

- Contains a set of text files, each of which with the identity {apkYear, apilevel}, that describe the total apks that are examined, the number of apks that are compatible, and the number of apks that are incompatible. Each compatible or incompatible apk is listed with its original identification number and further that

each incompatible apk listed with its failure message.

- How do I obtain these txt file? **NEED A.2 original source file**
  - python ClassifierScript/runtime_classify.py **RuntimeLogAddress**
  - will then saved a set of text files ./RuntimeResult (create folder if not existed)

F. ./DataParse/malware-minsdk
./DataParse/benign-minsdk

- Each folder contains set of text files, each of which with identity {apkType, apkYear}, describe each apk with its minSDKlevel.
- How do I obtain these text file? **NEED B original source file**
  - python DataParse/parseMinSdk.py FolderContainsApks "apkType" "apkYea"
  - will then saved the results in a folder at ./DataParse/

G. ./MultuCompatResult

- Contains a set of text files that describe apps that are multi-compatible (meaning at least some of its belonging apks are compatible with covered API19, 20-27) or multi-incompatible (all of its belonging apks did not compatible cover API19, 20-27).
- How do I obtain these text file? **NEED C and (D or E)**
  - python MultiClassify .csvDir InstallResult/RuntimeResultDir
  - will then saved the results in a folder at ./MultiCompatResult/

Tables and Figures

H. Data/Tables.xlsx

- How do I obtain these data?
  - Python TableOneStat.py A.1orA.2dir
  - Print to console then manually collect

Benign Data use	number of samples from each year within 2010-2019										Total
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
Installation-time incompatibility study	16835	9977	10991	9688	5300	5406	2431	2266	37838	30236	130968
Run-time incompatibility study	1531	2020	2054	1750	1335	327	1548	1680	13171	11787	37203

Malware Data use	number of samples from each year within 2010-2019										Total
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
Installation-time incompatibility study									32523	20270	52793
Run-time incompatibility study	16173	11226	11731	12094	14418	11411	10845	11926	12245	12205	124274

- How do I obtain these data? NEED SSPS data
  - Python SpearmanCof.py

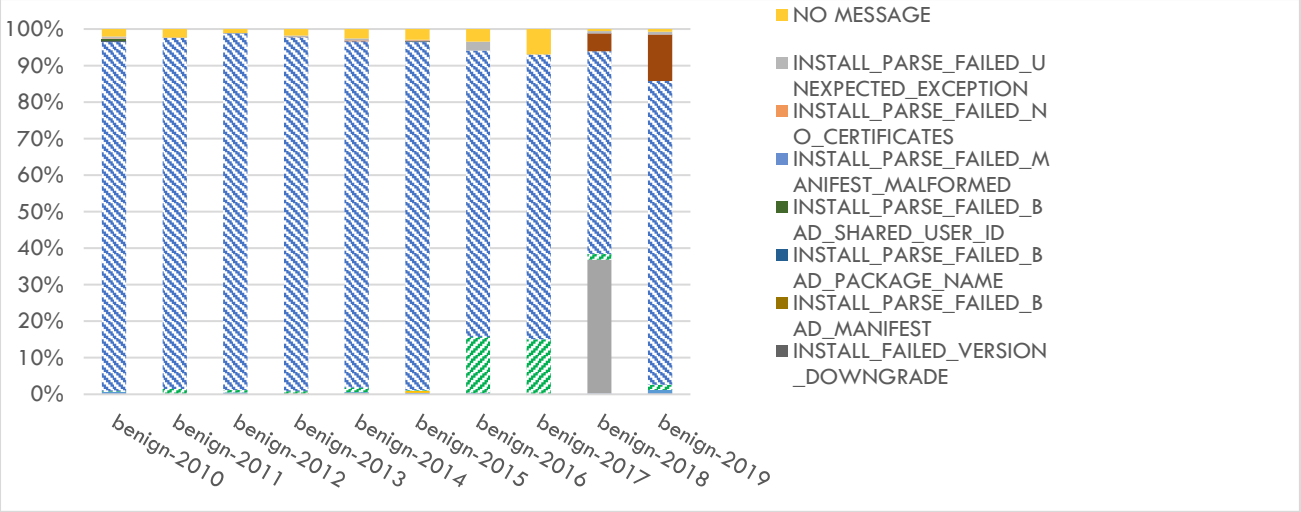
installation-time incompatible ratio (Malware 2018-2019)	App lapse	API lapse	minSDKVersion	SDK API Level	App year
overall	-0.407216	**-0.160562**	**0.048492**	-0.431729	0.026229
INSTALL_FAILED_NO_MATCHING_ABIS	-0.297657	**-0.078221**	**0.002708**	-0.311617	0.033715
INSTALL_FAILED_MISSING_SHARED_LIBRARY	-0.06168	**0.009026**	**-0.033328**	-0.07099	-0.046474

installation-time incompatible ratio (Benign 2010-2019)	App lapse	API lapse	minSDKVersion	SDK API Level	App year
overall	0.006858	**0.399003**	**-0.460116**	-0.199884	-0.05925
INSTALL_FAILED_NO_MATCHING_ABIS	Nan	**Nan**	**Nan**	Nan	Nan
INSTALL_FAILED_MISSING_SHARED_LIBRARY	-0.004548	**0.301023**	**-0.333095**	-0.110004	-0.022058

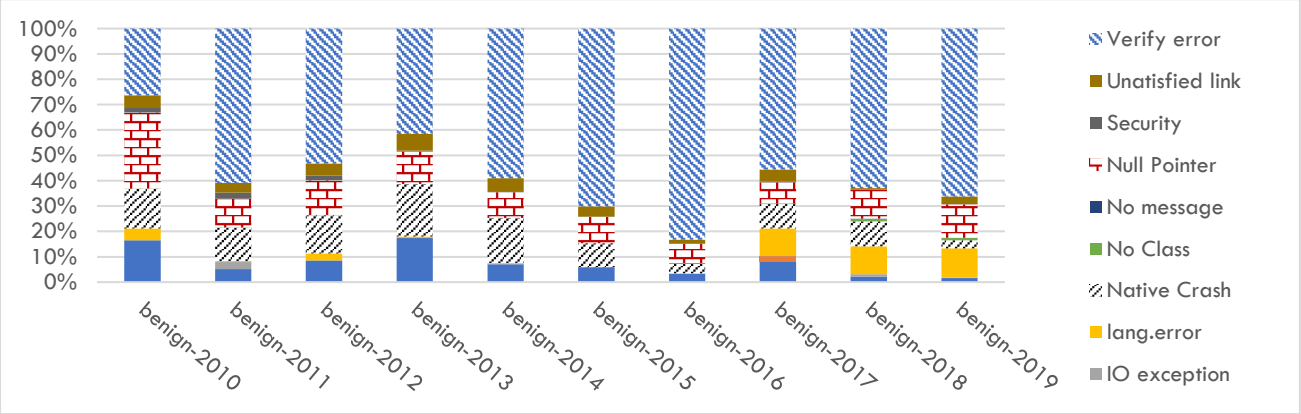
run-time incompatible ratio (Benign 2010-2019)	App lapse	API lapse	minSDKVersion	SDK API Level	App year
overall	**-0.289723**	0.106744	-0.258521	**-0.496454**	-0.04574
native crash	0.168285	0.181223	-0.212575	-0.057607	-0.212491
verify error	**Nan**	Nan	Nan	**Nan**	Nan

run-time incompatible ratio (Malware 2010-2019)	App lapse	API lapse	minSDKVersion	SDK API Level	App year
overall	**-0.216178**	0.176801	-0.271787	**-0.320694**	0.082923
native crash	-0.181285	0.087321	-0.132916	-0.153762	0.114699
verify error	**-0.172503**	0.149568	-0.227015	**-0.262857**	0.062887

I. Data/Bar(completed)-Benign.xlsx
- How do I obtain these data?
  - Python BarStatIns.py InstallResult
  - Print to console then manually collect

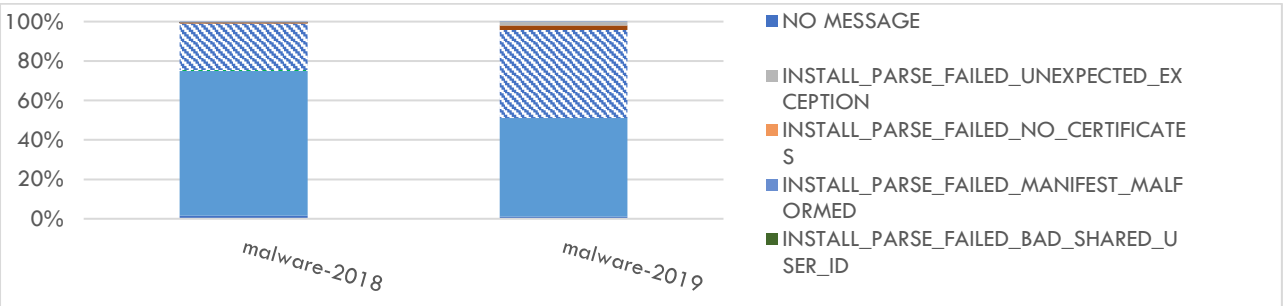


- How do I obtain these data?
  - Python BarStatRuntime.py RuntimeResult
  - Print to console then manually collect

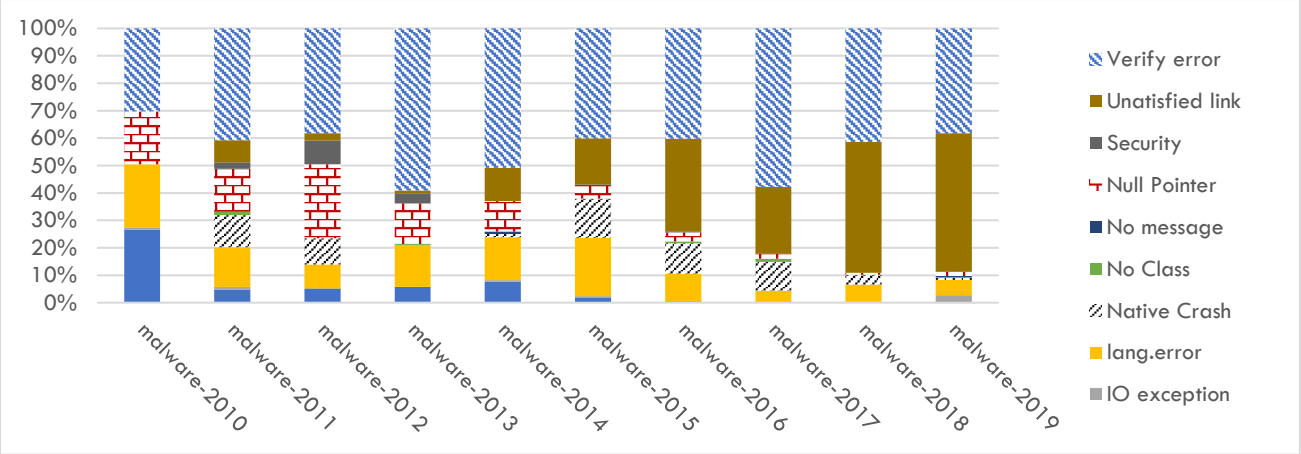


J. Data/Bar(completed)-Malware.xlsx

- How do I obtain these data?
  - Python BarStatIns.py InstallResult
  - Print to console then manually collect

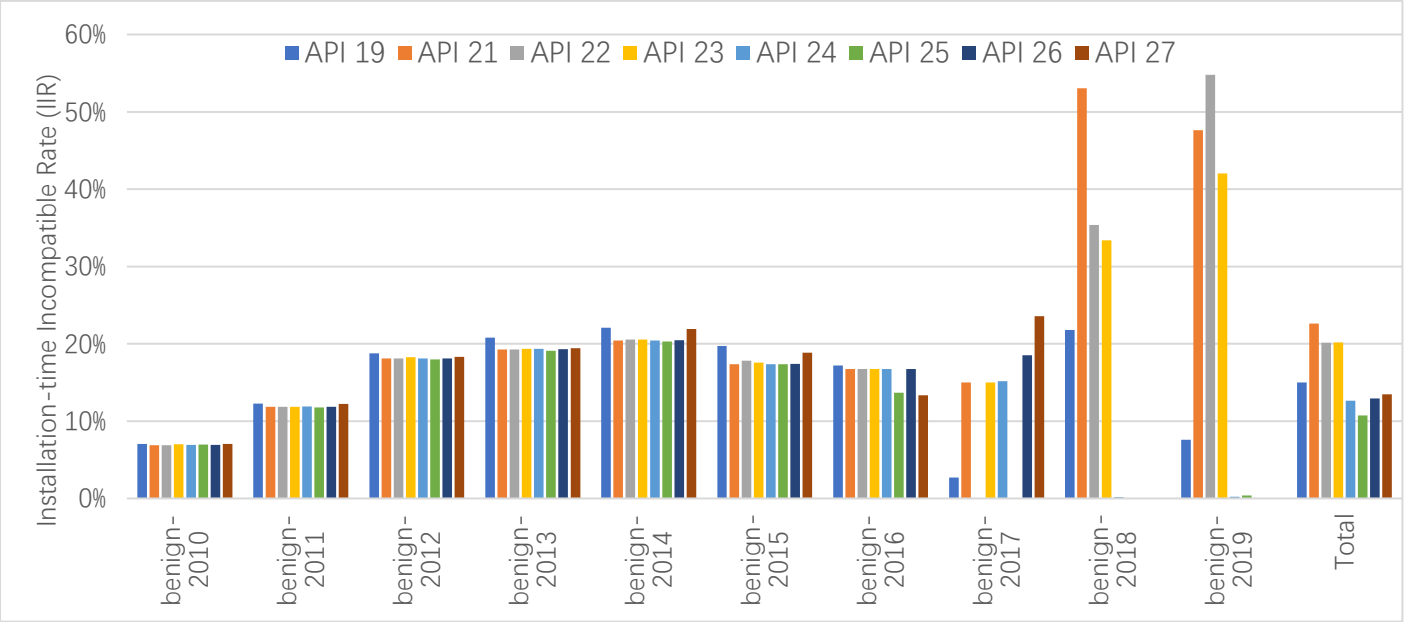


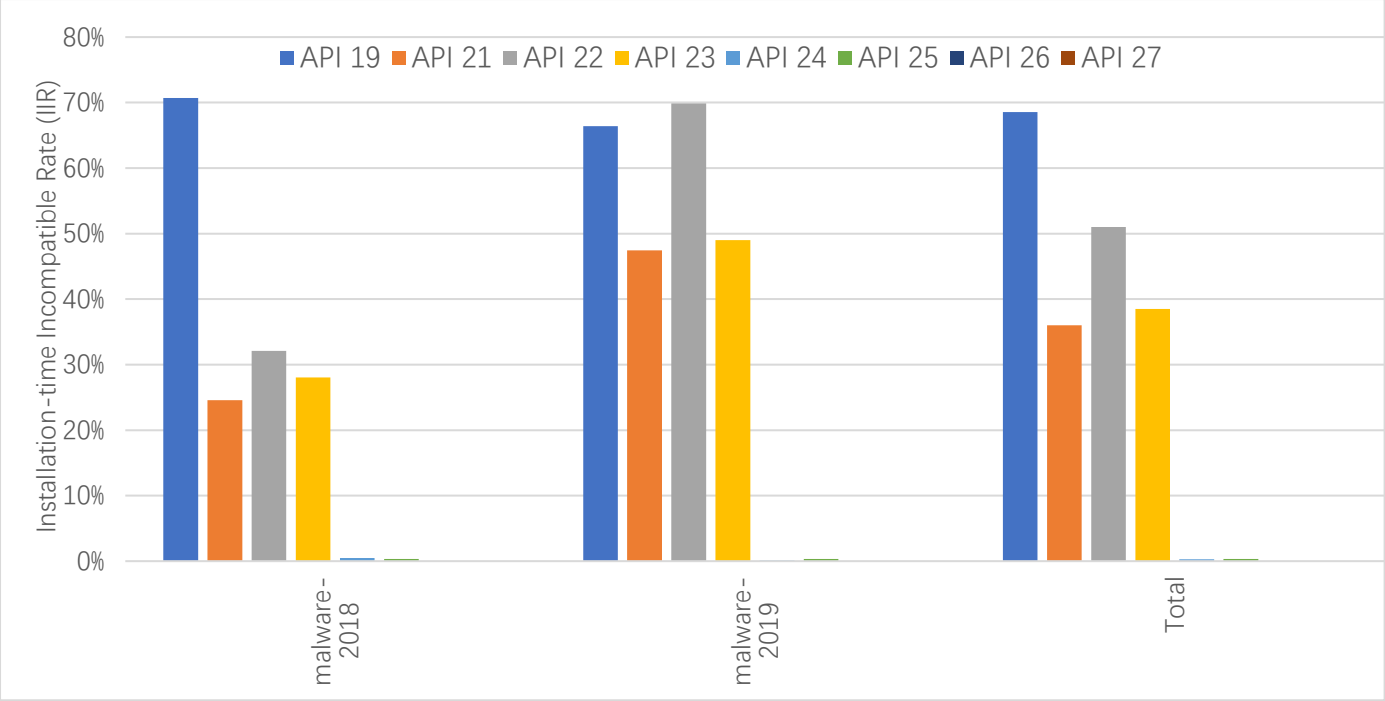
- How do I obtain these data?
  - Python BarStatRuntime.py RuntimeResult
  - Print to console then manually collect



K. Data/InstallationIIR(completed)-Benign.xslx
Data/InstallationIIR(completed)-Malware.xslx

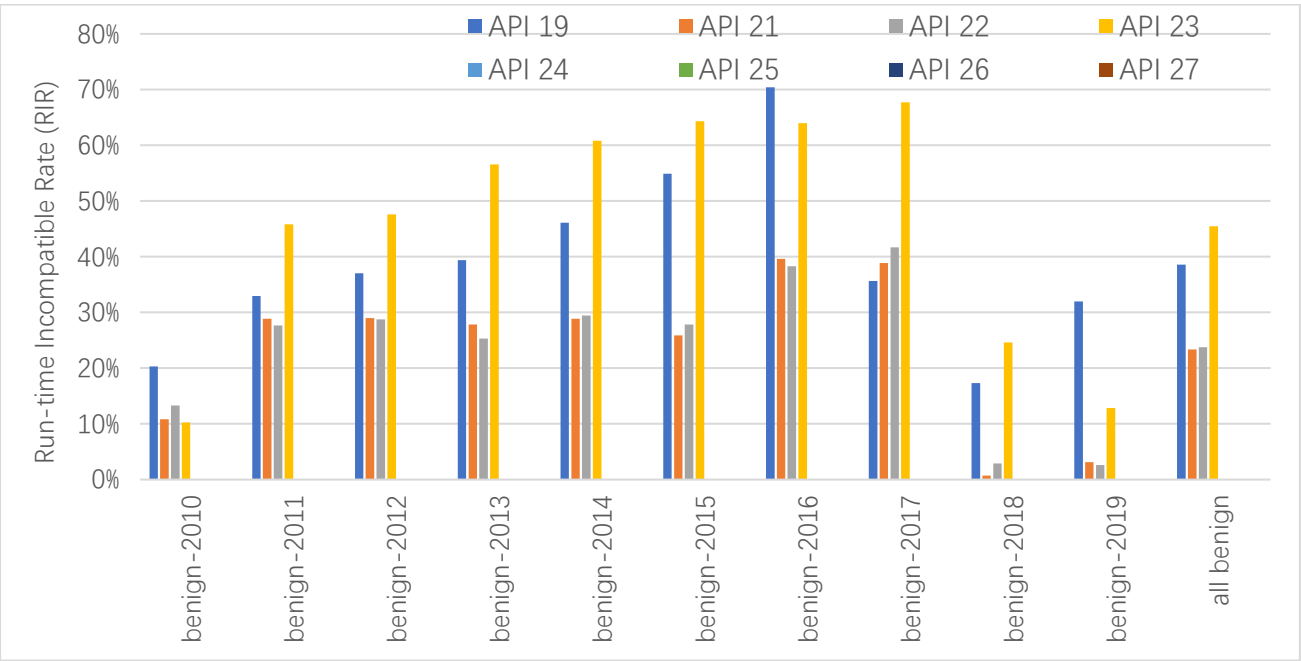
- Python Stats.py InstallResult
- Print to console then manually collect

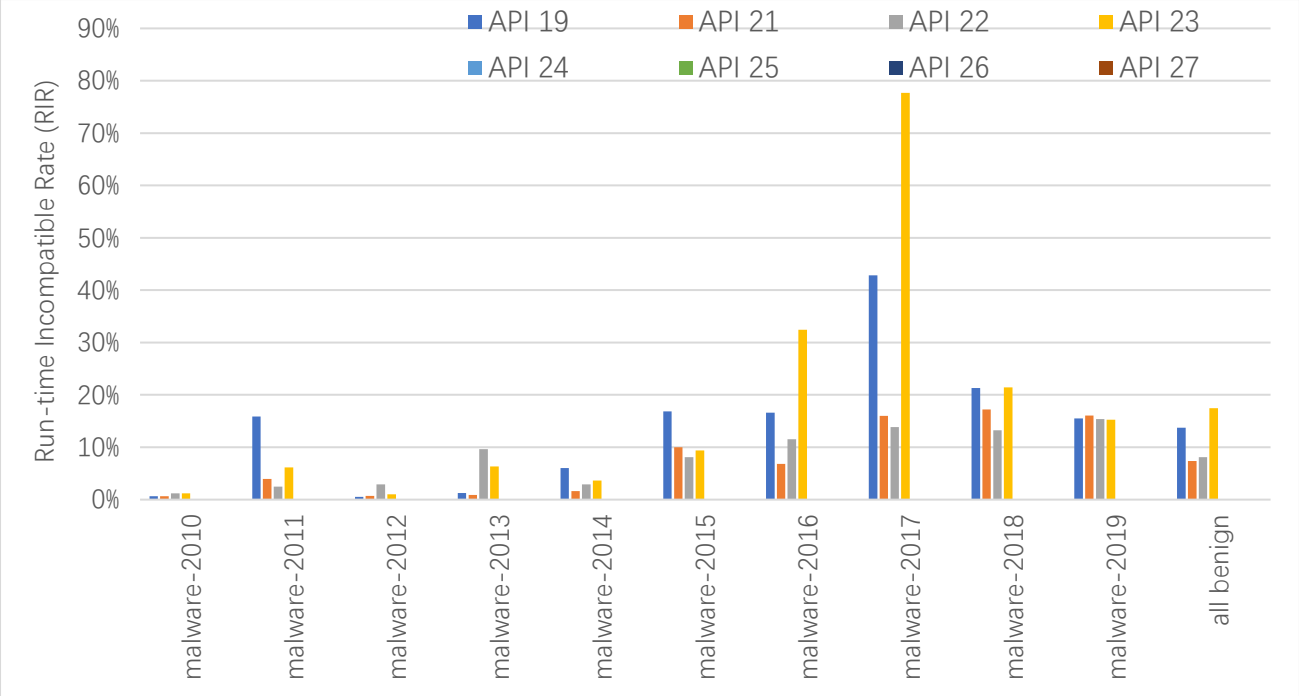




L. Data/RuntimeIIR(completed)-Benign.xlsx
Data/RuntimeIIR(completed)-Malware.xlsx

- Python Stats.py RuntimeResult
- Print to console then manually collect



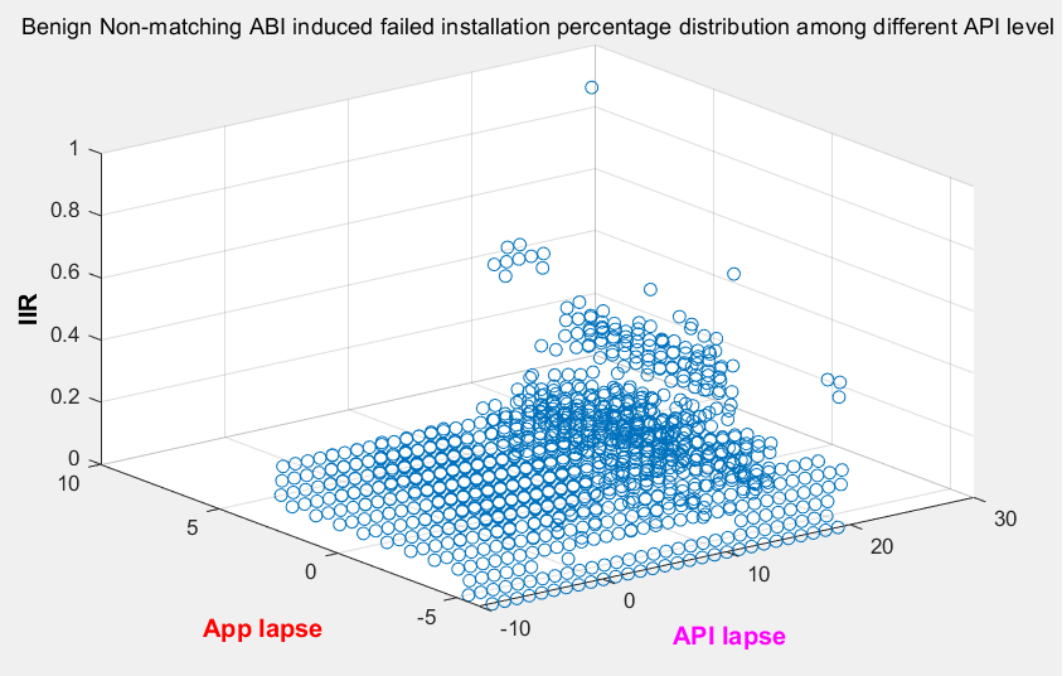


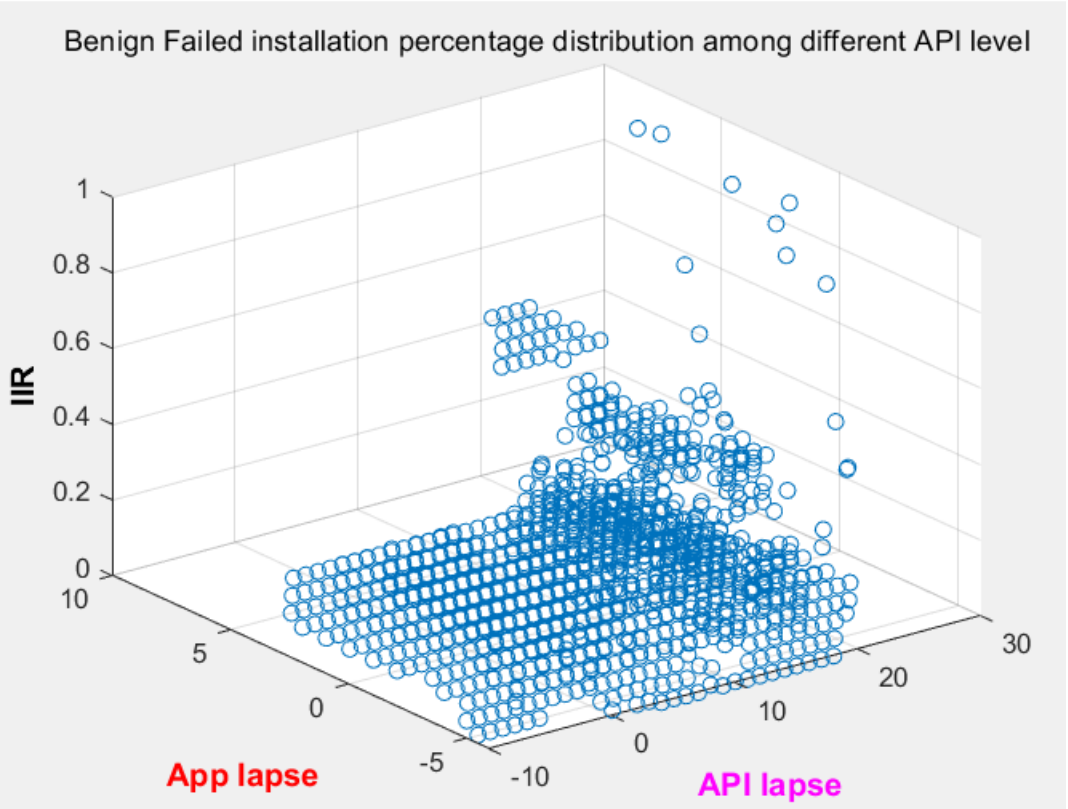
M. SSPS/*.m

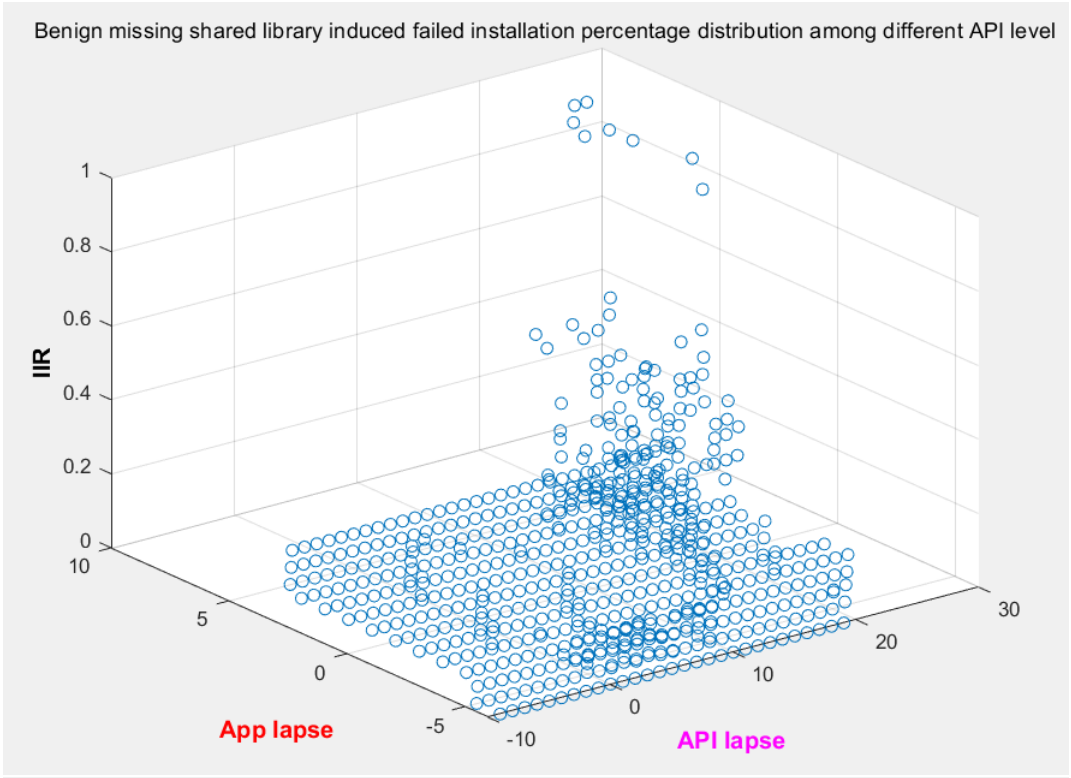
- Set of matlab script that generate the figures according the *.xlsx excel files in the current directory.

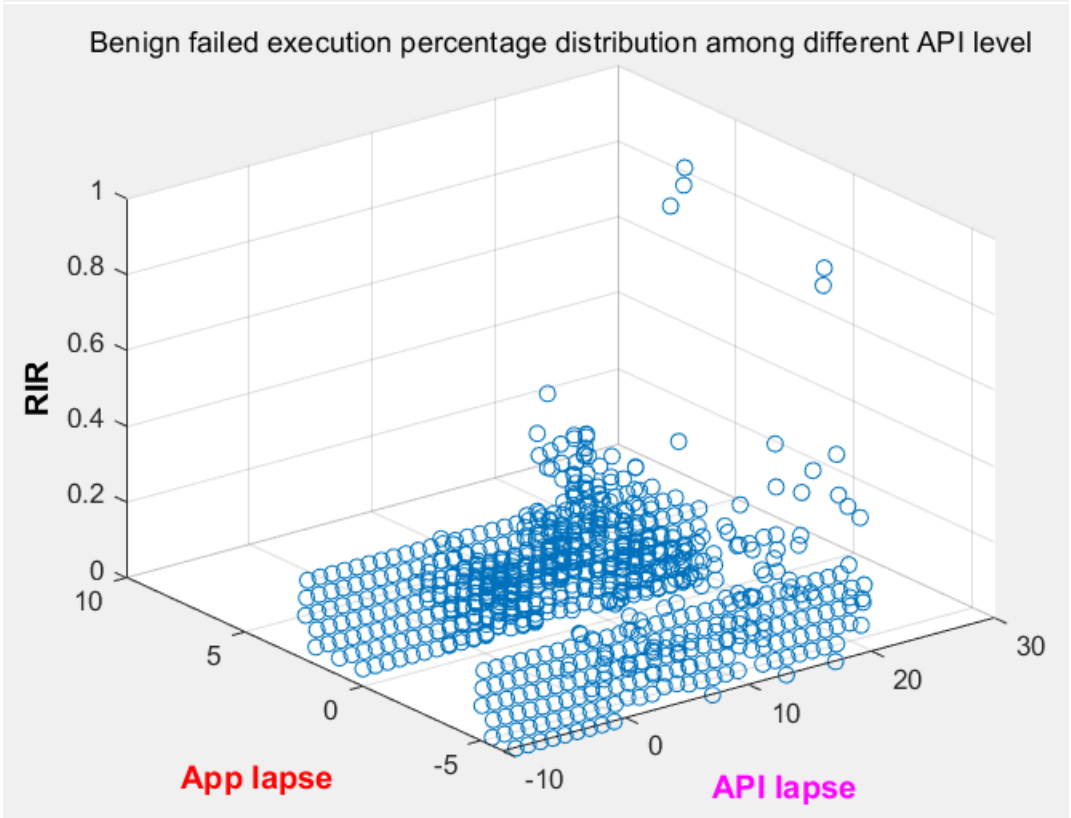
SSPS/*.xlsx

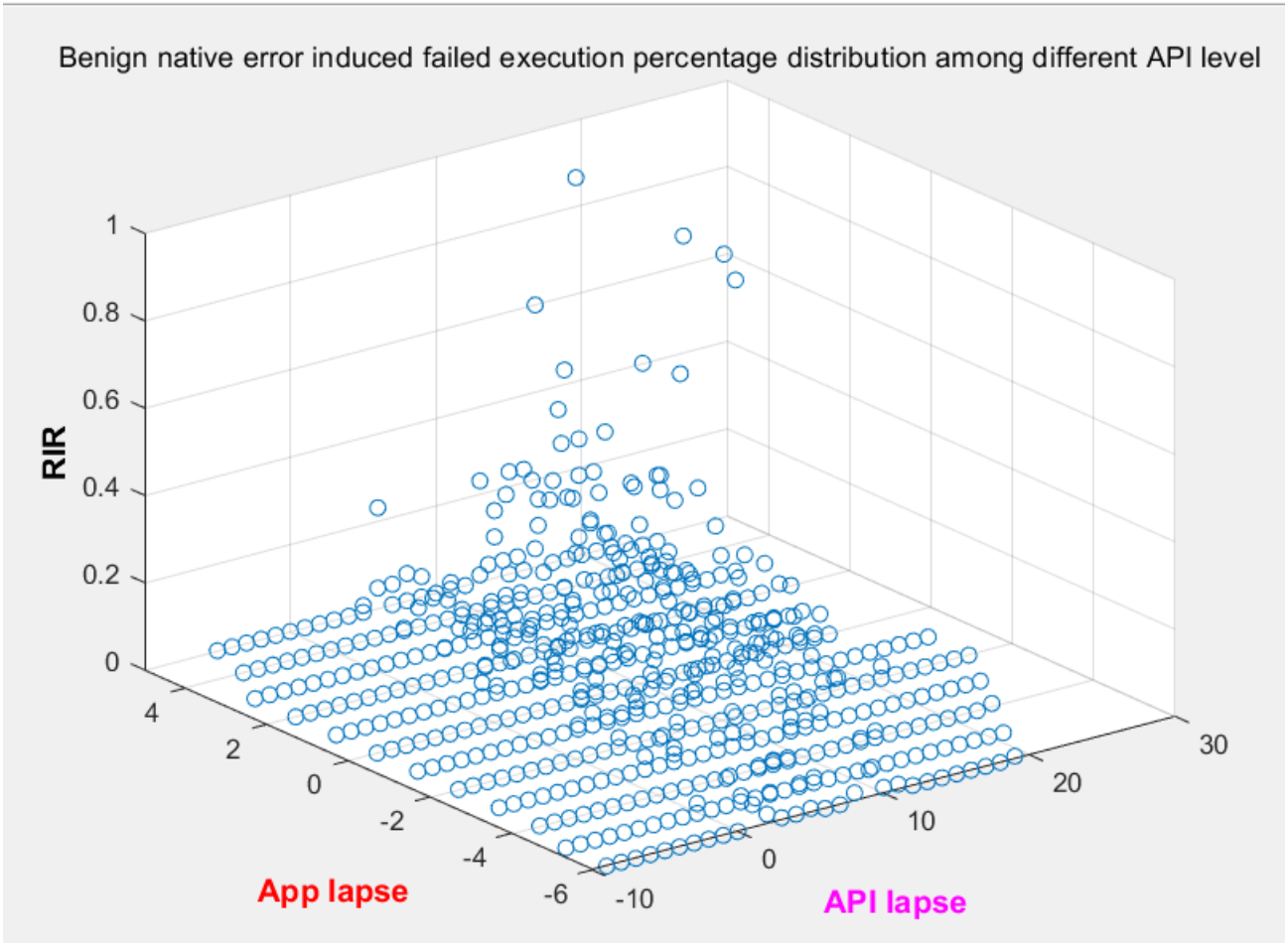
- Excel files for SSPS distribution.
- How do I obtain these data? **NEED A and C**
  - Python SSPSwrapper.py
  - Note that A and C dir are hardcoded.
- Generated:
  - SSPS/benignInstallTableABI.xlsx
  - SSPS/benignInstallTableLibrary.xlsx
  - SSPS/benignRuntimeTableVerify.xlsx
  - SSPS/benignRuntimeTableNative.xlsx
  - SSPS/malwareInstallTableABI.xlsx
  - SSPS/malwareInstallTableLibrary.xlsx
  - SSPS/malwareRuntimeTableVerify.xlsx
  - SSPS/malwareRuntimeTableNative.xlsx

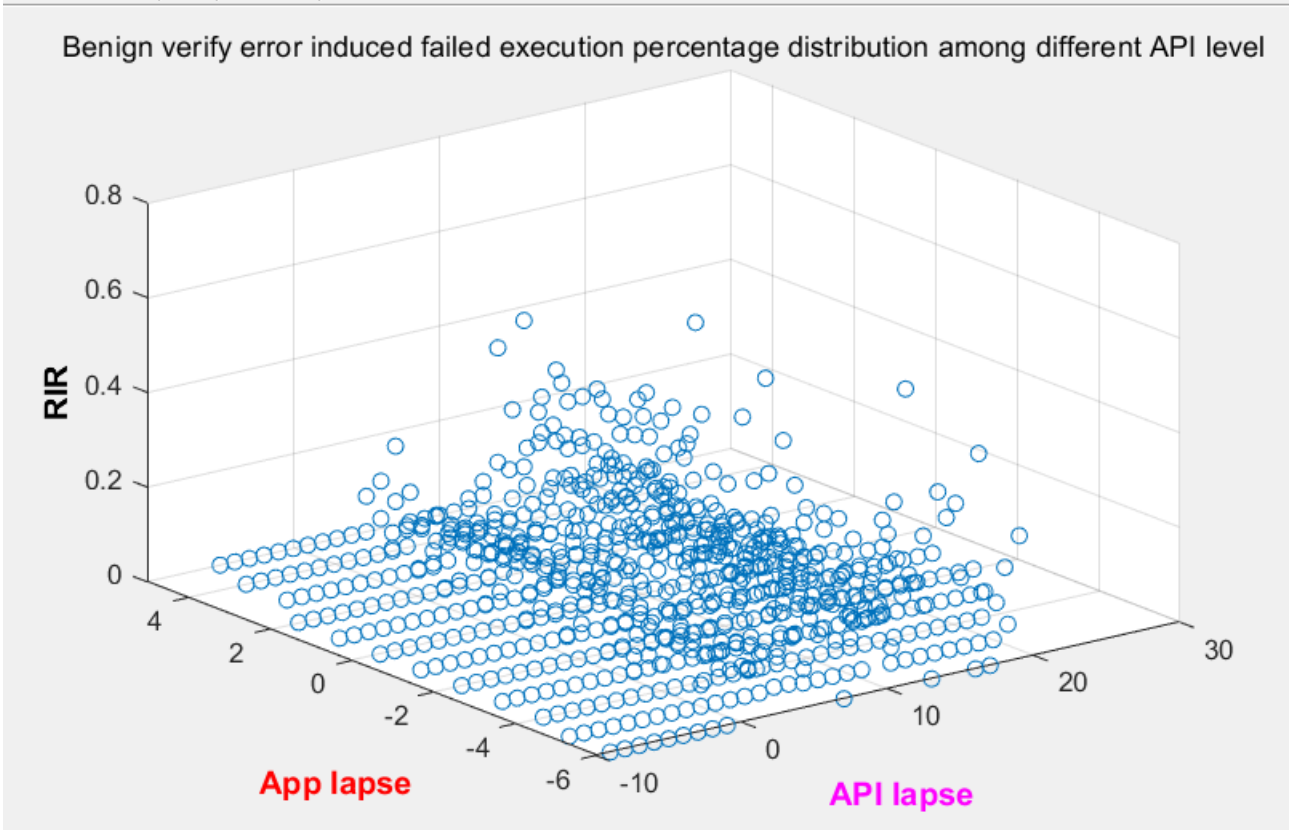

Benign Non-matching ABI induced failed installation percentage distribution among different API level
IIR
1
0.8
0.6
0.4
0.2
0
10
5
0
-5
-10
App lapse
30
20
10
0
API lapse

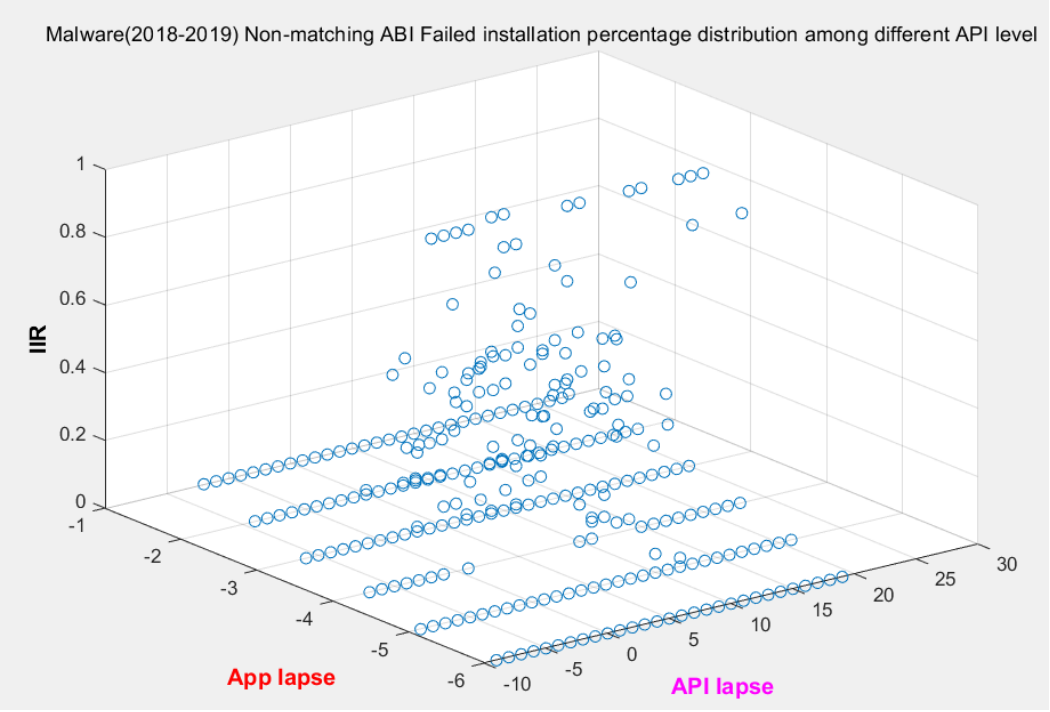

Benign Failed installation percentage distribution among different API level
IIR
1
0.8
0.6
0.4
0.2
0
10
5
0
-5
-10
App lapse
30
20
10
0
API lapse

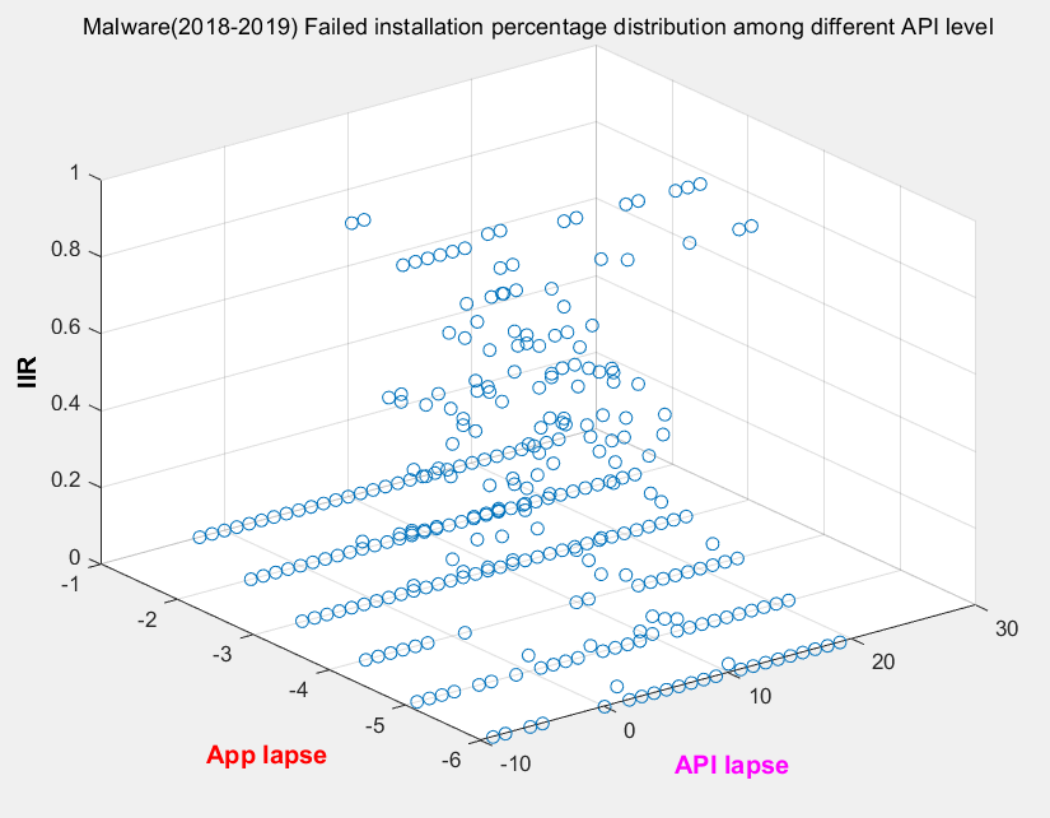

Benign missing shared library induced failed installation percentage distribution among different API level
IIR
1
0.8
0.6
0.4
0.2
0
10
5
0
-5
App lapse
-10
0
10
20
30
API lapse

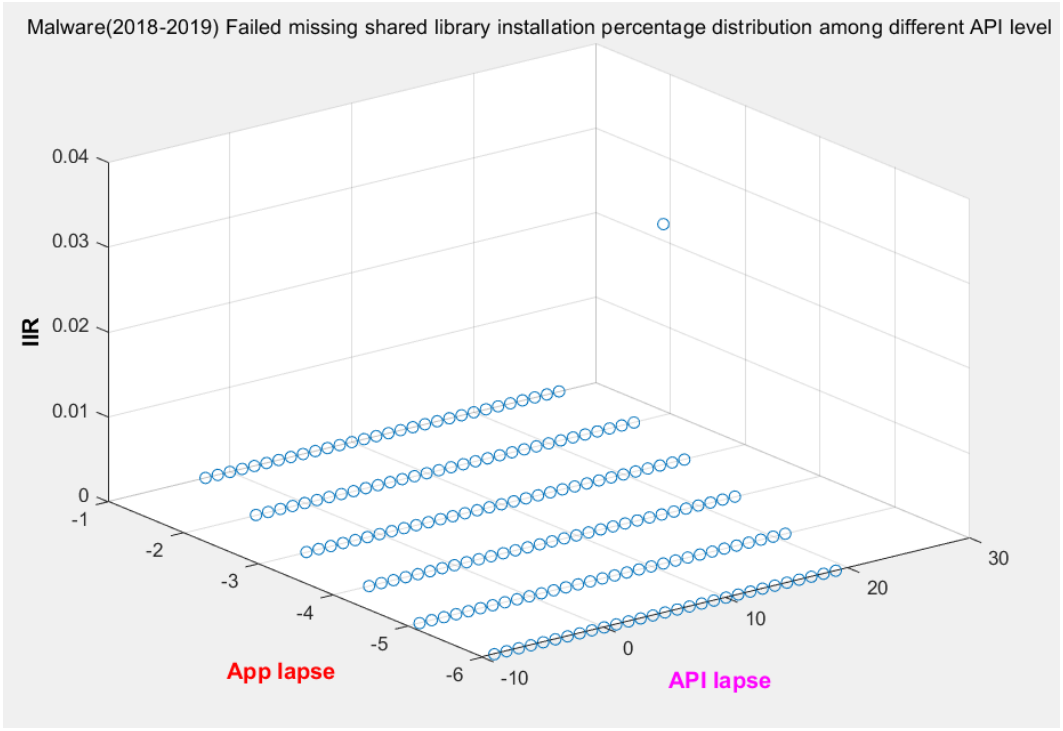

Benign failed execution percentage distribution among different API level
RIR
1
0.8
0.6
0.4
0.2
0
10
5
0
-5
App lapse
-10
0
10
20
30
API lapse

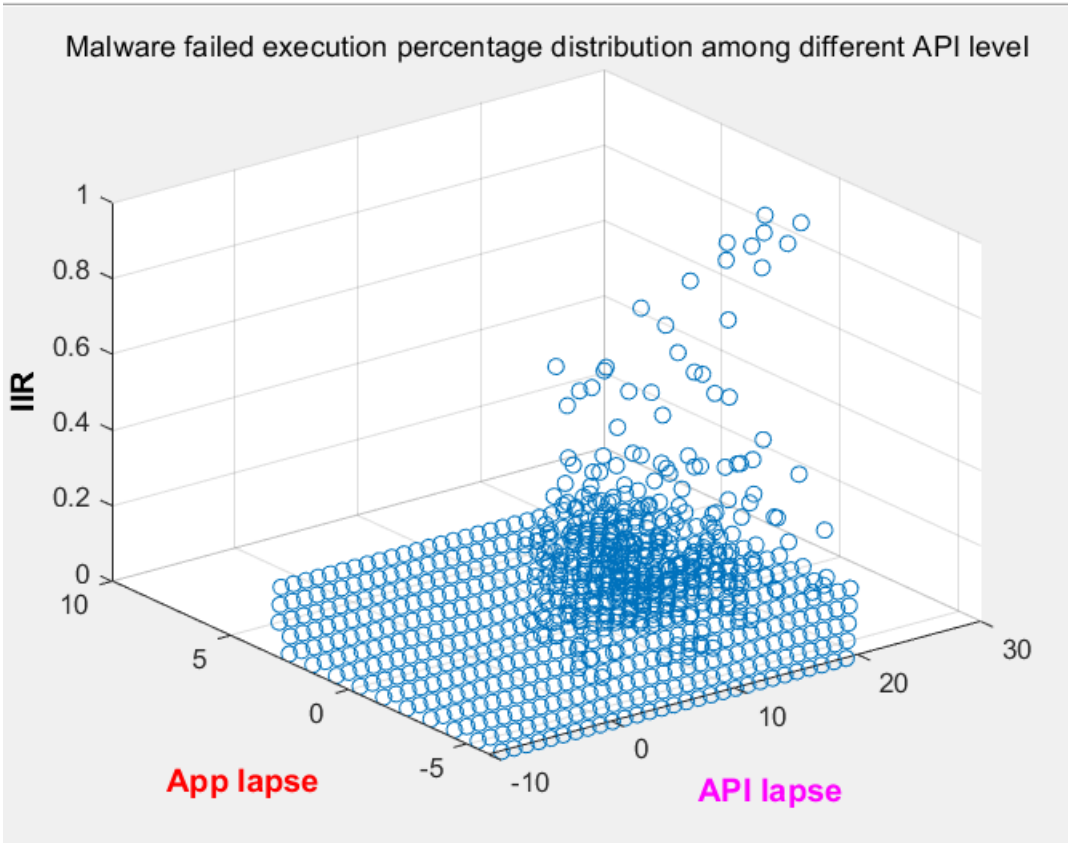

Benign native error induced failed execution percentage distribution among different API level
RIR
1
0.8
0.6
0.4
0.2
0
4
2
0
-2
-4
-6
App lapse
-10
0
10
20
30
API lapse

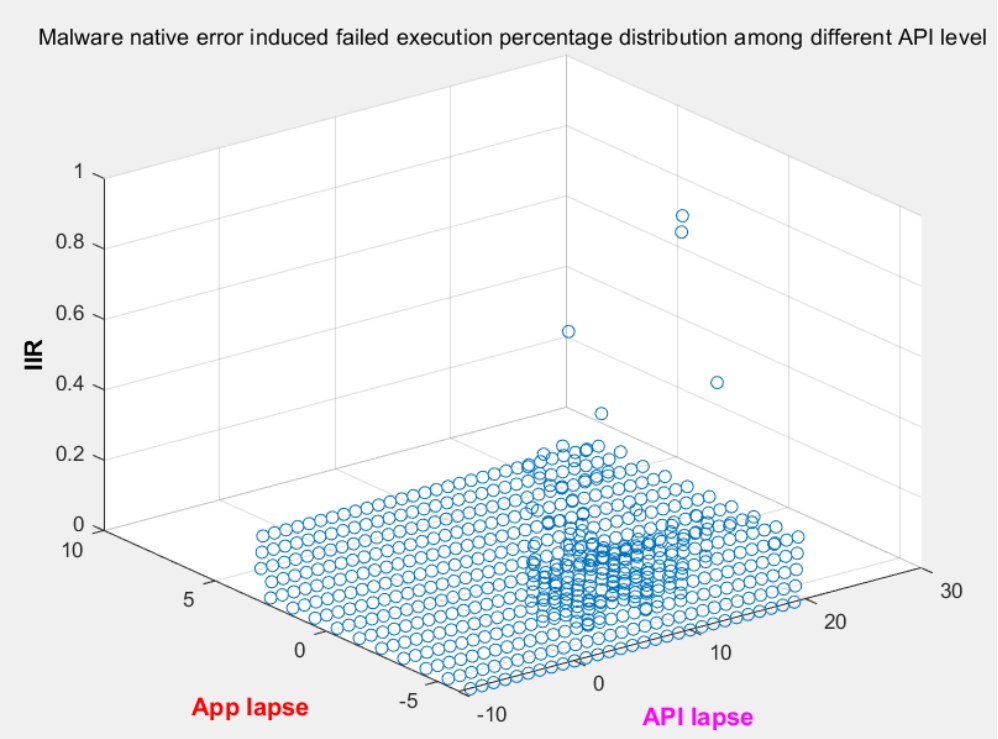

Benign verify error induced failed execution percentage distribution among different API level
RIR
0.8
0.6
0.4
0.2
0
4
2
0
-2
-4
-6
App lapse
-10
0
10
20
30
API lapse

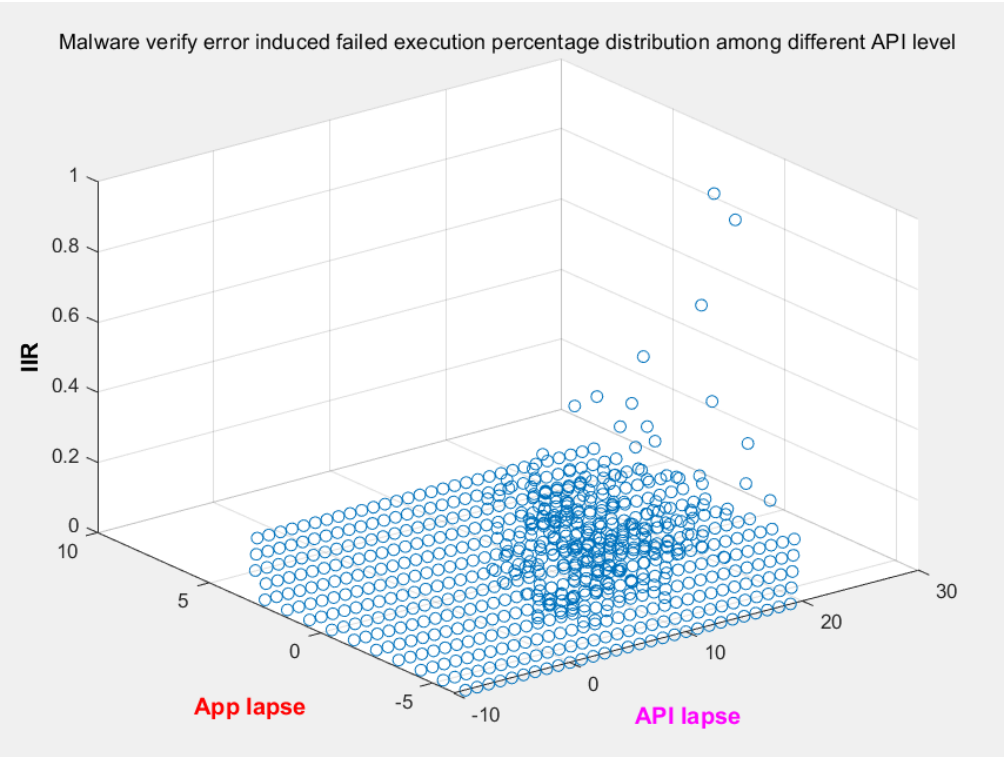

Malware(2018-2019) Non-matching ABI Failed installation percentage distribution among different API level
IIR
1
0.8
0.6
0.4
0.2
0
-1
-2
-3
-4
-5
-6
App lapse
-10
-5
0
5
10
15
20
25
30
API lapse


Malware(2018-2019) Failed installation percentage distribution among different API level
IIR
1
0.8
0.6
0.4
0.2
0
-1
-2
-3
-4
-5
-6
App lapse
-10
0
10
20
30
API lapse


Malware(2018-2019) Failed missing shared library installation percentage distribution among different API level
IIR
0.04
0.03
0.02
0.01
0
-1
-2
-3
-4
-5
-6
App lapse
-10
0
10
20
30
API lapse


Malware failed execution percentage distribution among different API level
IIR
1
0.8
0.6
0.4
0.2
0
10
5
0
-5
App lapse
-10
0
10
20
30
API lapse



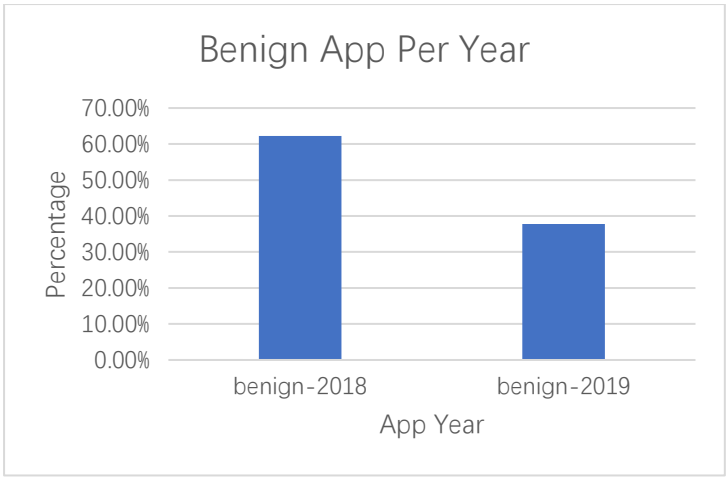


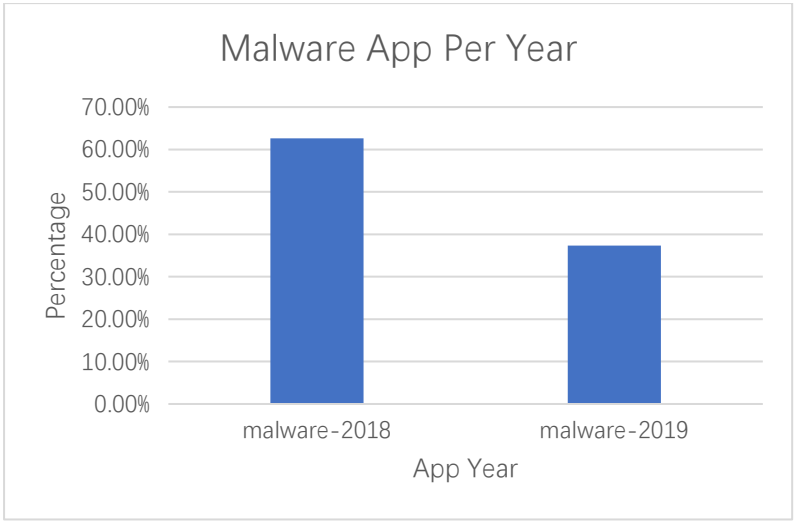
N. MultiCompat/MulClassify.py

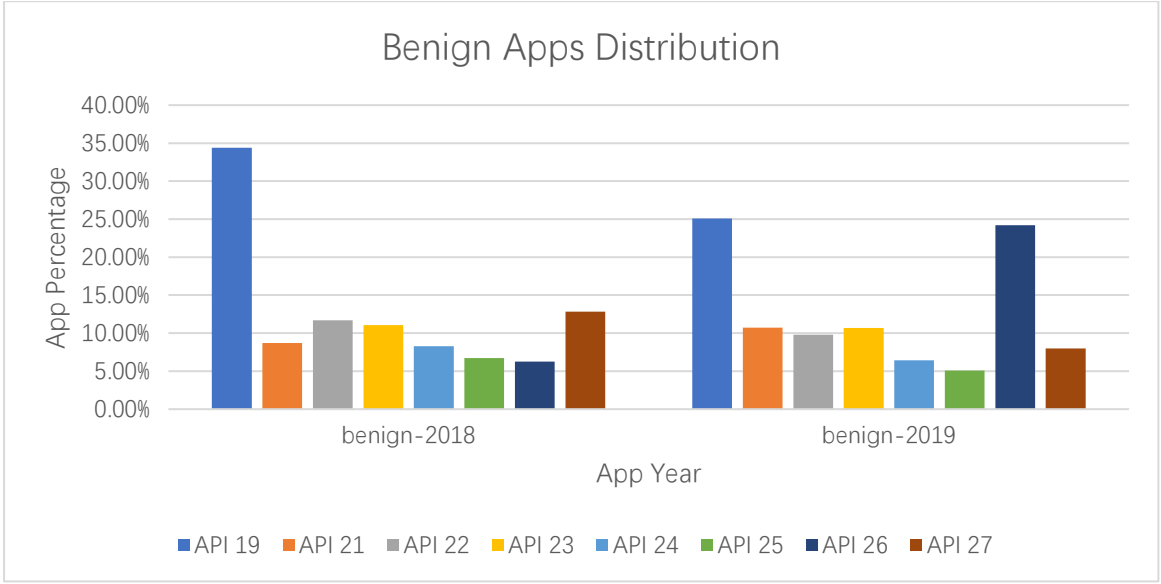
- Calculated the data of RQ1, RQ2, RQ3 of multi-compat.

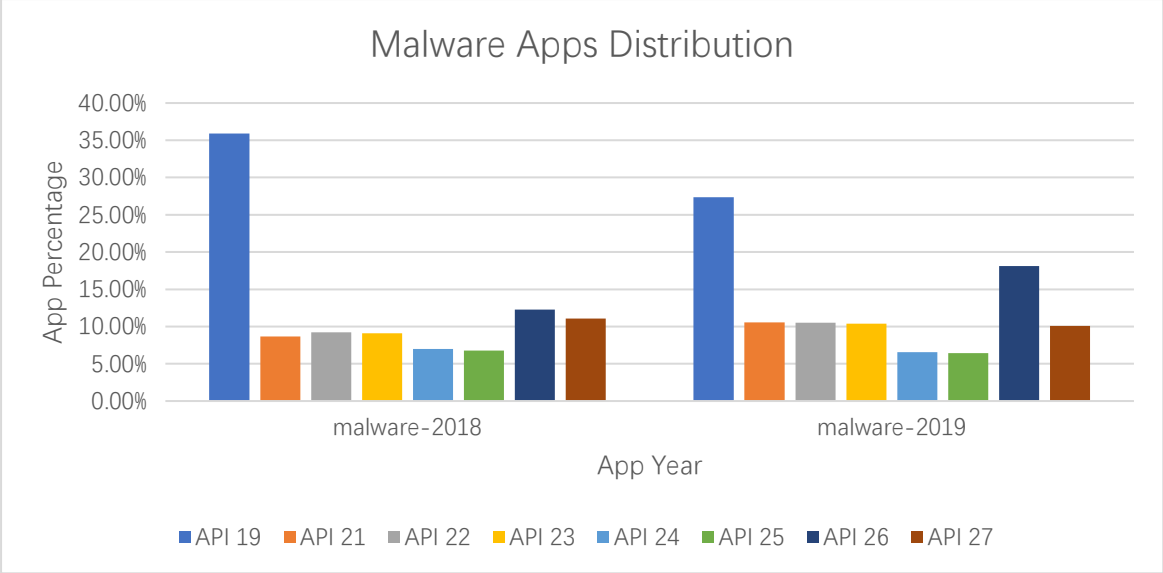
Data/MultiClassifyData.xlsx

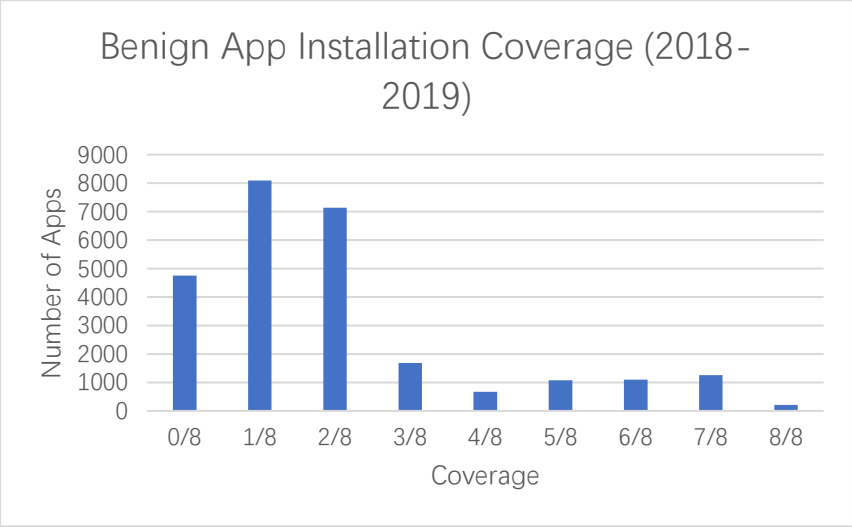
- Manually collect data from console to excel and plot the figures.

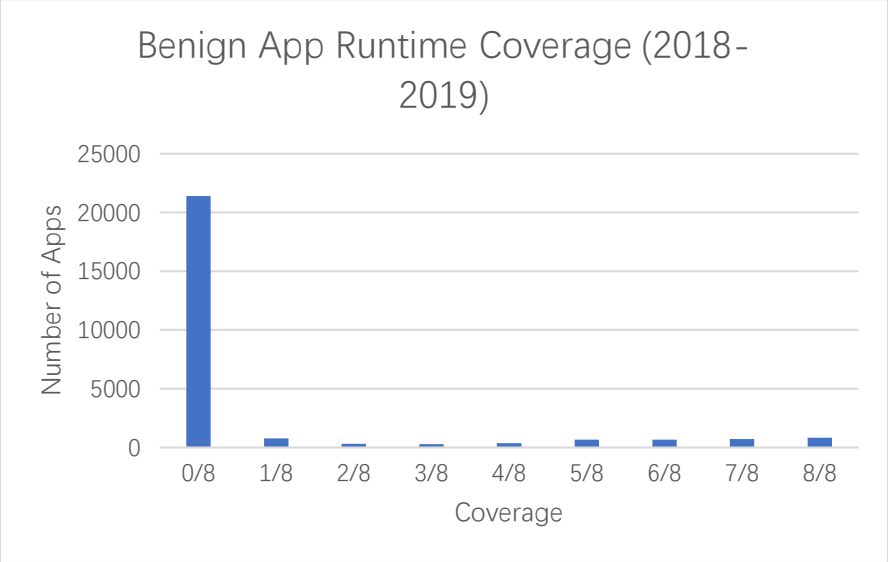

Benign App Per Year
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Percentage
benign-2018
benign-2019
App Year

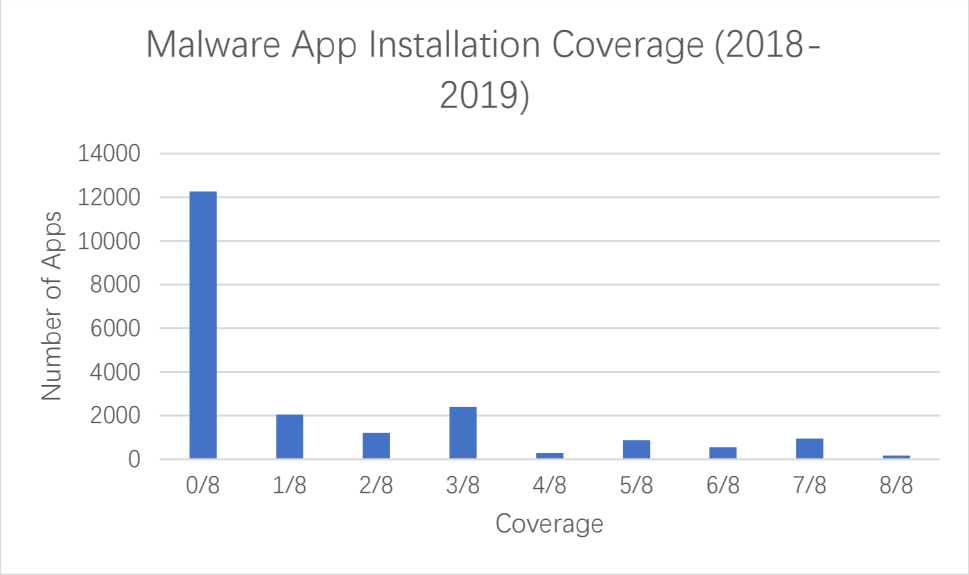

Malware App Per Year
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Percentage
malware-2018
malware-2019
App Year

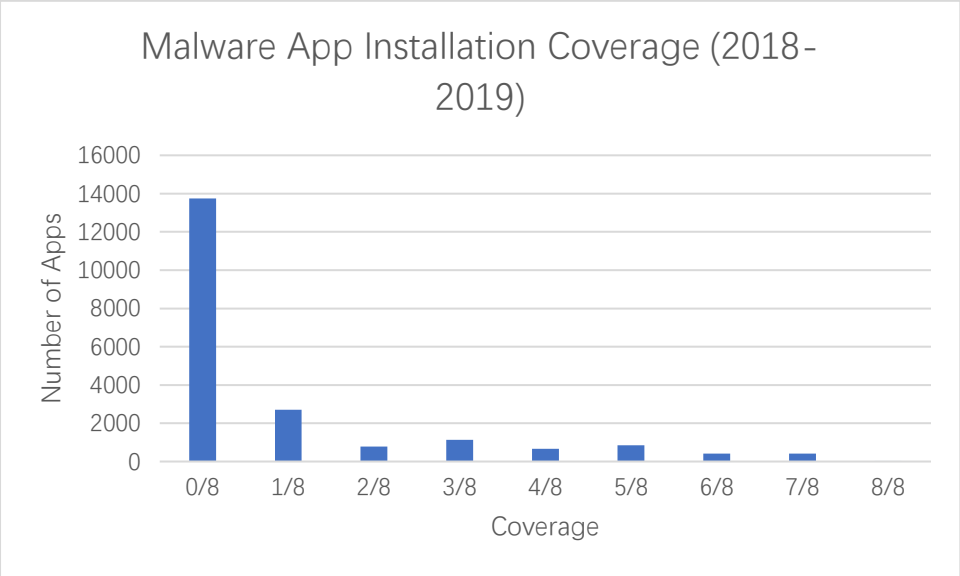

Benign Apps Distribution
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
App Percentage
benign-2018
benign-2019
App Year
API 19
API 21
API 22
API 23
API 24
API 25
API 26
API 27


Malware Apps Distribution
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
App Percentage
malware-2018
malware-2019
App Year
API 19
API 21
API 22
API 23
API 24
API 25
API 26
API 27


Benign App Installation Coverage (2018-2019)
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Number of Apps
0/8
1/8
2/8
3/8
4/8
5/8
6/8
7/8
8/8
Coverage


Benign App Runtime Coverage (2018-2019)
25000
20000
15000
10000
5000
0
Number of Apps
0/8
1/8
2/8
3/8
4/8
5/8
6/8
7/8
8/8
Coverage


Malware App Installation Coverage (2018-2019)
Number of Apps
14000
12000
10000
8000
6000
4000
2000
0
0/8
1/8
2/8
3/8
4/8
5/8
6/8
7/8
8/8
Coverage


Malware App Installation Coverage (2018-2019)
Number of Apps
16000
14000
12000
10000
8000
6000
4000
2000
0
0/8
1/8
2/8
3/8
4/8
5/8
6/8
7/8
8/8
Coverage