Eleven Paths

tacyt

Opcodes in Google Play

Tracing malicious Applications

Speakers

Alfonso Muñoz, PhD (@mindcrypt)
Senior Cybersecurity Researcher – alfonso.munoz@11paths.com

Sergio de los Santos (@ssantosv)
Head of Lab in ElevenPaths - ssantos@11paths

A Telefónica COMPANY

Eleven Paths

A Telefónica COMPANY

01

Android malware is coming!

True but… what is malware, anyway?

Report: 1 i… Android M…

More than one third of all apps were what Symatec calls "grayware" or…

Google guru: Android doesn't have malware,

TAG Malware , Operating System , Richard Stallman , iOS , Android , Windows

Malware, All Malware: How Free Software Advocate Richard Stallman Sees Windows, Android And iOS

By **Sumit Passary**, Tech Times | May 27, 7:39 AM

Richard Stallman, a free software activist, says that iOS, Android and Windows are malware. Stallman believes that

Richard Stallman, a computer programmer and free software activist, brands famous operating systems such as iOS, Windows and Android as malware.

In an opinion piece in The Guardian, Stallman suggests that nearly all operating systems, whether desktop operating system or mobile operating system, can be considered malware. Stallman argues that any software that is not distributed free of cost is malware.

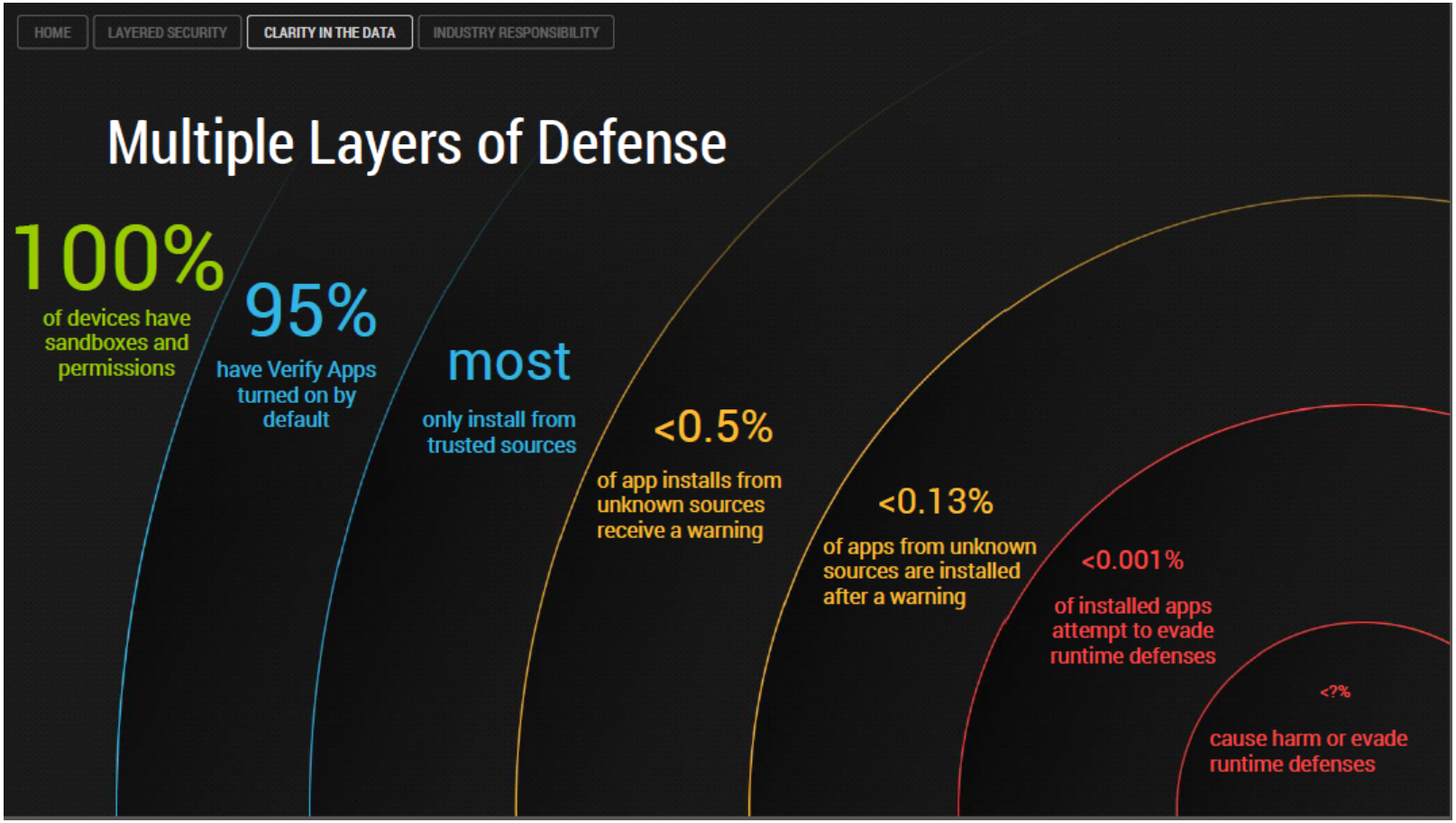
Stallman, who founded the Free Software Foundation, also made it clear in the opinion piece that he is not talking about any type of

RSA 2015 Malware doesn't exist on Android, Google says, but Potentially Harmful Applications™ do.

Analyzed | Classified as Malware | Classified as Grayware

7 roughly 700,000

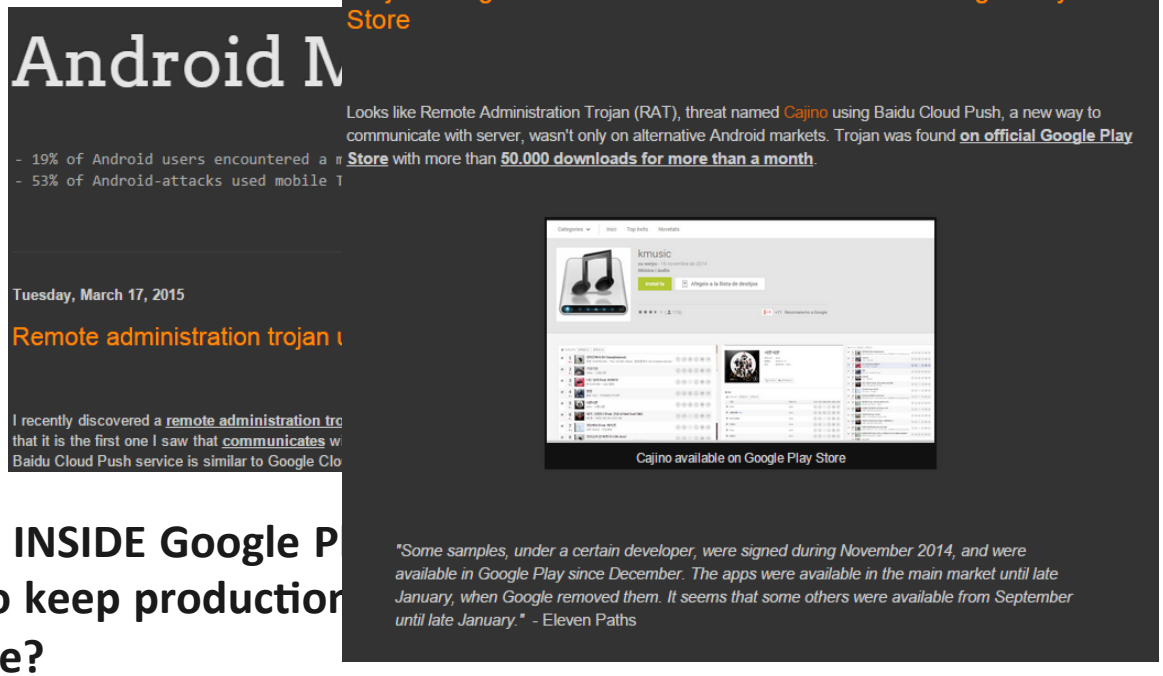
01 *Android malware is coming?*



- Android and Google Play, they do indeed have great security measures…
- … but what if… malware (ok, PHA) DO NOT NEED TO BREAK THEM?

01 *"Malicious code" in Android / Google Play*

- Malware/adware that sends SMS premium messages.
- Steals information
- Makes you part of a botnet
- Click fraud
- Insane ads
- RATs
- …



- **All found outside and INSIDE Google P**

rate. At least enough to keep productior

creating them otherwise?

01 *"Malicious code" in Android / Google Play*

- Antivirus is **not exactly** the technology we need for researching, discovering or analyzing new malware, frauds or threats.

- Researchs work with "intelligence"**.**

That is great, but…

- …to build this intelligence, they are very "malware set dependant" (relying on VirusTotal or ContagioDump…), so they will be good detecting this kind of malware, but not the new one.
- …regarding apps, traditionally, they have relied on **permissions, number of permissions, code, urls…. and just that.**
- Some of them work **with not so may apps** to train this intelligence.
- They feedback themselves with reputation, VT feedback, etc.
- And of course… adware.

01 What about adware?

Adware is a tricky matter.
Adware for Android may be very aggressive. Such aggressive that it could steal data, or flood you with popups since the telephone turns on.

Antiviruses

- Android developers use SDK to inject ads and have some revenues. They may configure this SDK so it is more or less aggressive.
- They usually mark this SDK as ADWARE. It does not matter how the developer have configured this ads to work into the app.
- They want to detect a lot, and get rid of ads for the user

Google Play

- Google Play IS OK WITH ADS, much more than Avs or even users. What may be wrong for an AV, is still ok with Google Play.
- Google Play wants to make money from this ads. But Google Play does not want to be “so OK that it bothers the user”…
- So, AFTER some complains, investigations, or any other criteria, it removes them.

So, basically, there is a grey area, and a conflict of interests.

01 *What to do?*

- *Android apps are APK, which are sets of Java files packaged in a ZIP structure signed with a self signed certificate. We have identified and dissected most of the technical characteristics.*

- *Most of Android apps are hosted in Google Play, with a developer, comments, descriptions, images, versions, categories…*

- *There is plenty of information in three stages:*

  Google Play -> The zip file itself -> cryptographic information

- ***An app is not just an app but an app and its circumstances. Focus on WHO and HOW, aside the WHAT***

- *Combining these three aspects we can get:*

  A lot “checkpoints” of data, that may identify and classify an app.

  From these data, we can deduce **the behavior of a developer.**

  We collect all these data, make a database, and “shake it”.

This has to be seen as a complementary method to detect malicious behaviour and it is not intended to replace any existing ones.

How to do it? Are there any previous researches about this?

- We need to find discriminant “features” to distinguish between goodware and “malware” apps
- **To be “really” fast, lets consider features from the apps and its environment, that do no require code revision or installing the apps (and execution)**
- Machine Learning using features (supervised learning) – SVM (Support Vector Machine)

Any difficulties?

Quantity:

We need a huge dataset of apps to test different researches

Quality:

Dataset Goodware and malware

Features selection

02

Previous researches

Hello?… Is there anyone out there?

Droid Permission Miner: Mining Prominent Permissions for Android Malware Analysis

Aswini A. M.
Department of Computer Science & Engineering
SCMS School of Engineering & Technology
Karukutty, Ernakulam, India
aswinimohan95@gmail.com

Vinod P.
Department of Computer Science & Engineering
SCMS School of Engineering & Technology
Karukutty, Ernakulam, India
pvinod21@gmail.com

Jacques Klein
Yves Le Traon

approach DREBIN draws the user's attention directly to rel-

Panda, Sophos). We flag all applications as malicious that

6

ıich
ors,
pli-
t as

A. Datasets

A total of 209 malicious samples are collected from the Contagiodump [2]. Likewise, benign applications were downloaded from various Internet sources (227 numbers).

	DREB
Full dataset	93.90
Malgenome	95.90

are detected by at least two of the scanners. This proce-

rate of 1%, corresponding to one false alarm when installing

02 Previous researches & results!

True but… what is malware, anyway?

These studies are all great. But, may we improve this in real life?

- They get a very good accuracy (up to 94%) and low False positives (down to 1%), detecting **specific malware**.
- They are very "malware set dependant" (relying on VirusTotal or ContagioDump…), so they will be good detecting this kind of malware.

- They have relied on **permissions, number of permissions and, in a way, code.**
- Some of them work **with not so many apps** to train.

02 But… What is malware?

Choosing malware set

VirusTotal is an excellent tool, but we think it needs to be “understood”… It is used for…

a) Comparing antivirus engines in a global or particular way. This is an awful idea. (Just read VirusTotal own help page…).

b) Cataloguing samples as malware. If it is “very detected” (*what does exactly that mean, anyway*) it is surely in the ”malware box”. Ok with it?

- **What is the right X? VT>1, VT>5, VT>10, VT>15?** Engines are growing in VT…
- If we use reputation of the engine as a factor… what is “reputation”? Vary famous?
- Do we really take into account that some Avs simply do not work with Android?
- Do we really penalize engines that detect “a lot”, so much that they may be false positives?

Setting as goodware samples with 0 detections: Depending of the freshness of the sample, samples are not detected by signatures.

02 What is malware?

Choosing the malware set

So far, for sure ...ive "quality" of engines... Did y

- VirusTotal se
- From an AV ...detecting such a flood of malw
- Later, when t ...le or analyze the sample.
- So, for some ... if researchers apply this crit ...g back itself.

Log in | Sign up

The Register®

Data Centre Software Networks Security Business Hardware Science Bootnotes Video Forums W

SECURITY

Kaspersky defends false detection experiment

Claws in copy cat dust-up

By John Leyden, 10 Feb 2010 Follow 3,100 followers

18

Kaspersky Lab has defended its handling of a controversial experiment criticised by some as a marketing exercise of questionable technical value.

The Russian anti-virus firm created 20 innocent executable files, adding fake malware detections for ten of the sample, before uploading the files to online online malware scanning service VirusTotal. VirusTotal routinely distributes samples of suspect files submitted to the service, which provides a useful tool for security pros to identify malware strains, to other vendors.

Ten days later, Kaspersky reports, 14 other vendors had added detection for the files it had deliberately (and falsely) labelled as malign.

Some representatives of security vendors, certainly those we spoke to at an event in Madrid last week, accused Kaspersky of attempting the suggest other vendors were copying its malware detections without applying anything like rigorous checks, in at least some cases. Kaspersky unveiled its findings to journalists attending its Moscow Labs rather than at a security conference or through peer review, further stoking negative sentiment about the experiment.

VirusTotal, which is run by Spanish penetration testing firm Hispasec Sistemas, accused Kaspersky of misusing its services. Kaspersky's actions particularly irked VirusTotal because in the past it was accused of running a service that was misused by virus writers in order to test if their creations would escape detection.

RELATED STORIES

Son of AV tycoon rescued following 'stupid' kidnapping

Misfiring Kaspersky update reduces servers to a crawl

Analysis Security firms plot revamp to minimise false alarms

Symantec slaps Trojan alert against Spotify

Updated Kaspersky update slaps Trojan warning on

Is there a better way?

We do not know in PE world… but, let's play in a smaller field… **Android world,** and **Google Play** town.

So, in this new field, some **considerations** should be taken into account.

So, what would be "**malware**" in Google Play?
How to build a good "malware set" so we can create a good classificator

Eleven Paths

A Telefónica COMPANY

03

Signature accuracy and Gregariousness

A proposal to get better malware sets

03 Signature accuracy

Google Play knows its bussiness, doesn't it?

- Signature accuracy may be seen as another factor of quality assigned to a detection.
  - It is based on **using Google Play as a Judge**, that, removing the app, validates in some way AVs detection. *"Hey, you were right, this app should not be here, (although it will take a while for me to remove it"*)
  - **"The more detected apps with a signature, that are eventually removed from Google Play, the more accurate the signature is"**
- If these detected apps are not removed from Google Play, it means Avs were too aggressive, (it was just "tolerated adware"). **This is why some of the studies say there is so much malware in Google Play.**
- That is easy: NumberOfRetiredWithASignature/DetectedWithASignature = accuracy of the Signature.

We could take "all the apps detected with high accuracy signatures"

This would give us a nice malware set.

03 Signature accuracy

Google Play knows its bussiness, doesn't it?

AVEng	AVEngine	Signature	nDead	detections	Accuracy	uracy
AegisLa	Ad-Aware	Android.Adware.Dowgin.D	1	1	1.0000	91
AntiVir	Ad-Aware	Android.Adware.Dowgin.X	1	1	1.0000	98
VIPRE	Ad-Aware	Android.Adware.Mulad.N	1	1	1.0000	56
ESET-N	Ad-Aware	Android.Riskware.SMSReg.AF	1	1	1.0000	99
VIPRE	Ad-Aware	Android.Riskware.SMSReg.CT	1	1	1.0000	54
ESET-N	Ad-Aware	Android.Riskware.SMSReg.W	1	1	1.0000	56
Comodo	Ad-Aware	Android.Trojan.Ewalls.A	1	1	1.0000	54
AVware	Ad-Aware	Android.Trojan.Mseg.A	1	1	1.0000	59
VIPRE	Ad-Aware	Trojan.Generic.KDV.690486	1	1	1.0000	41
AntiVir	Ad-Aware	Trojan.Script.603388	1	1	1.0000	17
VIPRE	AegisLab	BitCoinMiner	1	1	1.0000	37
Fortinet	AegisLab	Ganlet_1	1	1	1.0000	93
ESET-N	AegisLab	Kakay	1	1	1.0000	38
Jiangmi	AegisLab	SpyMob	1	1	1.0000	26
Sophos	Agnitum	Trojan.Agent!0TdSTBuha+8	1	1	1.0000	28
ESET-N	AhnLab-V3	Android-Adware/Gappusin	1	1	1.0000	84
AVware	AhnLab-V3	Android-Malicious/Eicar	1	1	1.0000	73
Sophos	AhnLab-V3	Android-Malicious/SmsForw	1	1	1.0000	82

sult 2 ×

03 Signature accuracy

Google Play knows its bussiness, doesn't it?

- Yes, "accuracy" has some "problems":
  - We have to improve it a bit, and introduce the concept of "Credibility" and "Participation", **so a "good" performance has been shown over time, detecting a minimum with success.** How many times a signature should be used so it shows a good performance over time?

Credibility Based on Participation X Accuracy

03 Signature accuracy /enhanced

Google P

AVEngine	Signature	accuracy	nDetections	enhancedAccuracy
MicroWorld-...	Android.Trojan.InfoStealer.DD	1.0000	5	0.6534395262100697
NANO-Antivi...	Trojan.Android.Funtasy.darrpp	1.0000	5	0.6534395262100697
NANO-Antivi...	Trojan.Android.Leadbolt.dfuxrl	1.0000	5	0.6534395262100697
Qihoo-360	Win32/Virus.DoS.44f	1.0000	5	0.6534395262100697
TrendMicro-...	TROJ_GEN.F47V0626	1.0000	5	0.6534395262100697
Ad-Aware	Android.Adware.GingerMast...	0.8000	5	0.5227516209680557
AhnLab-V3	Android-Spyware/Flexion	0.8000	5	0.5227516209680557
AVG	Android/Deng.ES	0.8000	5	0.5227516209680557
BitDefender	Android.Adware.GingerMast...	0.8000	5	0.5227516209680557
Emsisoft	Android.Adware.GingerMast...	0.8000	5	0.5227516209680557
ESET-NOD32	a variant of Android/Adware....	0.8000	5	0.5227516209680557
F-Prot	<AndroidOS/Airpush.C	0.8000	5	0.5227516209680557
F-Secure	Android.Adware.GingerMast...	0.8000	5	0.5227516209680557
F-Secure	Android.Adware.Minimob.A	0.8000	5	0.5227516209680557
F-Secure	Android.Adware.Minimob.B	0.8000	5	0.5227516209680557
GData	Android.Adware.GingerMast...	0.8000	5	0.5227516209680557
McAfee	Artemis!40A755127451	0.8000	5	0.5227516209680557
MicroWorld-...	Android.Adware.GingerMast...	0.8000	5	0.5227516209680557

sult 1

03 Gregariousness

Some Avs are more “reactive” than others

- What if the Avs detect a lot just when others do? (gregarious)
- What if Avs detect a lot of false positives? (eccentric)

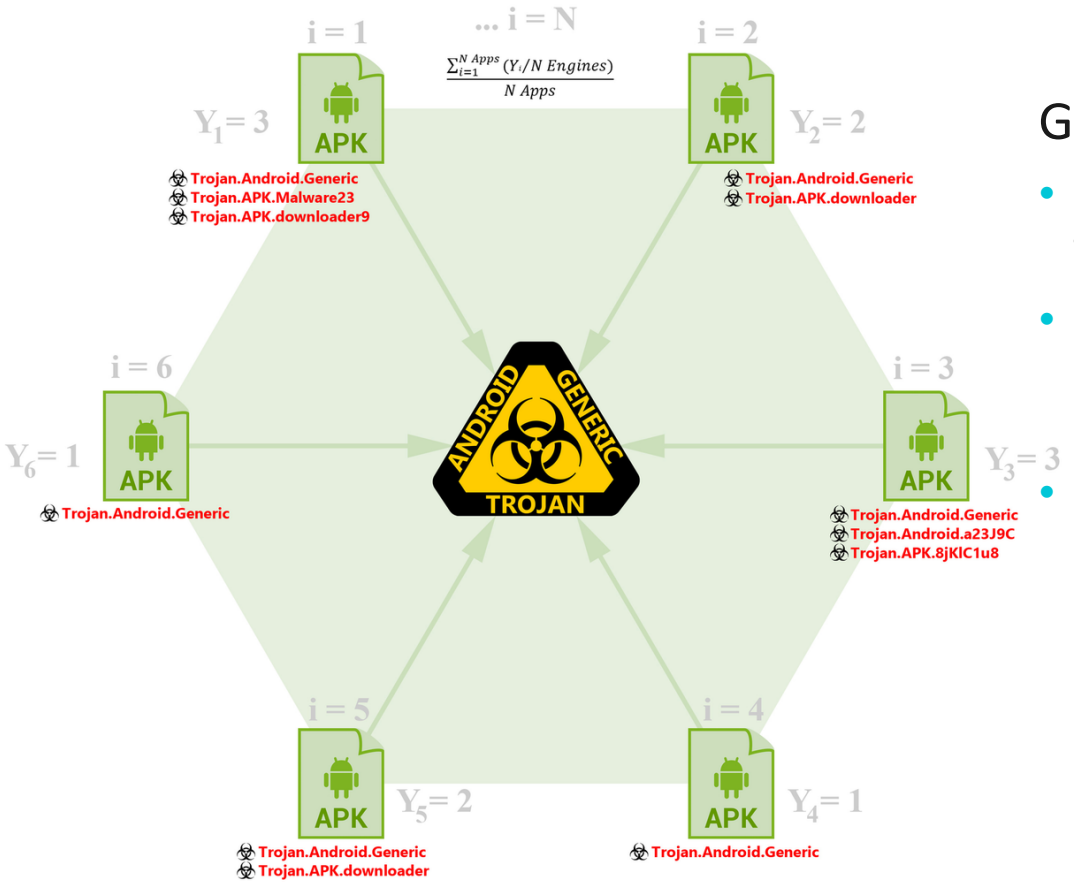
We may improve this a little bit, trying to determine de “nature” of the signature. For the future, we need to penalize signatures that work a lot in “single” detections (eccentric) or just when any others detect (gregarious).

- So the optimum point is somewhere in the middle.

03 Gregariousness

Some Avs are more “reactive” than others

What is the average number of engines detecting the same apps (in average) that this signature detects?



Given a signature:

- How many apps are detected by this signature?
- For this apps, how many others engines detect them (in average)?
- **Basically: It is the average of the coefficient of detection**

03 Gregariousness

Some Avs are more “reactive” than others

- If
- If
- **B**

AVEngine	Signature	nDetections	Gregariousness	Balanced
Avira	Eicar-Test-Signature	2	0.875	0.362195
AVware	EICAR (v)	2	0.875	0.362195
Cyren	EICAR_Test_File	2	0.875	0.362195
Sophos	Andr/Eicar-A	2	0.875	0.362195
Antiy-AVL	Test[:not-a-virus]/Win32.EICAR	3	0.866667	0.385769
Ikarus	EICAR-ANTIVIRUS-TESTFILE	2	0.866667	0.385769
Ikarus	Testfile.AndroidOS.EICAR	1	0.866667	0.385769
Jiangmin	EICAR-Test-File	3	0.866667	0.385769
McAfee-GW-Edition	EICAR test file	3	0.866667	0.385769
Zillya	EICAR.TestFile	3	0.866667	0.385769
Zoner	EICAR.Test.File-NoVirus	3	0.866667	0.385769
Ad-Aware	EICAR-Test-File (not a virus)	4	0.85	0.432373
Agnitum	EICAR_test_file	4	0.85	0.432373
AhnLab-V3	Android-Malicious/Eicar	1	0.85	0.432373
AhnLab-V3	Android-Test/Eicar	3	0.85	0.432373
Avast	EICAR Test-NOT virus!!!	4	0.85	0.432373
AVG	EICAR_Test	4	0.85	0.432373
Baidu-International	EICAR.Test.File	4	0.85	0.432373
BitDefender	EICAR-Test-File (not a virus)	4	0.85	0.432373
CAT-QuickHeal	EICAR.TestFile	4	0.85	0.432373
ClamAV	Eicar-Test-Signature	4	0.85	0.432373

03 Gregariousness / Balanced

Some Avs are more “reactive” than others

Sam

AVEngine	Signature	Balanced	nDetections	enhancedBalanced
Baidu-International	Trojan.Android.Plankton.I	0.906888	107	0.8764747302696492
Rising	DEX:PUF.Plankton!1.9DAE	0.893638	182	0.8757720740151775
F-Prot	AndroidOS/Apperhand.A	0.872509	230	0.8586480437113408
Fortinet	Android/Apperhand.AA!tr	0.868443	303	0.8579308874490152
NANO-Antivirus	Trojan.Dex.Startapp.ctgcbx	0.886432	110	0.8574892964166416
AntiVir	Android/Plankton.C.Gen	0.86681	336	0.8573365193301186
AVG	Android/Plankton	0.862285	216	0.8477139621037821
Ikarus	AndroidOS.AdWare.Plankton	0.908439	46	0.8405925646328143
Bkav	Android.Adware.ApperhandAds	0.888487	65	0.8404792425384358
F-Secure	Application:Android/Counterclank	0.850839	228	0.8372056125400107
DrWeb	Adware.Startapp.6.origin	0.857509	149	0.8366607529987036
Sophos	Android Apperhand	0.844874	321	0.8352138454825138
Commtouch	AndroidOS/Plankton.A.gen!Eldorado	0.847963	202	0.832658662261494
Ikarus	AndroidOS.AdWare.Apperhand	0.832447	179	0.8155312068458045
ESET-NOD32	a variant of Android/Plankton.I	0.826355	268	0.8150636092454455
F-Prot	AndroidOS/Plankton.D	0.836413	85	0.8014071129206428
Tencent	Adware.Android.Apperhand.a	0.872193	41	0.7997691393094328
Ikarus	AdWare.AndroidOS.Apperhand	0.824106	98	0.7940238691319053

03 All together now

Wh
you

AVEngine	nDetections	enhancedAccuracy	enhancedBalanced	enhancedAxenhancedB
ESET-NOD32	35404	0.400314294657568	0.3419182348281158	0.3762688582572053
AntiVir	28916	0.3857174313986034	0.35360439167922664	0.37249441504356595
Fortinet	11309	0.3076218211291671	0.45692378455436966	0.3690991001866034
Ikarus	11605	0.34531932780427554	0.3966825931148037	0.3664689076380225
VIPRE	34160	0.35927998509018105	0.32120627979087146	0.34360257702575947
AegisLab	19081	0.21864251575511276	0.5188919016693916	0.3422746158374629
TrendMicro-HouseCall	13272	0.3429611180911511	0.34099196900334583	0.3421502919961724
DrWeb	6334	0.3073604111923156	0.3901247352265287	0.3414398387358151
Sophos	7423	0.289326261661417	0.3868687818049352	0.3294908287793363
AVware	8767	0.23774311177693427	0.43153358820021026	0.3175391903041656
Comodo	6424	0.2879249629918774	0.3424215248144691	0.31036472374235635
F-Prot	9266	0.2645551939252682	0.37216099621172916	0.30886346545498744
McAfee	6361	0.27198170342155936	0.3584585881352673	0.30758983242132143
McAfee-GW-Edition	5371	0.2609017205111164	0.3433606928224435	0.2948554149922511
AVG	8876	0.2538577667579801	0.3428874138427105	0.2905170332046338
NANO-Antivirus	9384	0.2705734506205292	0.29343814688720826	0.27998832555386766
Baidu-International	5203	0.2571081797779555	0.1967620955397713	0.23225979215046794

Eleven Paths

A Telefónica COMPANY

04

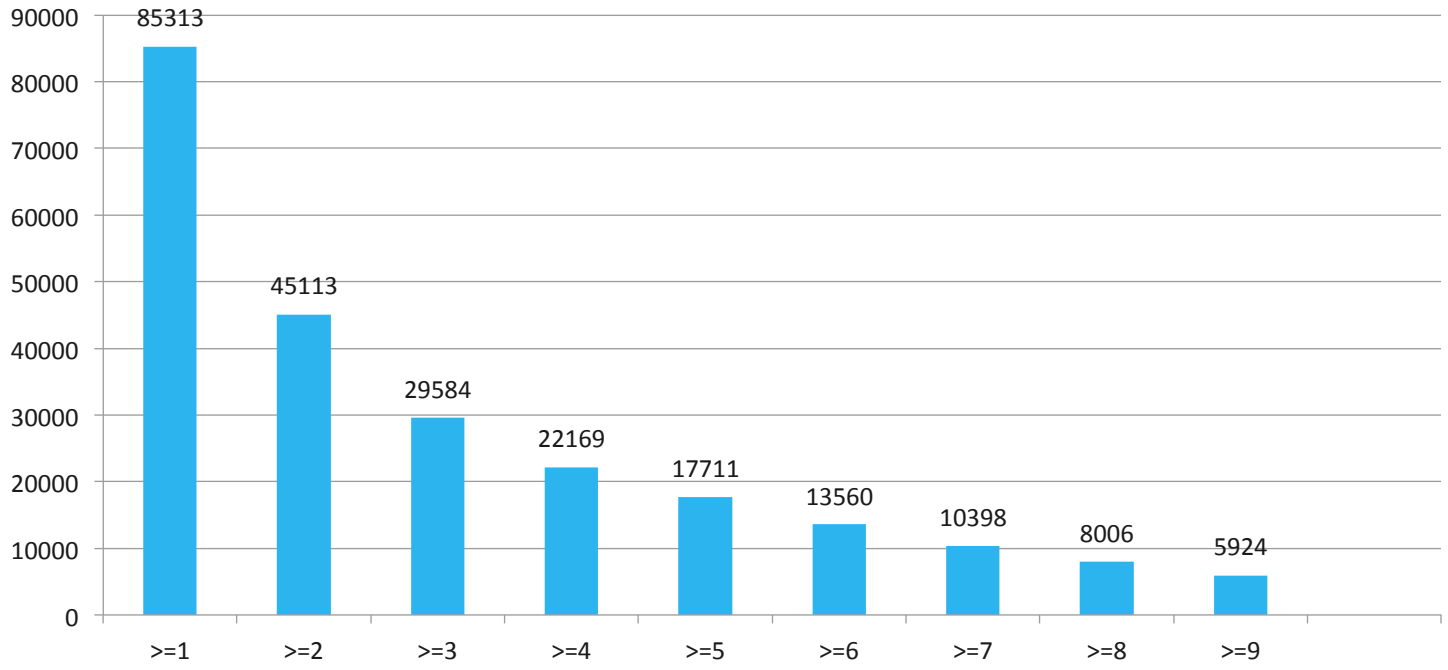
Datasets Goodware and Malware

Because… size matters!

04 Size matters…

We have a mega database of Android Apps with its market data associated and metadata of the APK itself… and the results in VirusTotal of the apps. It is used in intelligence and researching…

611,323 of our apps were found in VirusTotal, from a total of 742,344 that we got (about September 2014), extrictly from Google Play. Nowadays we have about 3M



04 All together now

What we have done so far. Any malware set may be ok depending on what you need

- We have created a criterion to improve the generation of malware sets based on information retrieved from Google Play.
- It is very customizable.
- This may improve researches about “detecting malware/adware” with machine learning techniques and Android.

For us, goodware dataset is:

- No detections.
- 4.3 stars or even better.
- 70,000 or more downloads.
- More than a month in Google Play.

Eleven Paths

A Telefónica COMPANY

05

Machine Learning (SVM)

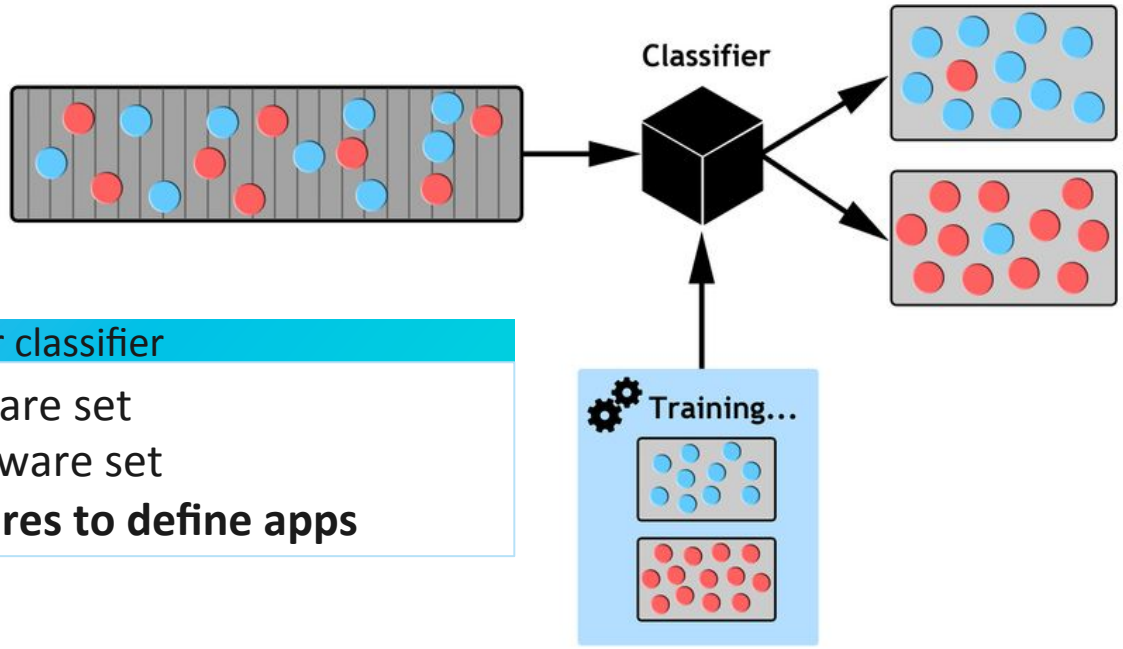
Detecting features and classifying malware

(Research in progress)

05 Target: Realistic and improved classifier

https://aws.amazon.com/es/machine-learning/details/

- Machine learning (ML) is a “commodity”. There are so much libraries (SPARK/ MLLIB, LIBSVM,…) ant it is easy to analyse data using ML.



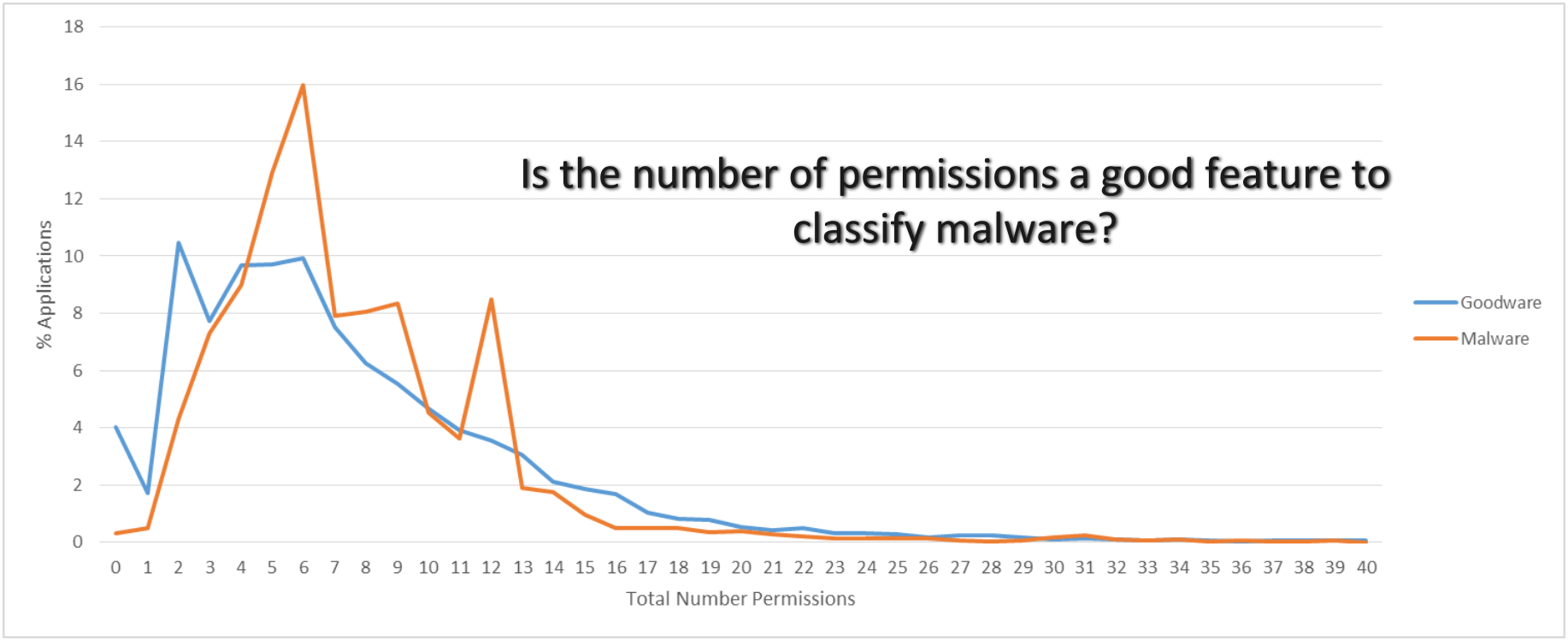
Recipe for classifier

- Malware set
- Goodware set
- **Features to define apps**

- We do not need only goodware/malware sets, but “features” of the apps to build different machine learning classifiers. Which are the bests features?

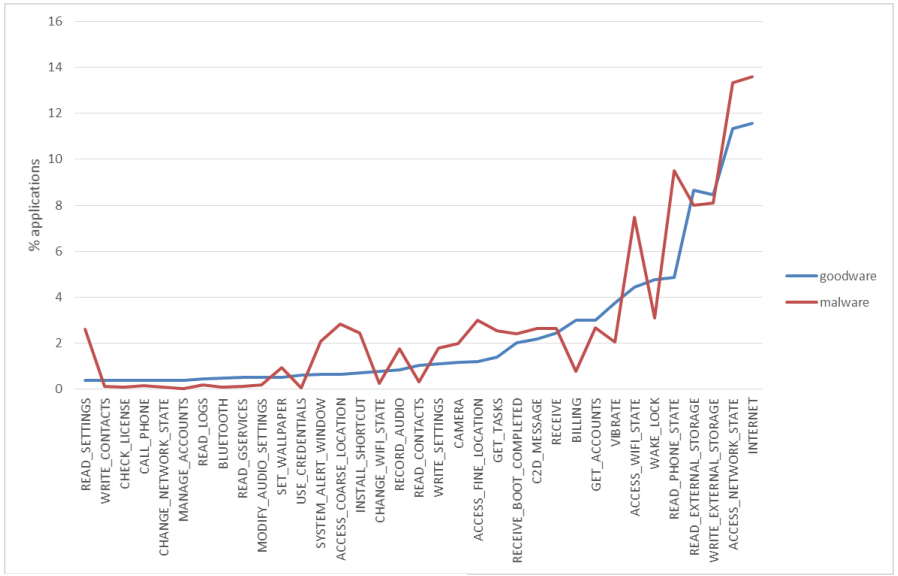
05 Previous researches…

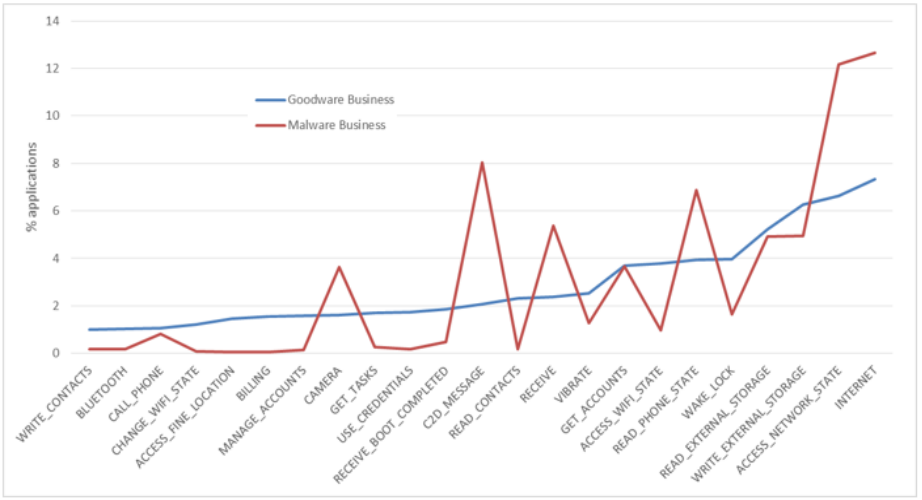
- With our huge database (around 3 million apps), we validate previous studies that use the number of permissions (apps) as a feature (goodware/malware).

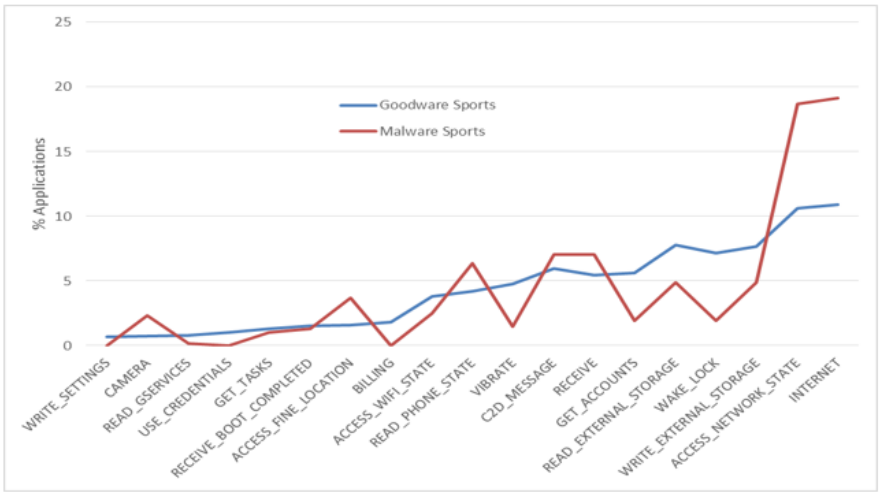


- Malware does not have more permissions than other apps (in general)…
- Malware is (only) more likely in a range from FOUR to THIRTEEN permissions.

- Specific combinations of permissions or specific permissions by category (apps) are useful to detect malware?



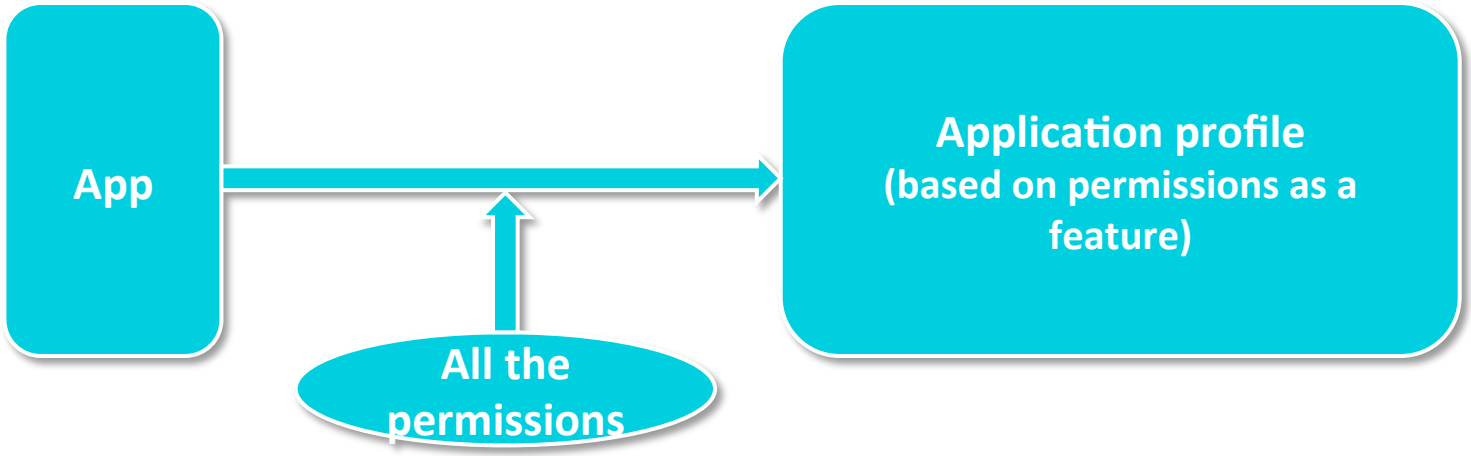




05 Features based on permissions…

Let's try to build a classifier using permissions as the only feature

- Well, the only way to know it is testing… what will happen?
- First: let's try to propose an app definition only by its permissions.



- **It does not seem a very accurate way of defining an app, but this is an experiment to see if it is enough for classifying only by this parameter**
- Second: we make a supervised learning process from the apps in both sets (goodware/malwae), just taking the permissions into account.

05 Features based on permissions…

Let's try to build a classifier only using permissions as a feature

- We have trained a machine learning system with supervised learning implementing Support Vector Machine (SVM) algorithms.
- **With our huge dataset (apps) from Google Play and their criteria (previous studies) for choosing a malware set, we were not able to guarantee an acceptable accuracy to distinguish and classify adware/malware and goodware relying only on the permissions.**
- Results:

Sorry… but as far as we know previous researches don't work with a huge dataset

- **What do we do now?**

05 Features based on permissions by category…

Come on! Let's try…

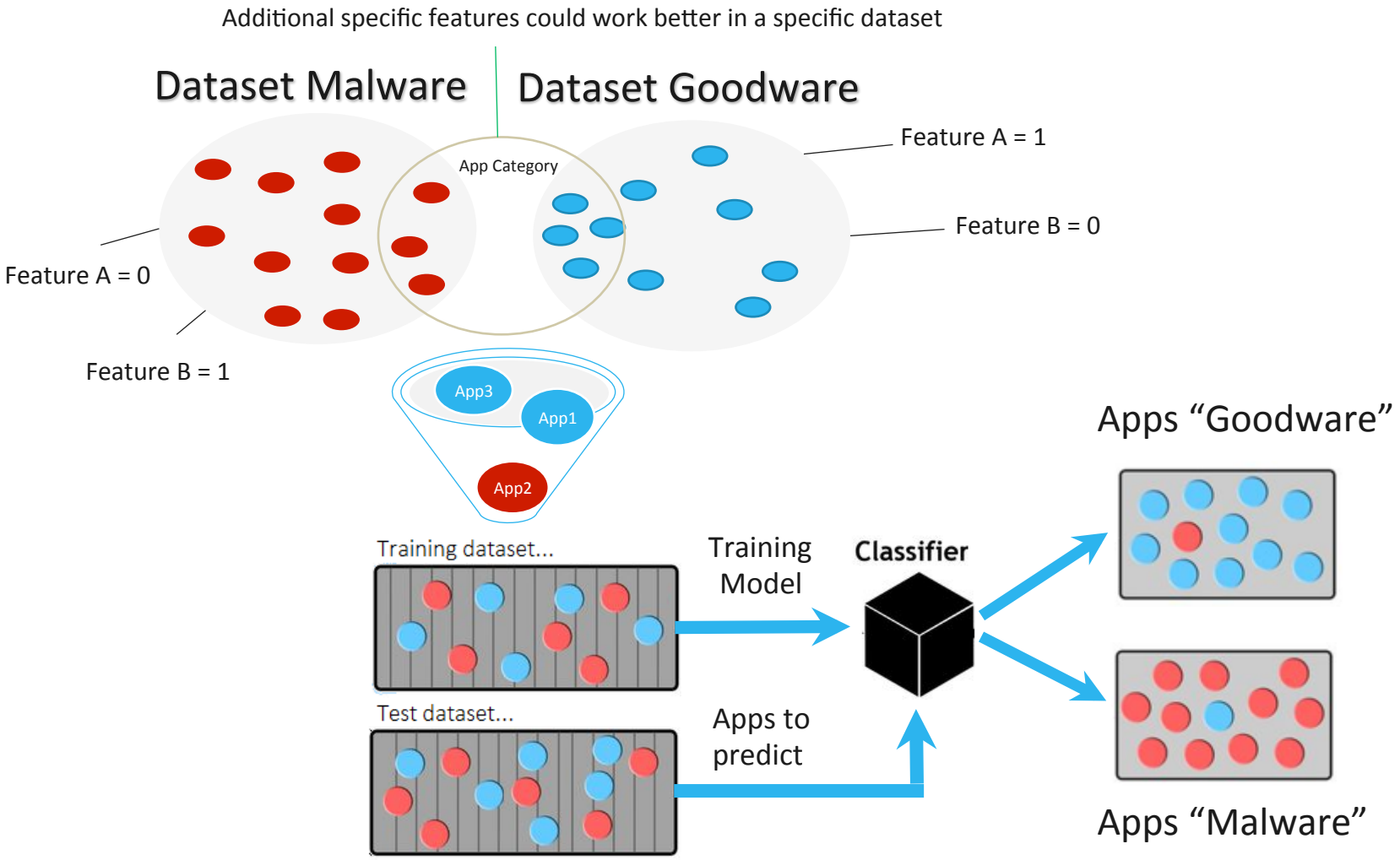
- **With subsets by category (apps) the classifier better detects goodware/malware than when it uses features based on permissions with general subsets** (in this last case we are not able to distinguish between goodware and malware)

Without Category (general dataset): Accuracy = 50%

Category	Accuracy %	Category	Accuracy %	Category	Accuracy %	Category	Accuracy %
Book and reference	77.74	Game arcade	72.57	Lifestyle	74.77	Shopping	86.11
Business	76.06	Game educational	66.88	Game action	66.66	Social	73.66
Communicaton	78.33	Game family	71.33	Game sports	68	Sports	77.86
Education	71.22	Game puzzle	84	Photography	76	Tools	66.33
Entertainment	66.66	Game Racing	74.92	Personalizatio n	80.26	Productivity	77.40

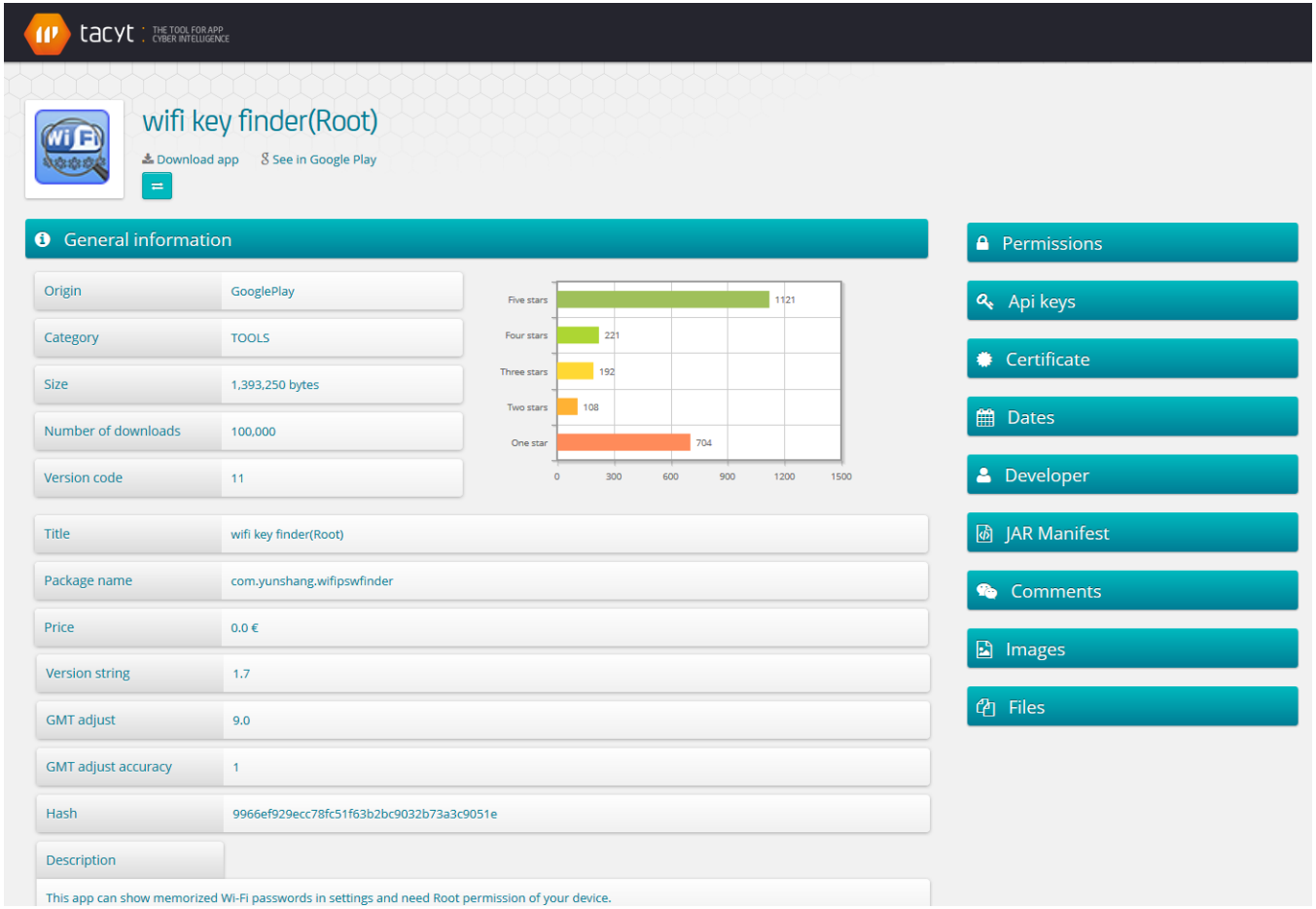
- Although, the results are bad, this study gives us a good idea to move forward: **What will happen if we use more features and specific categories (granularity)?**

05 What will happen if we use more features and specific categories (granularity)?



What features we use? “Meta-info features”

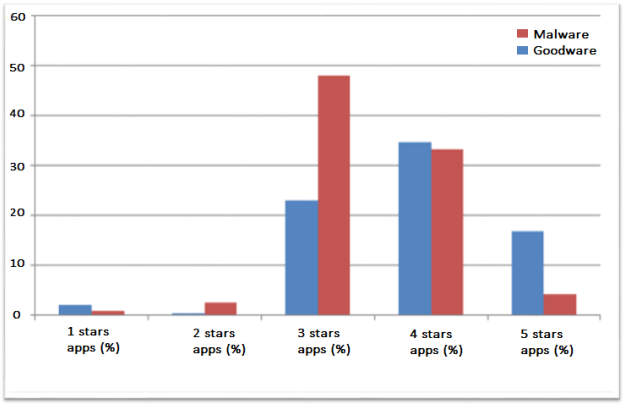
Let’s try to improve the classifier with some more features

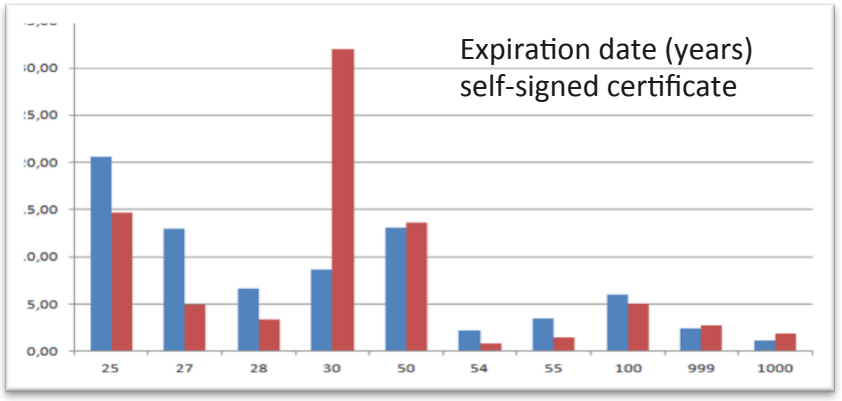


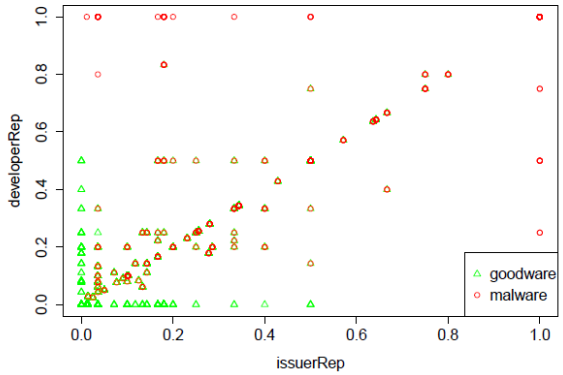
Do these features are useful to classify goodware and malware?

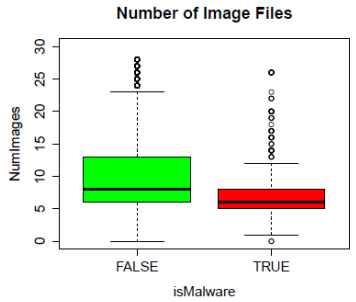
05 How to select features?... We will be honest! :D

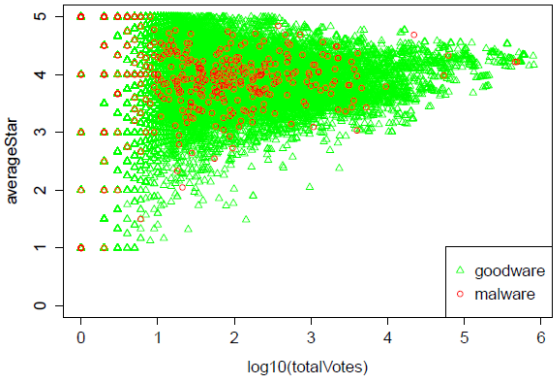
- Basic statistics and “intuitions”… and, recently, features selection algorithms

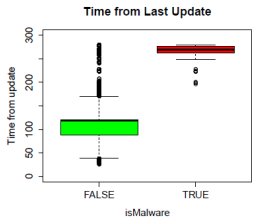


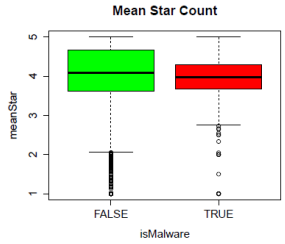












05 Classifying with more features (meta-info)…

Example dataset goodware/malware

- Trend: More features and specific categories (granularity)
- Goodware (True Positive), Malware (True Negative), SVM (Support Vector Machine)
- **10 features:** number permissions, size, certificate, info description, …
- Results are much better than with features based on permissions only…

- Train: 16,470 (apps goodware) + 8,236 (apps malware)
- Test: 8,236 (apps goodware) + 4,118 (apps malware)
- In total ("global" classification): 37,060 apps = 24,706 goodware + 12,354 malware

- **Accuracy = 91.015555% (11,244 apps ok predicted / 12,354 apps to predict)**

- Recall = True positive rate = 90.2507% (measure the proportion of GOODWARE which is correctly identified as such) TP/TP+FN
  True negative rate = 92.9468% (measure the proportion of MALWARE which is correctly identified as such) TN/TN+FP

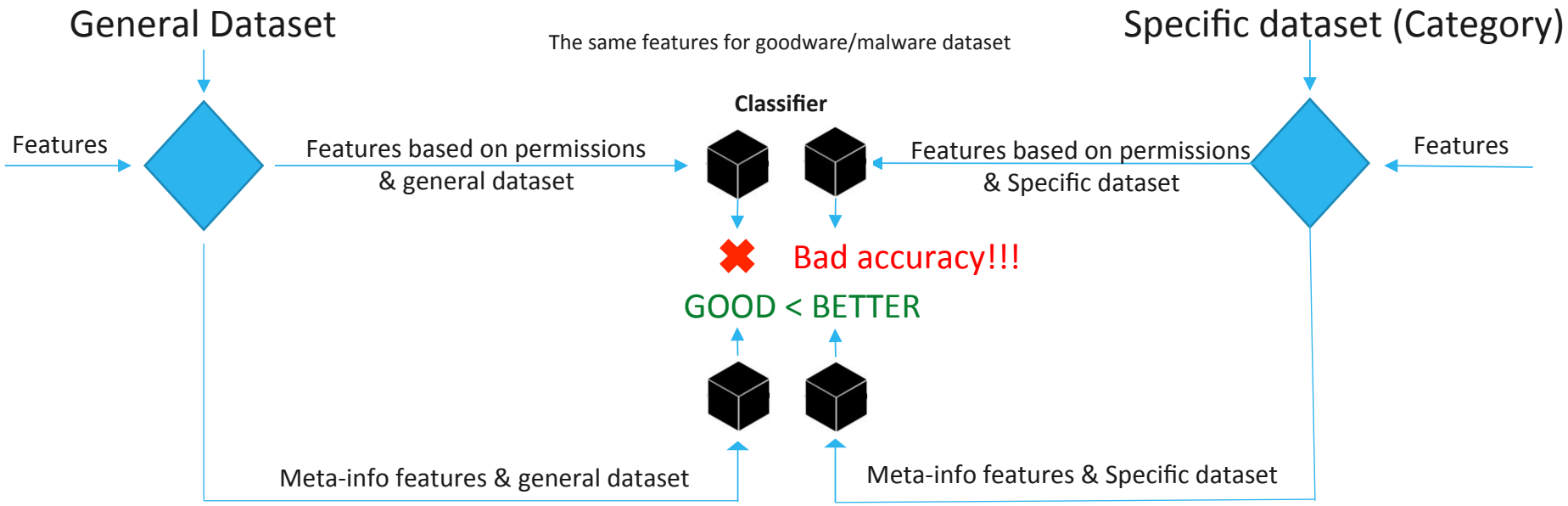
- **Precision = Positive Predictive value = 97%** (how good is predicting goodware) TP/TP+FP
  Negative predictive value = 79.0432% (how good is predicting malware) TN/TN+FN

- F1 score = f(precision, recall) = 93.5042%

- For us it is a good result as a complementary method to rank apps, especially, when you try detecting malware in a huge dataset, as the real world, and you do not know how to prioritize apps.

05 What will happen if we use more features and specific categories (granularity)?

- For some categories the results are better than the “global” classification (previous). Some examples will be showed later!
- For some categories the results are worse than the “global” classification
  - There aren’t enough samples to train/test… is a problem! How to solve it?
  - A good balance between general & specific dataset it is necessary to get an accuracy classifier.



05 How to improve the classifier...

- We are searching for new features…

- **We will be honest, the classifier is good but… I would like to improve the results analysing code… sometimes is the only way to detect malicious code with high accuracy.**

- But… analysing code is a slow process!!!

- Are there any solutions to analyse code quickly without reverse engineering?

- Could it be possible to add this kind of solutions as new features to our classifier?

Crazy idea?

- **Why not detecting patterns analysing .DEX files and Android bytecodes (opcodes) directly?**

Eleven Paths

A Telefónica COMPANY

06

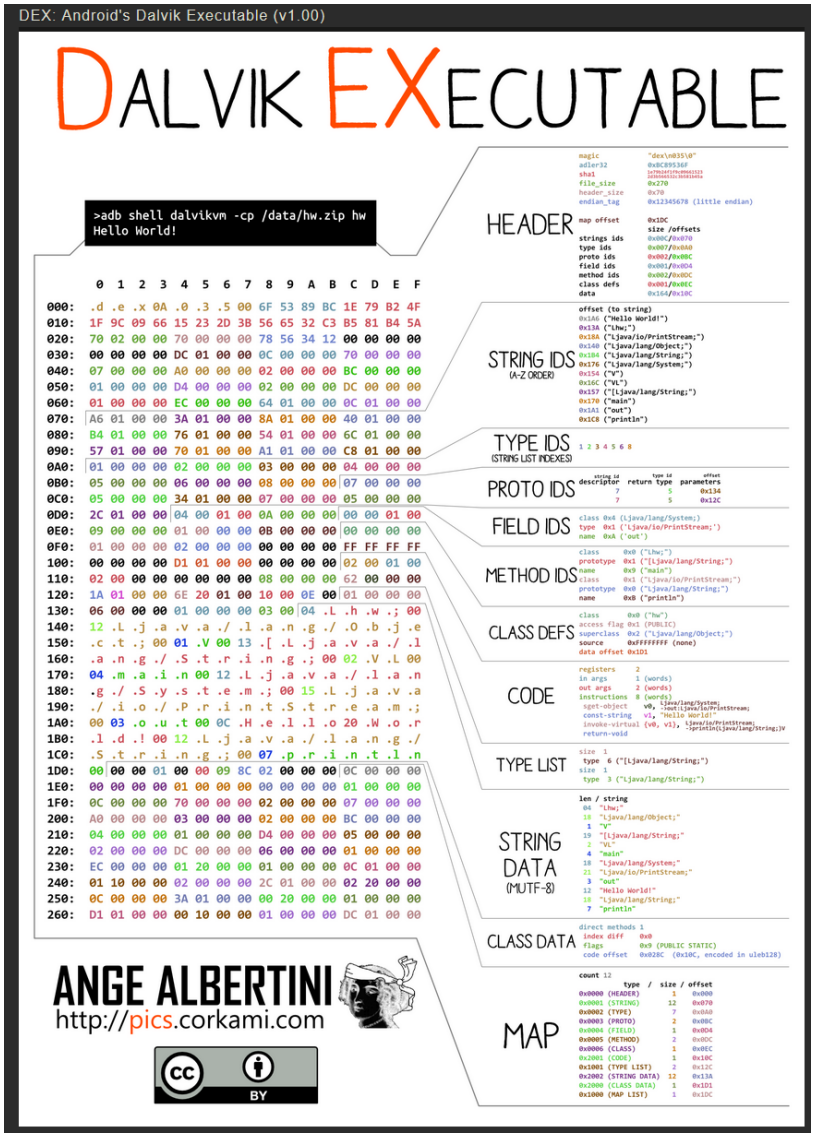
Machine Learning and Android bytecodes

Analyzing code without “reverse engineering or code execution”

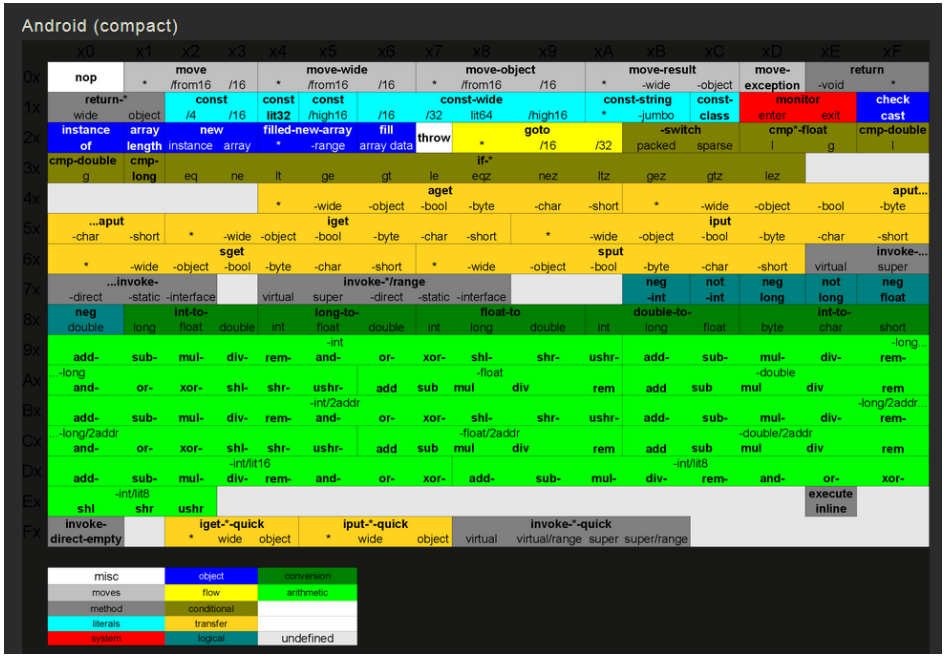
(Research in progress)

06 Analyzing DEX (Dalvik-bytecode)…

For the same datasets (goodware/malware) used “before”…



Example: NOP 0x00h, ADD 0x90h,…



OPCODES table: https://imgur.com/a/N5bqg#0

CODE

```
registers      2
in args        1 (words)
out args       2 (words)
instructions   8 (words)
 sget-object     v0, Ljava/lang/System;
                    ->out:Ljava/io/PrintStream;
 const-string    v1, "Hello World!"
 invoke-virtual  {v0, v1}, Ljava/io/PrintStream;
                    ->println(Ljava/lang/String;)V
 return-void
```

06 Analyzing DEX (Dalvik-bytecode)...

For the same datasets (goodware/malware) used "before"…

- **Why these features? Intuitions & "some" previous researches…**

Using opcode-sequences to detect malicious Android applications IEEE ICC 2014

Jerome, Q. Allix, K. ; State, R. ; Engel, T.
Interdiscipl. Center for Security Reliability & Trust, Univ. of Luxembourg,

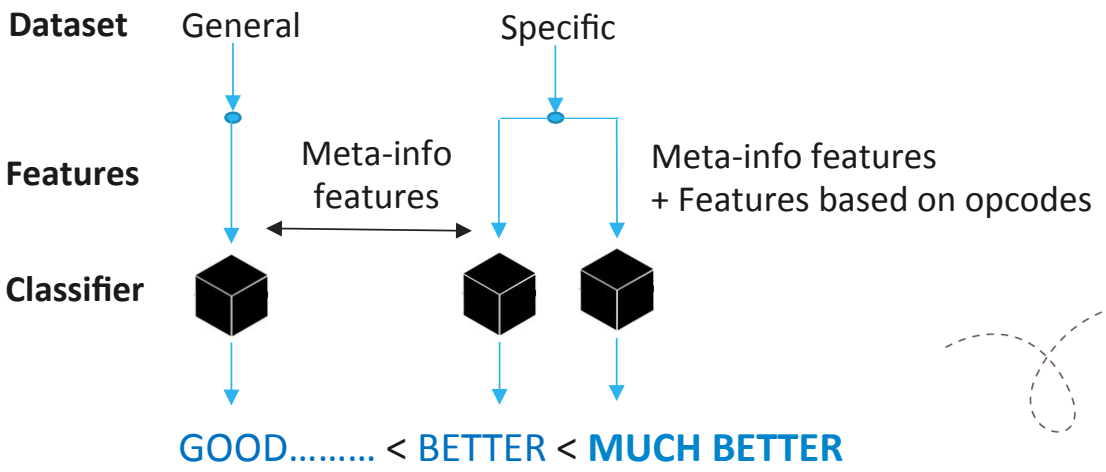
- **We collect new features per each app:**
  - N-gram opcodes (is a contiguous sequence of *n* opcodes in each .dex file)
  - Number of Opcodes (number of instructions)
  - Size of code section
  - % per each type of Opcode
  - % per each instruction format (30 formats)
    - https://source.android.com/devices/tech/dalvik/dalvik-bytecode.html
  - % per groups of instructions (13 formats)
    - Conditional, transfer, flow, arithmetic, moves, literals…
  - Entropy and Entropy per "blocks of N instructions"

- Does Malware have less code? Is it obfuscated? Opcodes pattern helps to distinguish malware? Developer's programming style?

06 Classifying using Opcodes...

For the same datasets (goodware/malware) used “before”…

- These new features do not classify malware better than our previously used meta-info features (permissions, size, description, certificate…) but, both combined per category they do!!!
- **Category classification (metainfo features + opcodes features):** We improve the previous results with opcodes patterns (we have again problems with some categories if there are not “enough” samples)



06 Category classification…

Metainfo features + Features based on opcodes…

- We are getting a very good results per category…
- Example: BOOK_AND_REFERENCE category

(Remember: with features based on permissions only → Accuracy=77.74% (for this category)

https://en.wikipedia.org/wiki/F1_score

"Global Classification Features based on META-INFO	Features based on META-INFO BOOK_AND_REFERENCE	META-INFO + OPCODES PATTERN Categ: BOOK_AND_REFERENCE
Accuracy = 91.015555%	Accuracy = 96.26 %	Accuracy = 97.06 %
Recall = 90.2507% (true positive rate) [true negative rate = 92.9468%]	**Recall = 96.0937% (tpr) [tnr = 96.6386%]**	**Recall = 96.8627% (tpr) [tnr= 97.5%]**
Precision = 97% (positive predictive value) [negative predictive value = 79.0432%]	**Precision = 98.4% (ppv) (npv = 92%)**	**Precision = 98.8% (ppv) (npv=93.6%)**
F1 score = 93.50%	F1 score = 97.2321%	F1 score = 97.8217%

- We proved, with different categories/datasets, that is possible to improve the classification results with opcodes... It is difficult to improve more by category :D

07
Conclusions

07 Conclusions

- Analyzing mobile malware is expensive. With 3k-5k new apps uploaded daily, it may be a good idea to priorize efforts filtering by suspicious apps first before deep analysis.

- Classifying apps without recurring to traditional ways, is usually based on characteristics of the app and Machine Learning classification. But this strongly depends on:

  - The malware/goodware set you choose, because:
    - You may have a “not big enough” set.
    - Not big enough set of characteristics so the apps in the set are properly “profiled”.
    - Not being very good at determining which is malware or not (VT dependency).
  - And what “features” you chose to define an app…
  - And what you consider as an “app” (take advantage of its circumstances)

- We discover a lot of features (meta-info and opcodes features) useful to classify adware/malware and goodware. We have a good detector, not relying on code, that offers accurate and high performance results.

tacyt : THE TOOL FOR APP CYBER INTELLIGENCE

Opcodes in Google Play

Tracing malicious Applications

Speakers

Alfonso Muñoz, PhD (@mindcrypt)
Senior Cybersecurity Researcher – alfonso.munoz@11paths.com

Sergio de los Santos (@ssantosv)
Head of Lab in ElevenPaths - ssantos@11paths