Revisiting Static Analysis of Android Malware

François Gagnon
frgagnon@cegep-ste-foy.qc.ca
www.cegep-ste-foy.qc.ca/cybersecurite

Frédéric Massicotte
Canadian Cyber Incident Response Centre (CCIRC)
Public Safety Canada

CSET 2017, Vancouver

CCIRC – Canadian Cyber Incident Response Center

CCIRC is Canada's national coordination centre responsible for reducing the cyber risks faced by Canada’s key systems and services. These systems, such as banks or phone service providers, are known as critical infrastructure.

CCIRC works within Public Safety Canada in partnership with provinces, territories, municipalities, private sector organizations and international counterparts. It also coordinates the national response to any serious cyber security incident.

Plan

- Intro
  - Objectives/Results
  - Android Static Analysis
- Experiment
  - Dataset
- Exploring Data:
  - Statistics & numbers
- Conclusion
  - Related work

Introduction

Intro - Android static analysis

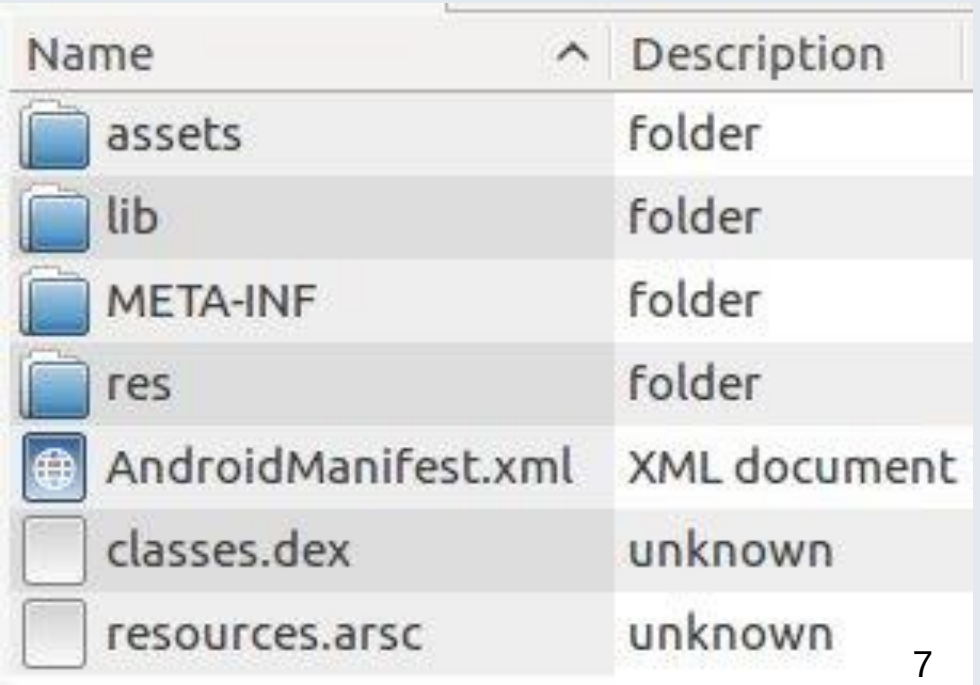
- Objectives :
  - Revisit previous static analysis results
    - Dataset (bigger, more recent)
    - Features (more)
  - Measure difference between legit and malicious apps
  - Find correlation between malware
  - Find information regarding malware production practices
  - Provide insights for cyber incident responders

Intro - Android static analysis

- Results :
  - Confirmed some previous findings, updated others.
  - General observations (statistics) for malware vs legitimate samples.
  - Specific observations of weird practices in malware production.

Intro - Android static analysis

- APK = archive
- What is inside an APK ?
  - A lot of files
  - .dex
  - Manifest
  - X509 Certificate



Stats - Experiment

Experiment- Dataset

Source	Status	Number
securityVendorX	Malware	208,221
Google Play	Legit	10,007

GooglePlay VirusTotal	
nbAVDetections	**nbApks**
0	8,301 (83%)
1 to 4	1,299 (13%)
5 to 9	189 (2%)
10 or more	218 (2%)
N/A	0 (0%)

Malware VirusTotal	
nbAVDetections	**nbApks**
0	144 (0.1%)
1 to 4	4,659 (2%)
5 to 9	9,830 (5%)
10 or more	191,611 (92%)
N/A	1,977 (1%)

- The dataset has not been pre-filtered.

Experiment- Dataset

	Malware		Legitimate	
Year	**nb**	**%**	**nb**	**%**
<2008	1,962	0.94%	35	0.35%
2008	4,581	2.20%	18	0.18%
2009-2013	6,022	2.89%	1,281	12.80%
2014	6,012	2,89%	1,617	16.16%
2015	**87,442**	**42.00%**	**3,156**	**32.54%**
2016	**102,155**	**49.07%**	**3,900**	**38.97%**
2017	2	0.00%	0	0.00%

Experiment - Extracted Information

Manifest	X509 Certificate	FileList
permissions freq.	nbCertFiles	nbFiles
duplicate perm.	nbCerts	nbDEX
mindSDKVersion	filename	sharedFiles
targetSDKVersion	serialNumber	EmbeddedAPK
appVersionCode	signature	
appVersionName	publicKey	
appName	certStartDate	
appPackage	certEndDate	Dates
appLabel	subject & issuer	creationDate
	validityPeriod	creationDelta

Stats - Manifest

Station of Loading	Flight Number	Date	Aircraft Registration	Prepared by
AMS	MH-017	17. July 2014	9M-MRD	RIK

DANGEROUS GOODS

Station of Unloading	Air Waybill Number	Proper Shipping Name	Class or Division For Class 1 compat. grp.	UN or ID Number	Sub Risk	Number of Packages	Net quantity or Transp. Ind. per package	Ra act M Cat
	There is no evidence that any damaged or leaking packages containing dangerous goods have been loaded on the aircraft.							

OTHER SPECIAL LOAD

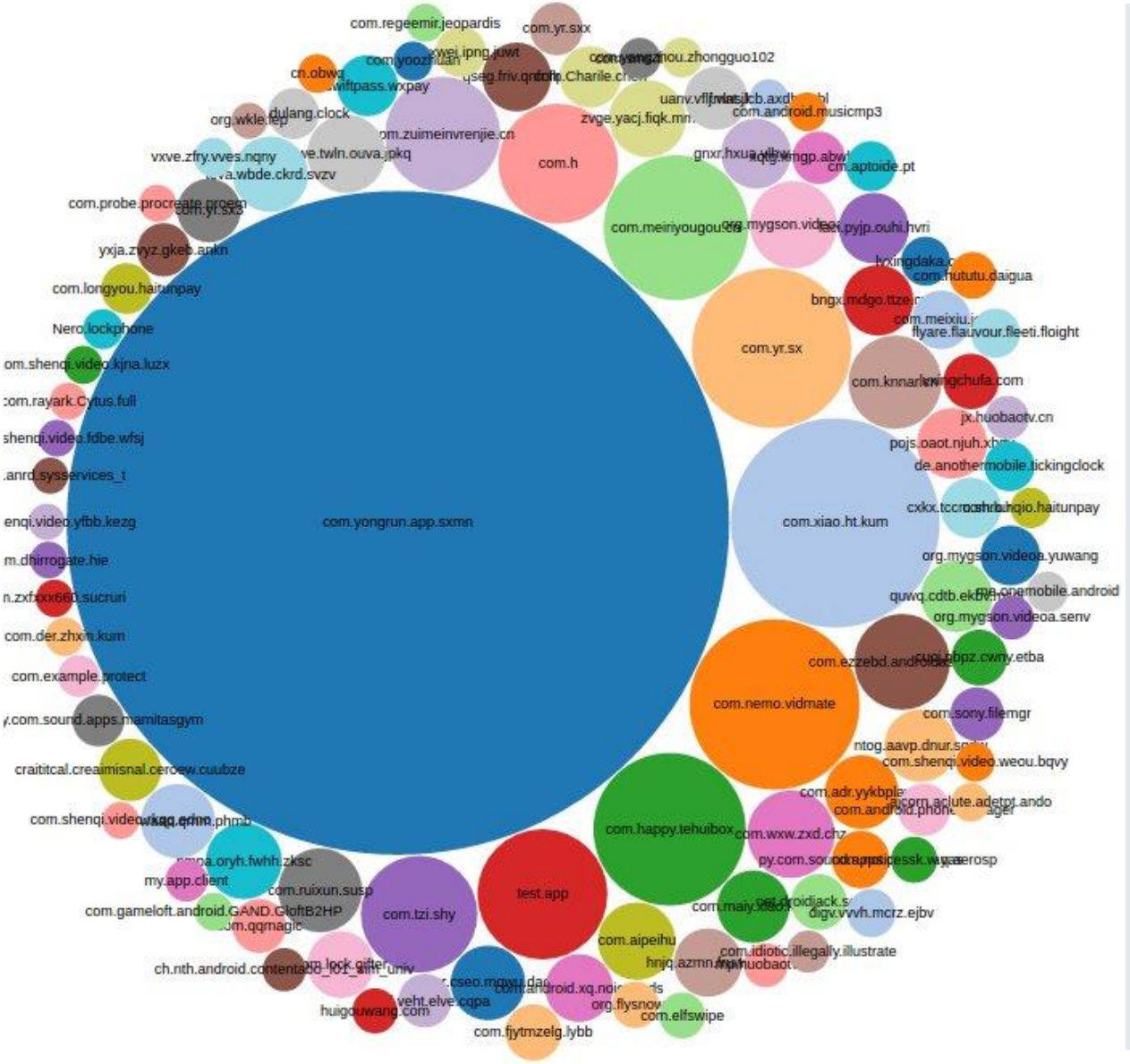
Station of Unloading	Air Waybill Number	Contents and Description	Number of Packages	Quantity	Supplementary Informa
KUL	232-12805085	MEDICALS	1	91.6 kg	KEEO COOL AT 2 AND 20 C
KUL	232-12790330	FRESH FLOWERS	12	215 kg	KEEP COOL AT 3 AND 5 C
KUL	232-12774134	LIVE PIGEONS	4	82 kg	80X60X30CM
KUL	232-11342295	LIVE BIRDS	5	70 kg	
KUL	232-12809591	LIVE DOG	1	30 kg	
KUL	232-12809635	LIVE DOG	1	20 kg	

Stats - Manifest - appPackage

Dataset	NbDistinct	NbTotal	Ratio
Legit	9,996	10,006	0.999
malware	106,574	205,008	0.51

Malware	
appPackage	Nb
“com.yongrun.app.sxmn”	33,729
prefix of “com.ym.ref.package.jxyq”	7,223

Stats - Manifest - appPackage

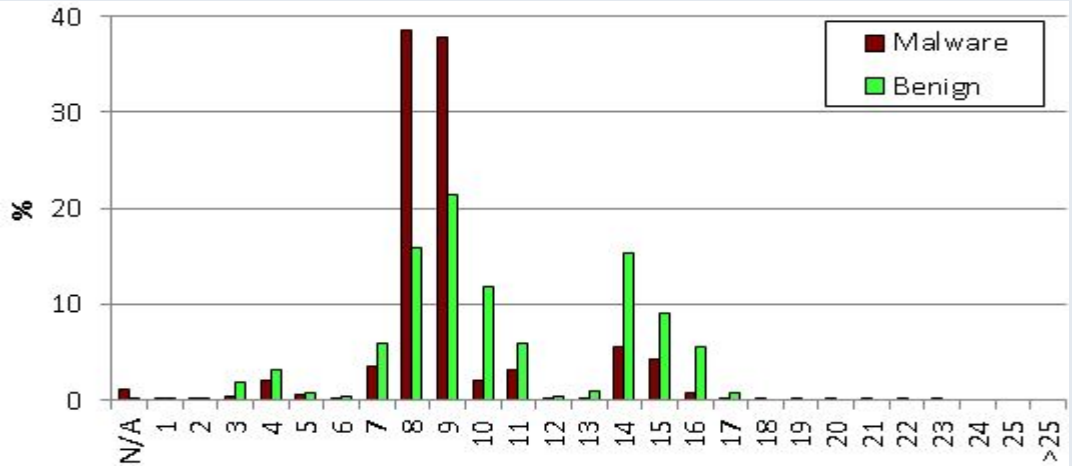


Stats - Manifest - appVersionCode

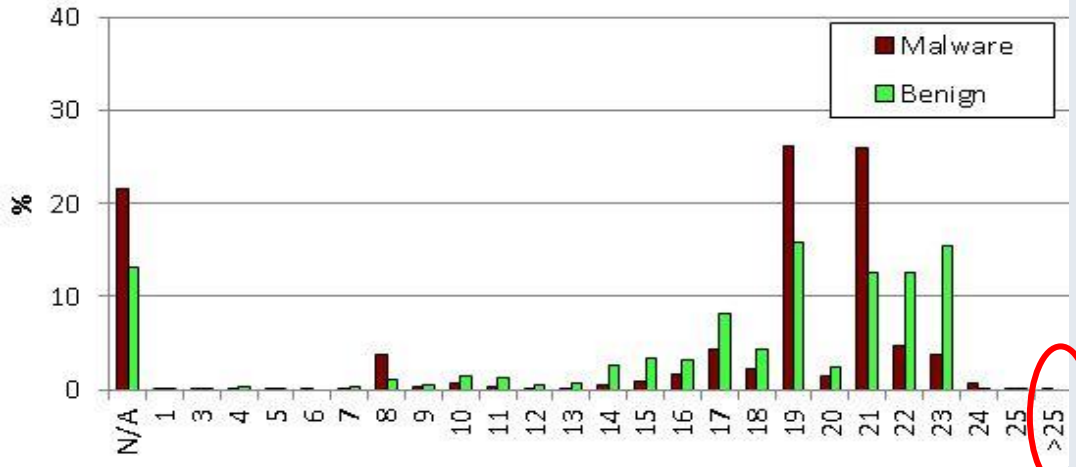
Dataset	NbDistinct	Ratio
Legit	1,624	0.162
malware	9,757	0.048

appVersionCode	Nb Malware	Nb Legit
= 2 147 483 647	3.5%	0% (1)
> 2 100 000 000	4%	0% (6)
> 1 000 000	8%	7%
< 10	54%	47%
= 1	42%	14%

Stats - Manifest - (min/target)SDKVersion



minSDKVersion



targetSDKVersion

25 samples > 25

1 sample = 999

Stats - Manifest - duplicatePermissions

	Nb Malware	Nb Legit
has duplicatePermissions	**65%**	**7.6%**
avg duplicatePermissions	**17**	**0.2**

CONJECTURE: Repackaging process

Stats - Certificates

Stats - Certificates - signature

Dataset	NbDistinct	Ratio
Legit	8,496	0.85
malware	126,024	0.605

Same certificate signs more than X samples	Nb Malware	Nb Legit
10,000	1	0
1,000	8	0
100	59	0

Same certificate
⇒
Same origin

Different certificates
⇏
Different origins

Stats - Certificates - publicKey

- 9 malware samples have distinct signatures but still the same public key.
  - Most likely same origin

00940bb5f7bd124e77213402b118cf19bda6426b84c1bdf7c90157edab0dc91c7d70147
207cbfc170fc0671a456c33e5e962f462f3324c0094d1d807301f0a0639341ceb4e7b2f6e
01414241d6920cd44f64726614351b6dec9c8218ae4b6b23b4fdf39d6e57221793cdfe3
4cf23e1104f63065b82e0a7317b883d208dddfa93f9dc53da769aea7bdd8c64ae7a2dcd
3c302f7baf9b4428d3a3ec19c3cd47c5872182260f96f0af1e772603357c5a99f905334cb
fe09362f10dfadf78254a484df33ca13e8f269ecb1ba121f45e5cff4332fbe4be94a3ebd71
25a2d3018fa4897dcd92b2d4bd547ecd32d5efcee6636cc333f67f6a1e1f9e8822712927
248111f8f

CONJECTURE: Generate a new certificate for each sample, but always reuse the same key for all those certificates...

Stats - Certificates - subject

Dataset	NbDistinct	NbDistinct Signatures
Legit	8,159	8,496
malware	62,093	126,024

Most Popular Subjects in Malware Dataset

Type	Certificate Subject	nbSamples
Generic	C=y, ST=y, L=y, O=y, OU=y, CN=y	35,673
Generic	C=CN, ST=Unknown, L=Unknown, O=Unknown, OU=Unknown, CN=Unknown	19,562
Generic	C=Unknown, ST=Unknown, L=Unknown, O=Unknown, OU=Unknown, CN=Unknown	2,000
Specific	ST=haitunpay, L=haitunpay, O=haitunpay, OU=haitunpay, CN=haitunpay	4,569
Random	C=llpfihhdydkfffovoaxbl1bm, ST=llpfihhdydkfffovoaxbl1bm, L=llpfihhdydkfffovoaxbl1bm, O=llpfihhdydkfffovoaxbl1bm, OU=llpfihhdydkfffovoaxbl1bm, CN=llpfihhdydkfffovoaxbl1bm	3,324

Stats - Certificates - fileName

Dataset	NbDistinct	NbDistinct Signatures
Legit	926	0.093
malware	17,905	0.086

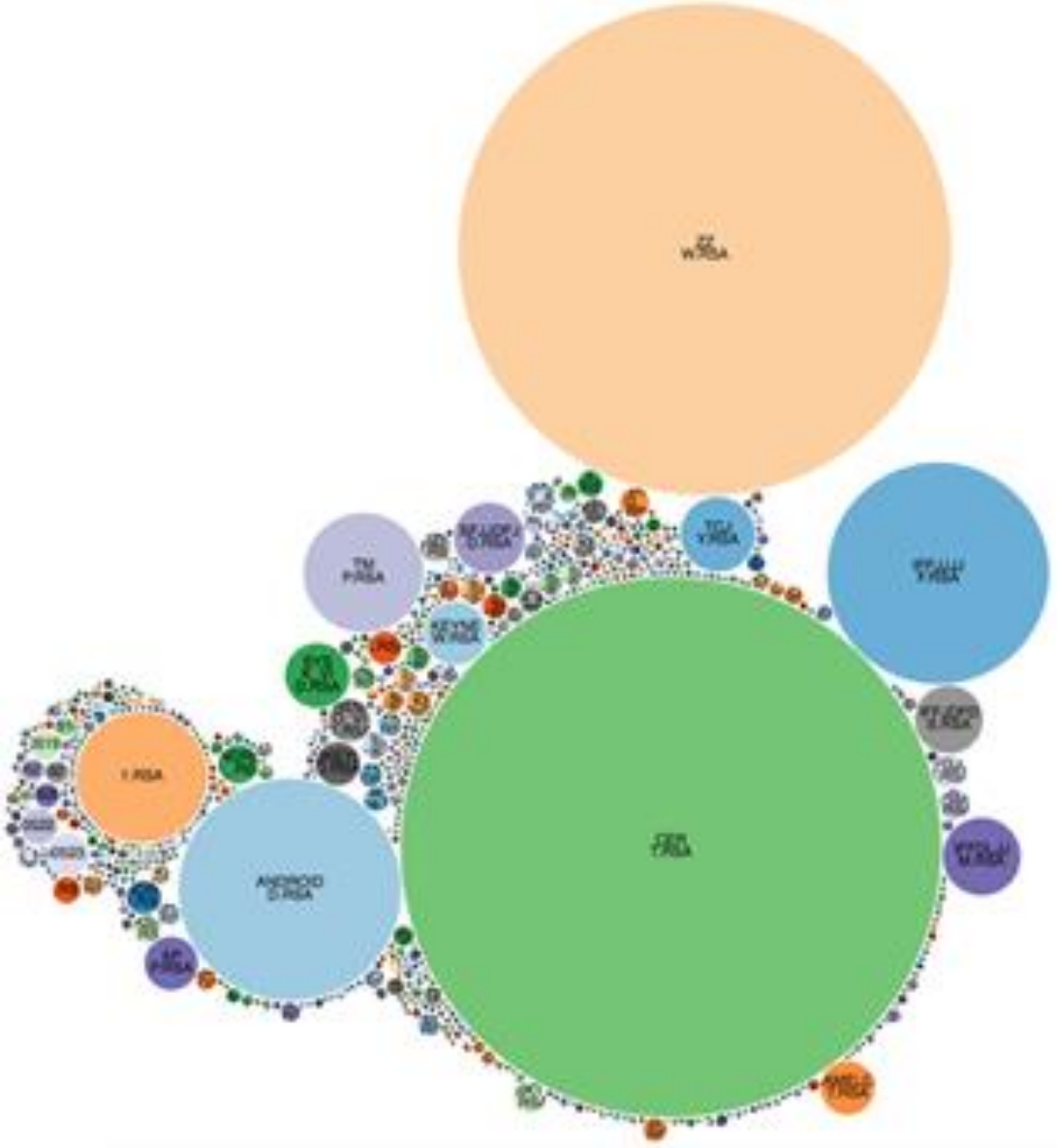
Popular certificate file names	Nb Malware	Nb Legit
CERT.RSA	89,330 (42.9%)	8,379 (83.7%)
ZZW.RSA	36,232 (17.4%)	0 (0%)
Android.RSA	13,505 (6.5%)	10 (0%)

- 2,249 malware samples have a file name following the pattern “8 digits”.RSA

1,594 of those have a creation date that exactly matches the fileName

yyyymmdd

Stats - Certificates - fileName



Stats - Certificates - (start/end)Date

- It is not entirely clear to me what is the impact of those dates in Android.
  - An app can be updated only if it’s certificate is not expired
  - To be accepted on Google Play:
    - The certificate of an app must not expire before 2033-10-22
    - The certificate must have started its validity.
    - But they could still be installed on a device
- This info is harder to leverage.

Stats - Certificates - (start/end)Date

- Two strategies:
  a. Delta between end and start date (validityPeriod)
  b. Delta between start date and APK packaging date

	Malware	Legit
min value	1 day	18 years
nb with validity < 1 year	13,568 (6.5%)	0 (0%)
avg value	10 years	189 years

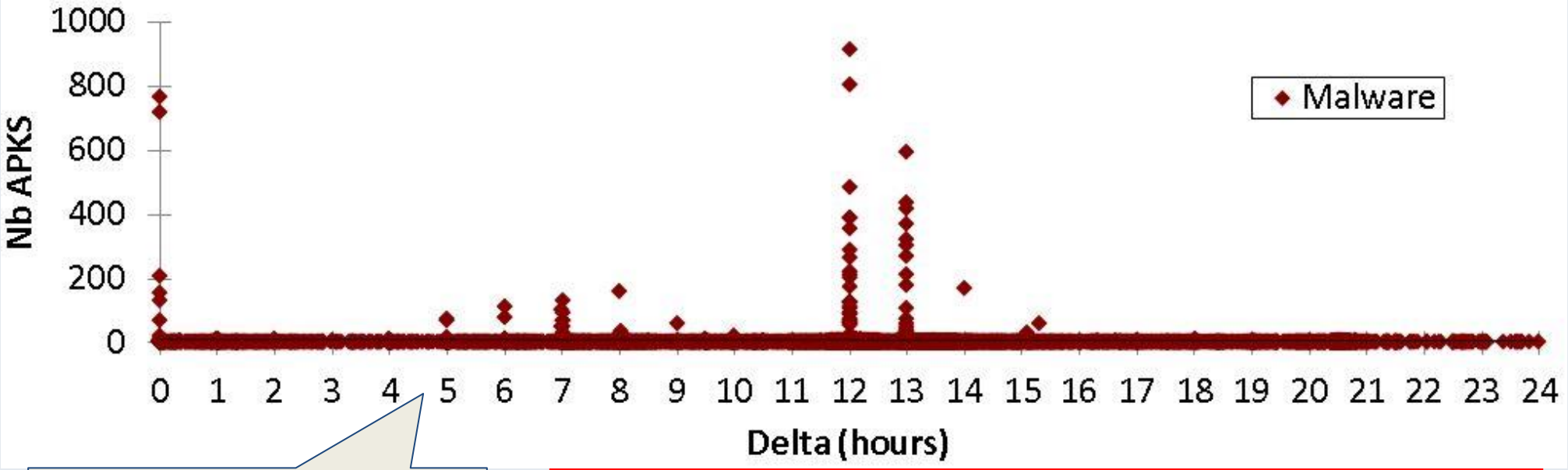
OpenSSL:
-in days
-default 30

AndroidStudio:
-in years
-default 25

Stats - CreationDate

Stats - Dates - creationDate

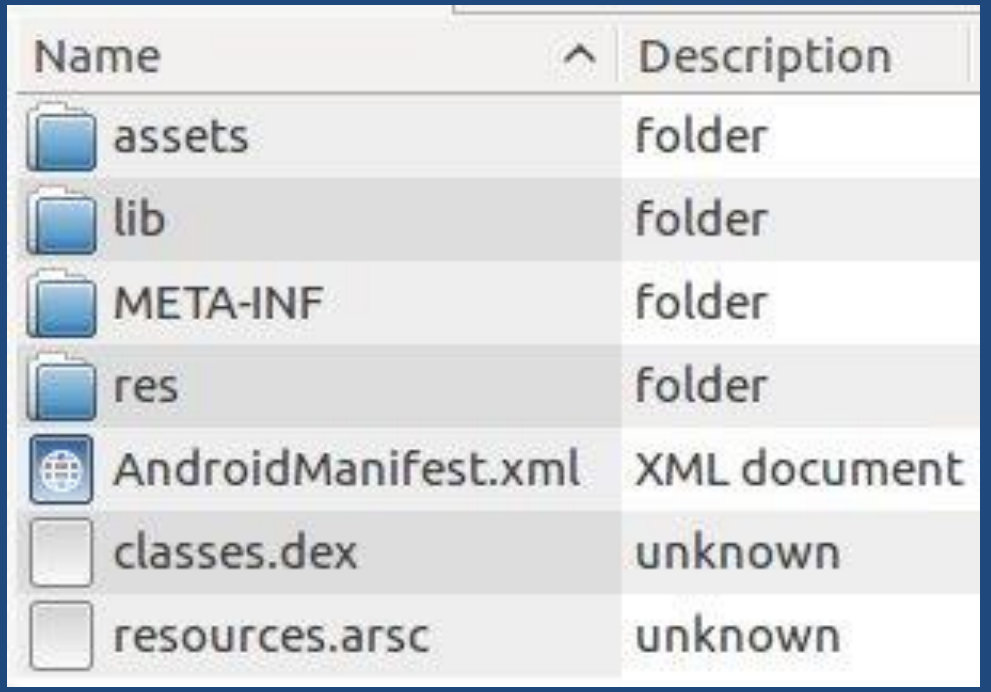
- Now, let’s observe delta between creationDate and certStartDate (with second-level precision):
  - To get insight on the creation process...



Deltas tend to be grouped right on the hour.

CONJECTURE: certificate and apk are created on different machines in different timezones

Stats - FileList



Name	Description
assets	folder
lib	folder
META-INF	folder
res	folder
AndroidManifest.xml	XML document
classes.dex	unknown
resources.arsc	unknown

Stats - FileList - misc

	Malware	Legit
avg number of files inside APK	247	606
number of samples containing a specific DEX file [DEX file md5: cfdba92d344b57fecabadab26296f84c]	8,250	0
number of samples containing an APK inside	13,211	97

Conclusion

Conclusion

- Malware creation is automated but still careless
  - Many artifacts
  - Easy correlation/identification for low hanging fruits
- Revisiting is good as some things evolve
- Some interesting weird stuff:
  - APKs inside APKs
  - Time delta (12h) between cert creation and apk packaging
  - Duplicate permissions

Conclusion – Related Work

- Scope
  - <= 5 feature vs > 25
  - most < 5,000 malware vs > 200,000
  - most used older malware.
- [8]: malware appVersion < benign appVersion
  - just for appVersion = 1
  - appVersionName instead of appVersionCode
- [12]: malware request more permissions
- [10&15]: most requested permissions mal vs ben
  - mal vs benign profiles are still different
  - changes on most popular permissions

Conclusion – Related Work

- [9] used serial number to distinguish certificates
  - signature should be used instead
  - 5 groups of distinct certificates (signature, public key) have the same serial number
- [9] number of distinct certificates seen
  - 622/4,554 (0.14) vs 126,024/208,221 (0.61)
- [3] malware are created on a Mon-Fri schedule