Institut
Mines-Telecom

Identifying Unknown Android Malware with Feature Extractions and Classification Techniques

Ludovic Apvrille
ludovic.apvrille@telecom-paristech.fr

Axelle Apvrille
aapvrille@fortinet.com

TRUSTCOM'2015

Many Android Applications (and Malware!)

Application repositories

- Google Play: 1.6 million+
- F-Droid, APPSAPK, APKTOP, ...

Many Android Applications (and Malware!)

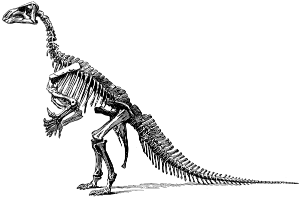
Application repositories

- Google Play: 1.6 million+
- F-Droid, APPSAPK, APKTOP, . . .

Malware

- Aug. 2015. **2 millions+** malicious Android **samples**
- **2,000+ new malicious Android samples every day**

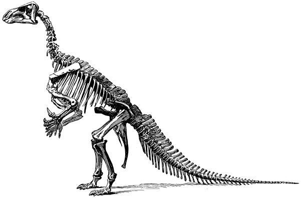
Malware: Android Carbon 14 Dating ;)



Shortest detection delay for some samples by **all AV vendors**

Name	**Creation date**	**Detection date**
Android/Wroba	June 16 2014	June 21 **+5d**
Android/Curesec	July 3 2014	July 11 **+8d**
Android/ScarePakage	July 13 2014	July 24 **+11d**

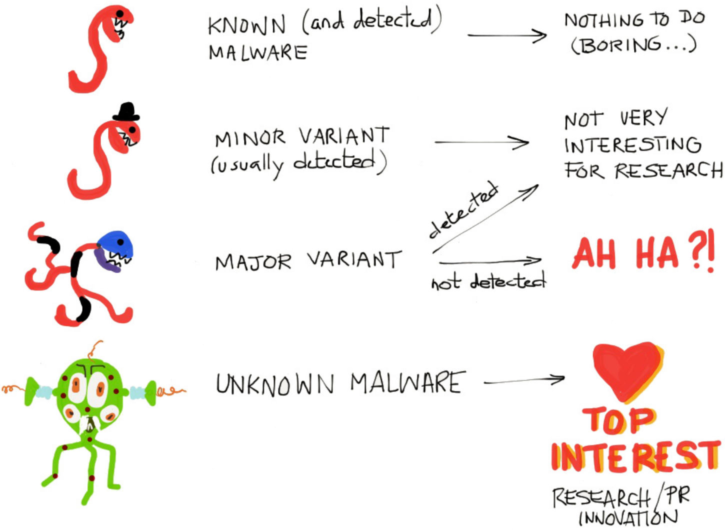
Malware: Android Carbon 14 Dating ;)



Shortest detection delay for some samples by **all AV vendors**

Name	Creation date	Detection date
Android/Wroba	June 16 2014	June 21 **+5d**
Android/Curesec	July 3 2014	July 11 **+8d**
Android/ScarePakage	July 13 2014	July 24 **+11d**
Android/Ganlet	**Nov 1 2013**	**May 15 2014 +6 months!!!**

What Are We Interested In?


KNOWN (and detected) MALWARE
NOTHING TO DO (BORING...)
MINOR VARIANT (usually detected)
NOT VERY INTERESTING FOR RESEARCH
MAJOR VARIANT
detected
not detected
AH HA ?!
UNKNOWN MALWARE
TOP INTEREST
RESEARCH / PR INNOVATION

Problems with Manual Search

Too many apps and marketplaces to crawl

Waste time on clean apps

Even a **team of 100 analysts is insufficient**

Problems with Manual Search

Too many apps and marketplaces to crawl
Waste time on clean apps
Even a **team of 100 analysts is insufficient**

We need an automated system that helps identifying unknown malware with less effort

→ SherlockDroid

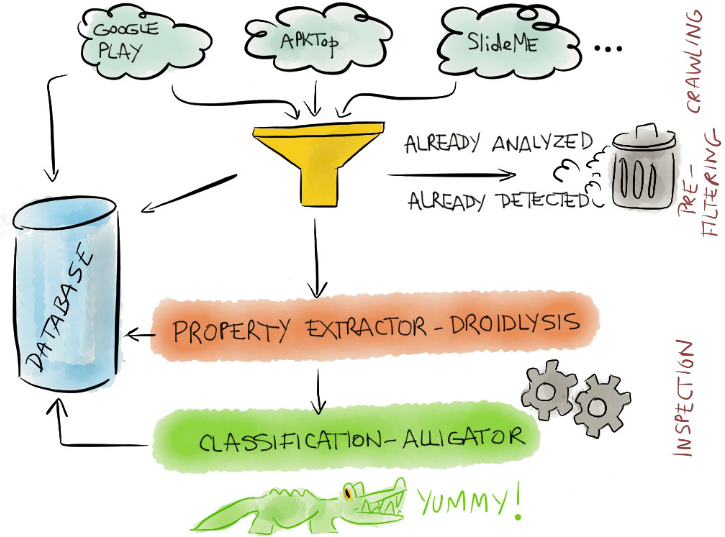
SherlockDroid to the Rescue!

Crawl Android marketplaces

Spot suspicious apps

Focus on major variants and unknown malware

SherlockDroid Architecture



Remarks on SherlockDroid

It is not an AV scanner

because SherlockDroid does not handle known malware / minor variants

Remarks on SherlockDroid

It is not an AV scanner

because SherlockDroid does not handle known malware / minor variants

We will miss some malware

We're not (yet) perfect ;-)

Remarks on SherlockDroid

It is not an AV scanner

because SherlockDroid does not handle known malware / minor variants

We will miss some malware

We're not (yet) perfect ;-)

But we would have missed them without SherlockDroid too

DroidLysis - Extracting Properties/Features

Static extraction of 289 properties

54 File-related properties

- Permissions, certificate, . . .

22 Resource properties

- Native code, resource risky calls (*su*, *mount*, etc.), Javascript, URLs, . . .

70 Dalvik code properties

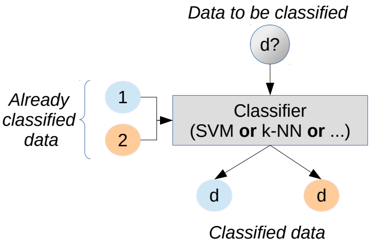
- API usage, actions, intents, constants, implementation techniques (e.g., *junk bytecode injection*)

143 Third party kits properties

- Advertisements, statistics reporting, error reporting

Alligator - Classification (1/2)

Usual classification approach



Data to be classified

d?

Already classified data

1

2

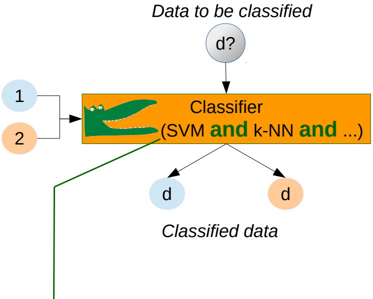
Classifier
(SVM **or** k-NN **or** ...)

d

d

Classified data

Alligator



Data to be classified

d?

1

2

Classifier
(SVM **and** k-NN **and** ...)

d

d

Classified data

Alligator automatically combines classification algorithms in order to obtain better classification results

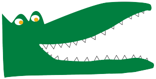
Alligator - Classification (2/2)

Other capabilities

- Favor a cluster over another
- Forget/boost too abnormal elements

Other

- Shown to better classify than other classifiers (e.g., SVM) in various application domains (e.g., image classification)
- Free and open-source, easy to install and configure, scriptable



`alligator.telecom-paristech.fr`

SherlockDroid: Hall of “Fame”

- Android/MisoSMS.A!tr.spy
- Android/Odpa.A!tr.spy
- Adware/Geyser!Android
- Riskware/Flexion!Android
- Riskware/SmsControlSpy!Android
- Riskware/Zdchial!Android
- Riskware/SmsCred!Android
- Riskware/Blued!Android
- Riskware/SneakFont!Android

Descriptions: http://www.fortiguard.com/encyclopedia/

SherlockDroid: Unknown Malware Identified

Do you known any other framework who identified real unknown malware?

SherlockDroid: Unknown Malware Identified

Do you known any other framework who identified real unknown malware?

Answer: DroidRanger: **2**

SherlockDroid: Unknown Malware Identified

Do you known any other framework who identified real unknown malware?

Answer: DroidRanger: **2**

AAS, Andromaly, CopperDroid, Crowdroid, Drebin, MADAM, MAST, pBMDS, PUMA...
tested on *artificial* or *known malware*

Learning and Classification Results

Typical results we expect

- FP/FN shall be as low as possible (Obviously)
- FP shall be much lower than FN (Missing a malware is not a big deal w.r.t. wasting time on false alerts)

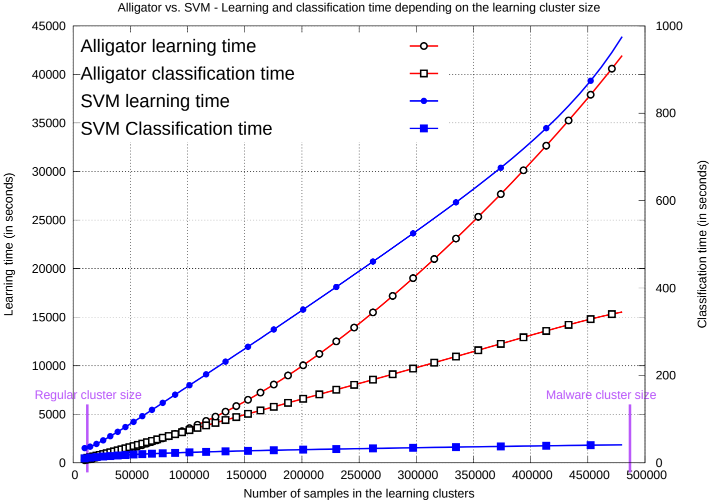
Samples

- **Learning clusters**: 500k samples used in the learning clusters
  - $\sim$ 487k malware, $\sim$ 12k clean
  - Gathered before June 2014
- **Testing clusters**: 1.5k clean and 3k malware gathered after Sept. 2014

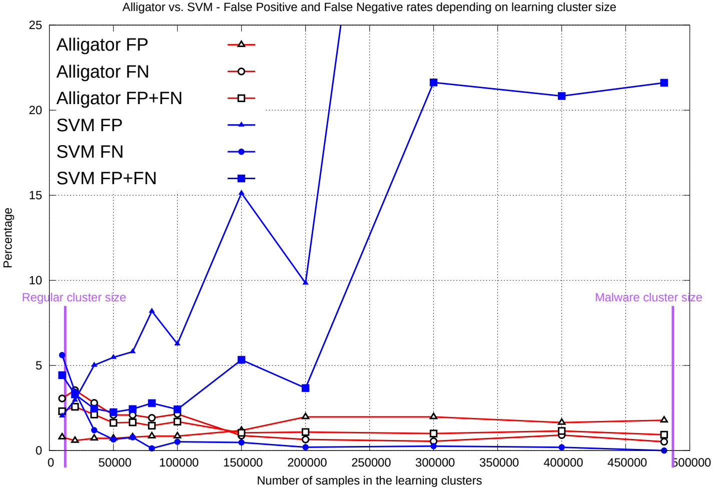
Learning and Classification Results (Cont.)

Learning cluster size	Learning time	Classification time	FP	FN	Average
480,000	∼ 11 hours	6 mn	1.78%	**0.52%**	**0.93%**
50,000	20 min	∼ 34 s	**0.72%**	2.1%	**1.67%**

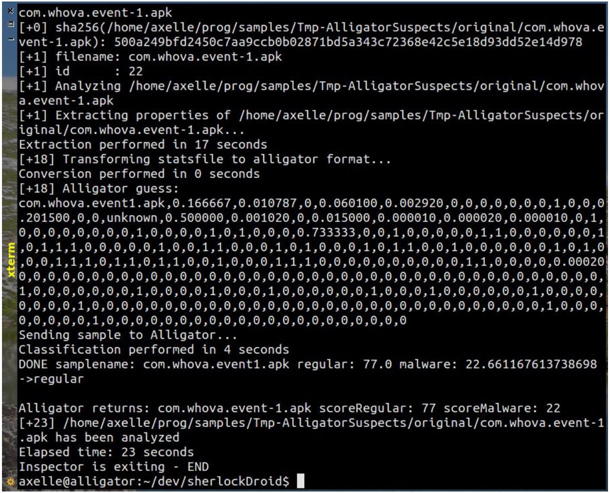
Learning and Classification Results: Comparison



Learning and Classification Results: Comparison (Cont.)



What About the TrustCom Android App?



```
com.whova.event-1.apk
[+0] sha256(/home/axelle/prog/samples/Tmp-AlligatorSuspects/original/com.whova.e
vent-1.apk): 500a249bfd2450c7aa9ccb0b02871bd5a343c72368e42c5e18d93dd52e14d978
[+1] filename: com.whova.event-1.apk
[+1] id      : 22
[+1] Analyzing /home/axelle/prog/samples/Tmp-AlligatorSuspects/original/com.whov
a.event-1.apk
[+1] Extracting properties of /home/axelle/prog/samples/Tmp-AlligatorSuspects/or
iginal/com.whova.event-1.apk...
Extraction performed in 17 seconds
[+18] Transforming statsfile to alligator format...
Conversion performed in 0 seconds
[+18] Alligator guess:
com.whova.event1.apk,0.166667,0.010787,0,0.060100,0.002920,0,0,0,0,0,0,0,1,0,0,0
.201500,0,0,unknown,0.500000,0.001020,0,0.015000,0.000010,0.000020,0.000010,0,1,
0,0,0,0,0,0,0,0,1,0,0,0,0,1,0,1,0,0,0,0.733333,0,0,1,0,0,0,0,0,1,1,0,0,0,0,0,0,1
,0,1,1,1,0,0,0,0,0,1,0,0,1,1,0,0,0,1,0,1,0,0,0,1,0,1,1,0,0,1,0,0,0,0,0,0,1,0,1,0
,0,0,1,1,1,0,1,1,0,1,1,0,0,1,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0.00020
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
1,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,
0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,
0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Sending sample to Alligator...
Classification performed in 4 seconds
DONE samplename: com.whova.event1.apk regular: 77.0 malware: 22.661167613738698
->regular

Alligator returns: com.whova.event-1.apk scoreRegular: 77 scoreMalware: 22
[+23] /home/axelle/prog/samples/Tmp-AlligatorSuspects/original/com.whova.event-1
.apk has been analyzed
Elapsed time: 23 seconds
Inspector is exiting - END
axelle@alligator:~/dev/sherlockDroid$ 
```

- Property extraction: 17 seconds
- Classification: 4 seconds
- Regular score: 77.0
- Malware score: 22.7
- → **Classified as regular ;-)**

Conclusion and Future Work

- SherlockDroid is operational, tested on a huge number of applications from various application markets
- 9 unknown malware identified
- For classification purpose, it relies on the Alligator meta-classifier

What's next?

- Feature extraction: mix contextual information (e.g., call stack) and the related features
    - Sending an email is not the same if it is for a bug report or for a connection to a C&C server
- Differentiate between malware and Potentially Unwanted Applications

Thank You

Contact info

SherlockDroid: aapvrille at fortinet dot com
Alligator: ludovic dot apvrille at telecom minus paristech dot com

References

Alligator Release: alligator.telecom-paristech.fr

A. Apvrille, L. Apvrille, "SherlockDroid: a research assistant to spot unknown malware in Android marketplaces", Journal of Computer Virology and Hacking Techniques, vol. 11, No. 39, pages 1-11, pub. Springer, july 2015

Powerpoint slides? No way! This is LATEX- Beamer !