TELECOM ParisTech

Institut Mines-Telecom

FORTINET

Identifying Unknown Android Malware with Feature Extractions and Classification Techniques

Ludovic Apvrille
ludovic.apvrille@telecom-paristech.fr

Axelle Apvrille
aapvrille@fortinet.com

Seminar at SAP

Many Android Applications (and Malware!)

Application repositories

- Google Play: 1.7 million+
- F-Droid, APPSAPK, APKTOP, . . .

Many Android Applications (and Malware!)

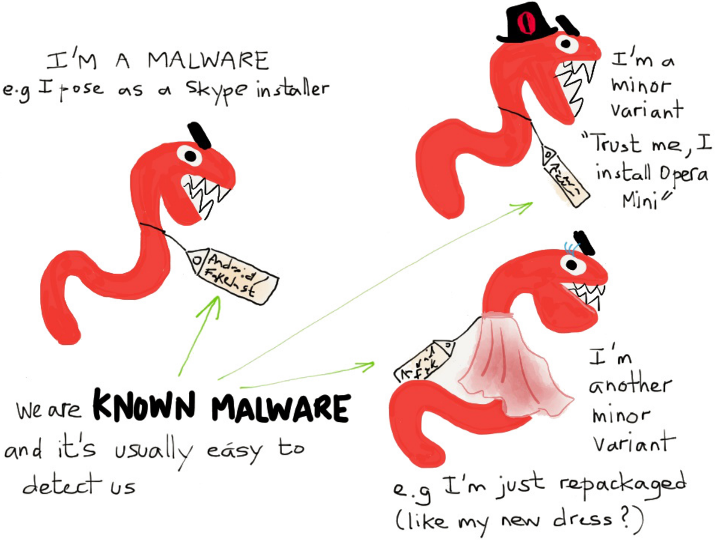
Application repositories

- Google Play: 1.7 million+
- F-Droid, APPSAPK, APKTOP, . . .

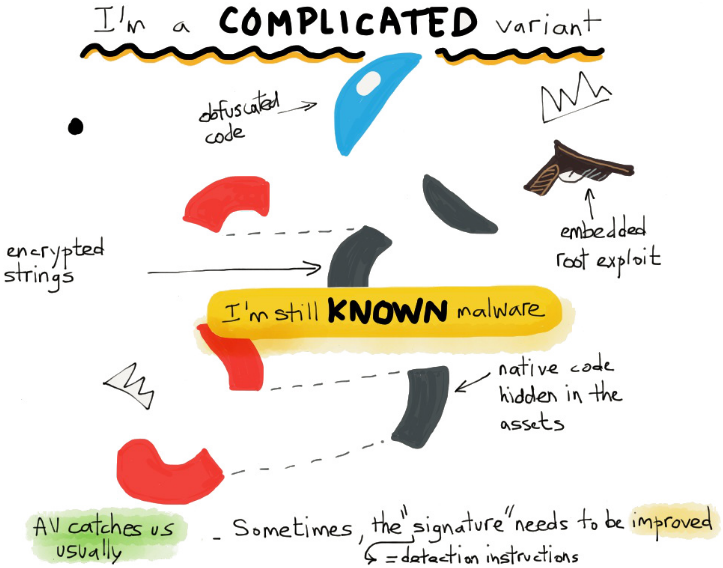
Malware

- Aug. 2015. **2.7 millions+** malicious Android **samples**
- **2,000+ new malicious Android samples every day**

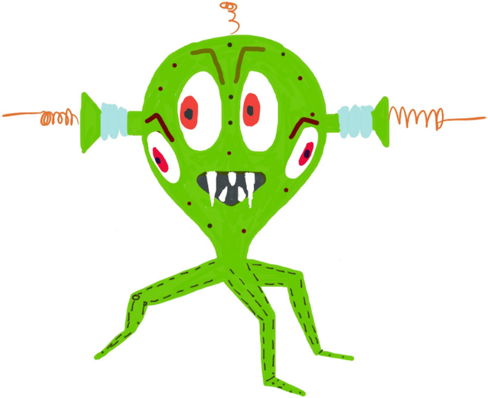
Known Malware


I'M A MALWARE
e.g I pose as a Skype installer
Android FakeInst
I'm a minor variant
"Trust me, I install Opera Mini"
We are KNOWN MALWARE and it's usually easy to detect us
I'm another minor variant
e.g I'm just repackaged (like my new dress?)

Known Malware

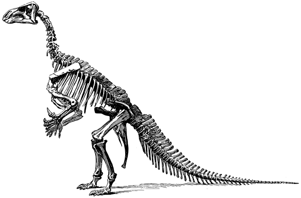


Unknown Malware



Do they exist? YES

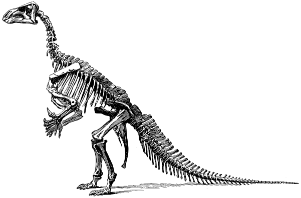
Malware: Android Carbon 14 Dating ;)



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**

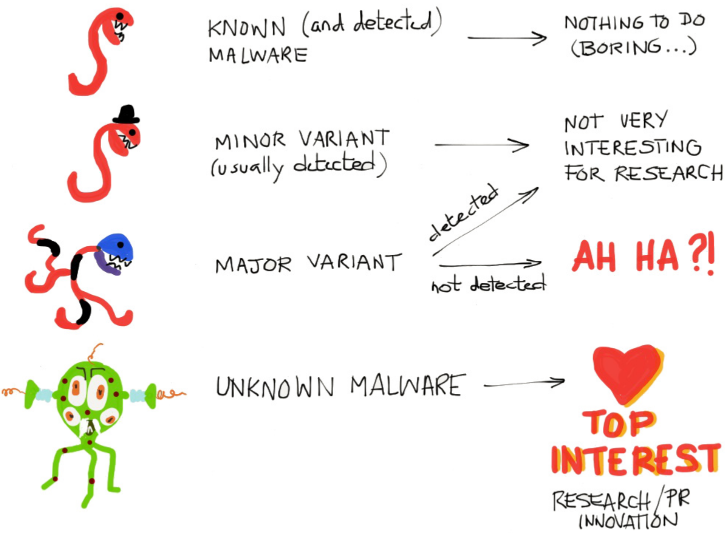
Malware: Android Carbon 14 Dating ;)



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

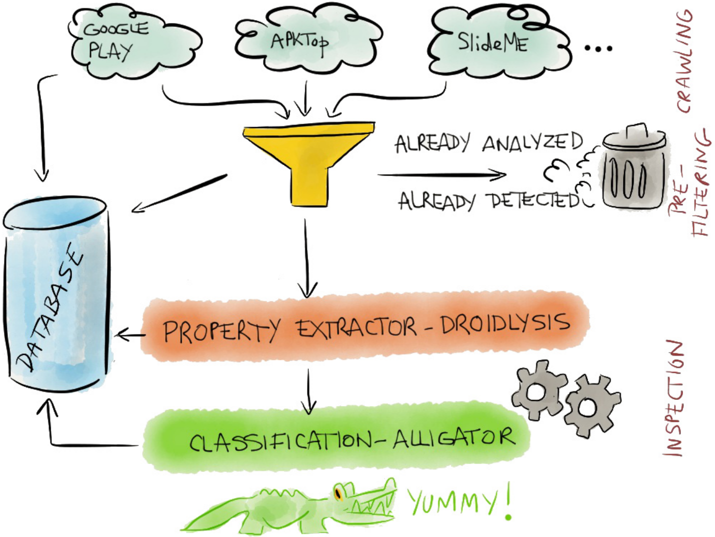
SherlockDroid to the Rescue!

Crawl Android marketplaces

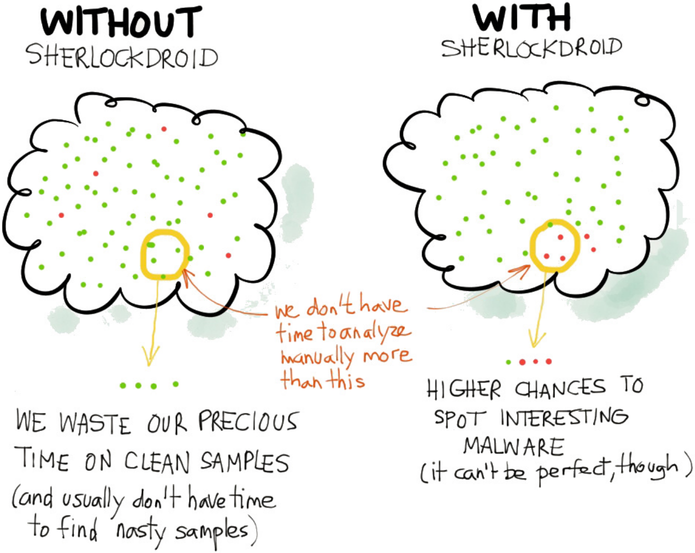
Spot suspicious apps

Focus on major variants and unknown malware

SherlockDroid Architecture



SherlockDroid (Unbiaised) Benefits


WITHOUT SHERLOCKDROID
WITH SHERLOCKDROID
we don't have time to analyze manually more than this
WE WASTE OUR PRECIOUS TIME ON CLEAN SAMPLES
(and usually don't have time to find nasty samples)
HIGHER CHANCES TO SPOT INTERESTING MALWARE
(it can't be perfect, though)

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

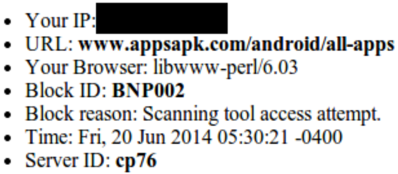
We will miss some malware

We're not (yet) perfect ;-)

But we would have missed them without SherlockDroid too

Crawlers - Evading Detection

Easy to implement
but constantly needs to be maintained :(



- Your IP: [redacted]
- URL: **www.appsapk.com/android/all-apps**
- Your Browser: libwww-perl/6.03
- Block ID: **BNP002**
- Block reason: Scanning tool access attempt.
- Time: Fri, 20 Jun 2014 05:30:21 -0400
- Server ID: **cp76**

- Search Limit
- Download activity per IP address
- User Agent verification
- Android ID verification https://github.com/Akdeniz/google-play-crawler

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, . . .

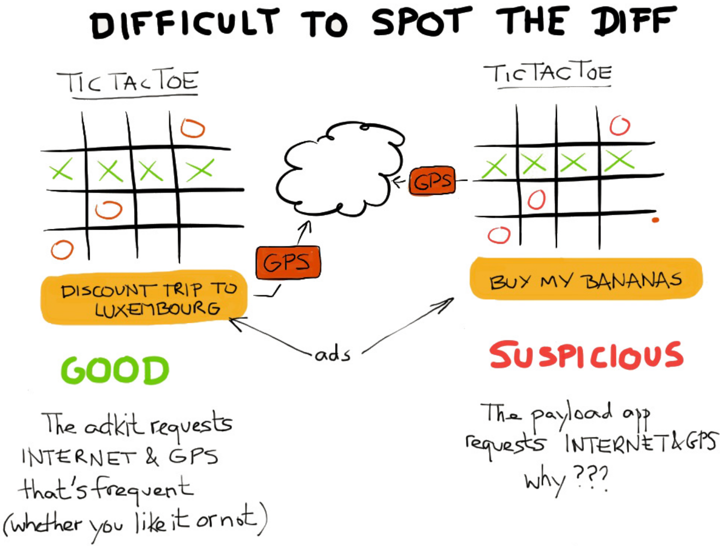
70 Dalvik code properties

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

143 Third party kits properties

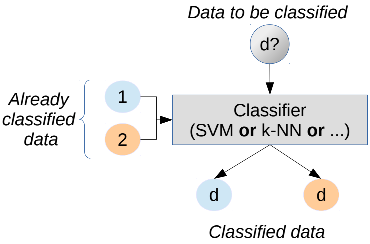
- Advertisements, statistics reporting, error reporting

Ruling out Third Party Code

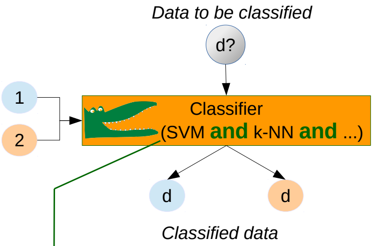

DIFFICULT TO SPOT THE DIFF
TICTACTOE
TICTACTOE
GPS
GPS
DISCOUNT TRIP TO LUXEMBOURG
BUY MY BANANAS
ads
GOOD
SUSPICIOUS
The adkit requests INTERNET & GPS that's frequent (whether you like it or not)
The payload app requests INTERNET&GPS why ???

Alligator - Classification (1/2)

Usual classification approach



Alligator



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

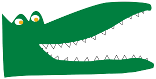
Alligator - Classification (2/2)

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**

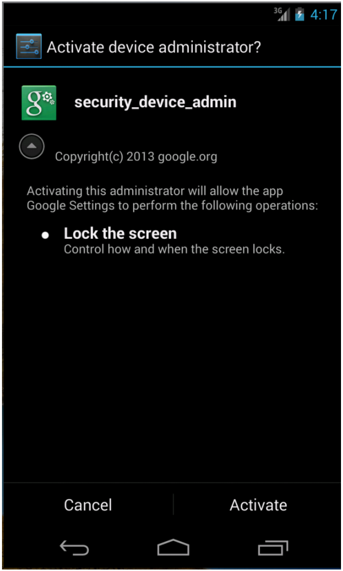
SherlockDroid: Unknown Malware Identified

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*

Into Android/MisoSms Trojan Spyware


4:17

Activate device administrator?

security_device_admin

Copyright(c) 2013 google.org

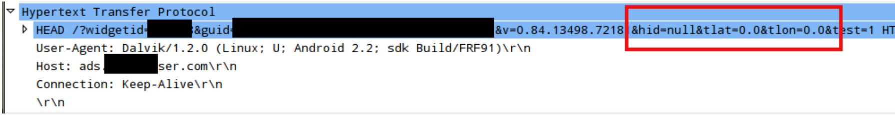
Activating this administrator will allow the app Google Settings to perform the following operations:

- **Lock the screen**
  Control how and when the screen locks.

Android/MisoSms.A!tr.spy

- Poses as Google Settings app
- Sends 1 initial email with phone number of victim
- Listens to incoming SMS
- Forwards them by email to attackers

Into Geyser Adware



```
▽ Hypertext Transfer Protocol
  ▷ HEAD /?widgetid=[redacted]&guid=[redacted]&v=0.84.13498.7218 &hid=null&tlat=0.0&tlon=0.0&test=1 HT
    User-Agent: Dalvik/1.2.0 (Linux; U; Android 2.2; sdk Build/FRF91)\r\n
    Host: ads[redacted]ser.com\r\n
    Connection: Keep-Alive\r\n
    \r\n
```

Adware/Geyser!Android

Posts GPS location in clear text
http://blog.fortinet.com/post/alligator-detects-gps-leaking-adware

LOL - In falsepositives.txt:

"Reputable companies including banks, US Government/ Military sector are using our tools"

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)

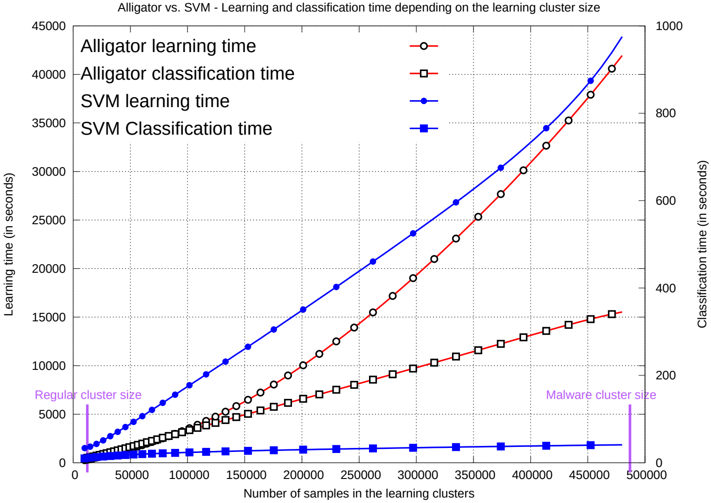
Samples

- **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

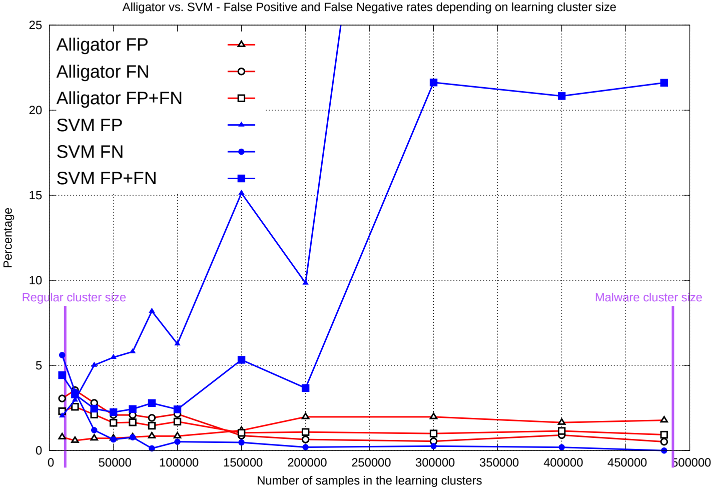
Learning and Classification Results (Cont.)

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

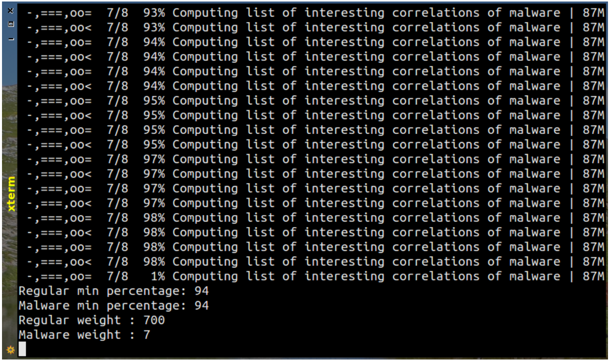
Learning and Classification Results: Comparison



Learning and Classification Results: Comparison (Cont.)



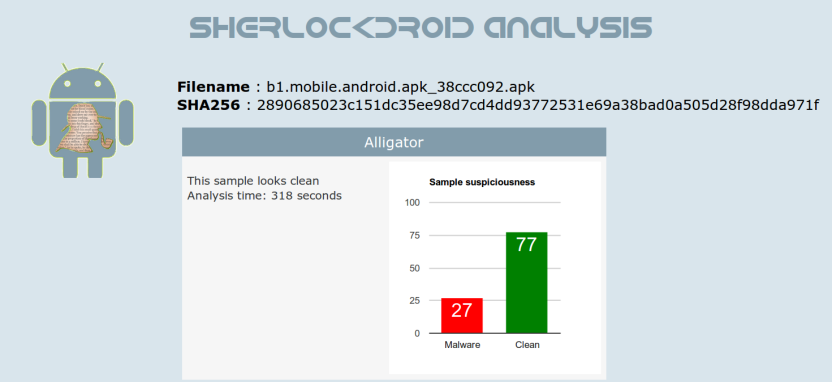
What About the SAP Android App?



```
 -,===,oo=  7/8  93% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  93% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  94% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  94% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  94% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  94% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  95% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  95% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  95% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  95% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  97% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  97% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  97% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  97% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  98% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  98% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8  98% Computing list of interesting correlations of malware | 87M
 -,===,oo<  7/8  98% Computing list of interesting correlations of malware | 87M
 -,===,oo=  7/8   1% Computing list of interesting correlations of malware | 87M
Regular min percentage: 94
Malware min percentage: 94
Regular weight : 700
Malware weight : 7
```

- Property extraction: 17 seconds
- Classification: 4 seconds

What About the SAP Android App? (Cont.)



SHERLOCKDROID ANALYSIS

Filename : b1.mobile.android.apk_38ccc092.apk
SHA256 : 2890685023c151dc35ee98d7cd4dd93772531e69a38bad0a505d28f98dda971f

Alligator

This sample looks clean
Analysis time: 318 seconds

→ 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 (Actually: 10!)
- 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 Command&Control 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 !