



# SherlockDroid, an Inspector for Android Marketplaces

Axelle Apvrille - FortiGuard Labs, Fortinet  
Ludovic Apvrille - Institut Mines-Telecom,  
Telecom ParisTech, LTCI CNRS

Hack.Lu, Luxembourg  
October 2014

**FORTINET**®

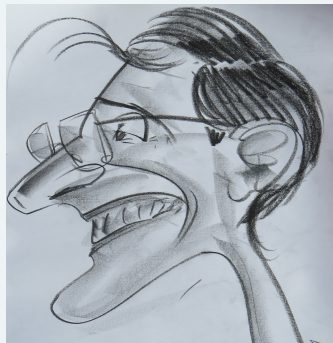


# Who are we?

Axelle



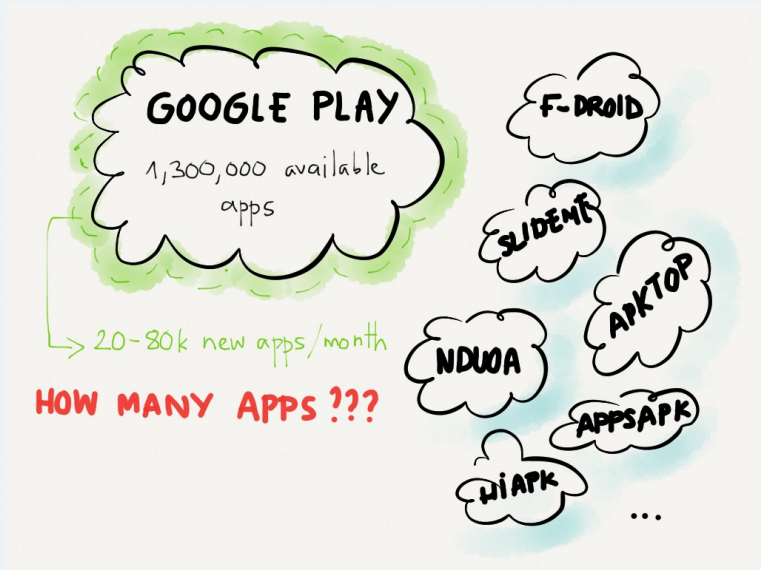
Ludovic



A hand-drawn cloud graphic with a black outline and a light blue watercolor-like background. Inside the cloud, the text "GOOGLE PLAY" is written in bold, uppercase letters. Below it, the text "1,300,000 available apps" is written in a smaller, lowercase font.

**GOOGLE PLAY**  
1,300,000 available  
apps

# Many Android Applications







**We don't know exactly  
how many apps there are**



## **We don't know exactly how many apps there are**

- ▶ Precise number of Android marketplaces????
- ▶ How many duplicate apps?
- ▶ How many old/retired apps?



## **We don't know exactly how many apps there are**

- ▶ Precise number of Android marketplaces????
- ▶ How many duplicate apps?
- ▶ How many old/retired apps?

**but it's BIG NUMBERS**

**We don't know...** (exactly)

**We don't know...** (exactly)

## What we **do** know

- ▶ Oct 2014. **840k** malicious Android **samples**

**We don't know...** (exactly)

## What we **do** know

- ▶ Oct 2014. **840k** malicious Android **samples**
- ▶ **1,000+** new malicious **Android sample every day**

# Known Malware

I'M A MALWARE  
e.g I pose as a Skype installer



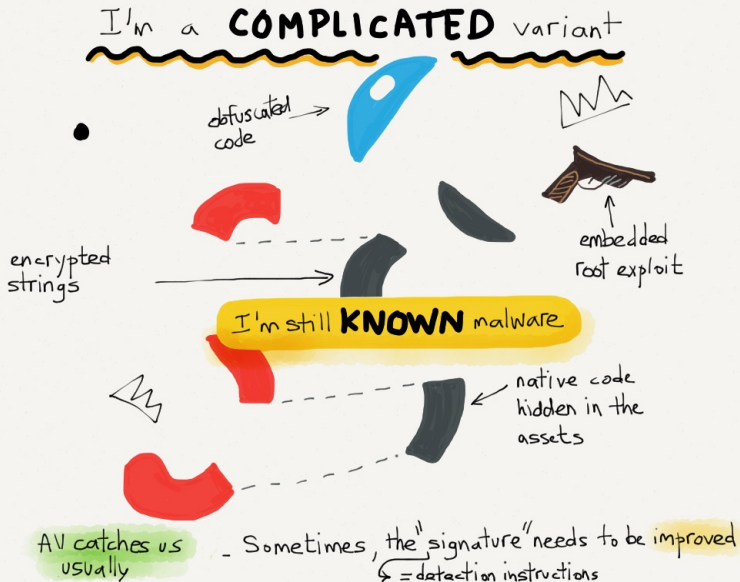
We are **KNOWN MALWARE**  
and it's usually easy to  
detect us



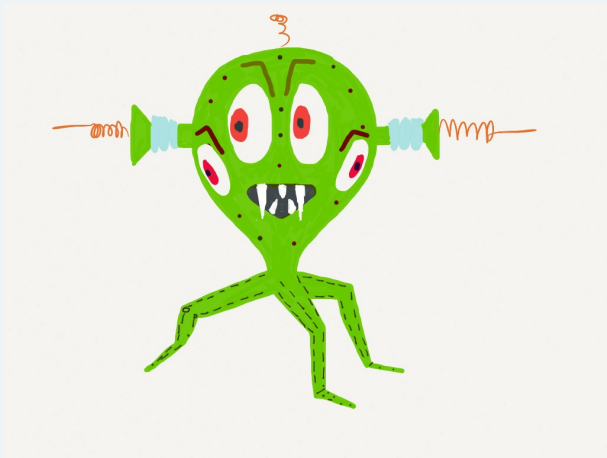
I'm a  
minor  
variant  
"Trust me, I  
install Opera  
Mini"



I'm  
another  
minor  
variant  
e.g I'm just repackaged  
(like my new dress?)







**Do they exist? YES**



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

<b>Name</b>	<b>Creation date</b>	<b>Detection date</b>
Android/Wroba	June 16	June 21 <b>+5d</b>
Android/Curesec	July 3	July 11 <b>+8d</b>
Android/ScarePackage	July 13	July 24 <b>+11d</b>



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

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

# So, What Are We Interested In?



KNOWN (and detected)  
MALWARE



NOTHING TO DO  
(BORING...)



MINOR VARIANT  
(usually detected)



NOT VERY  
INTERESTING  
FOR RESEARCH



MAJOR VARIANT



**AH HA?!**



UNKNOWN MALWARE



  
**TOP  
INTEREST**  
RESEARCH/PR  
INNOVATION

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

**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**



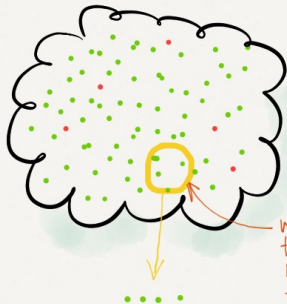
**Crawl Android marketplaces**

**Spot suspicious apps**

**Focus on major variants and unknown malware**

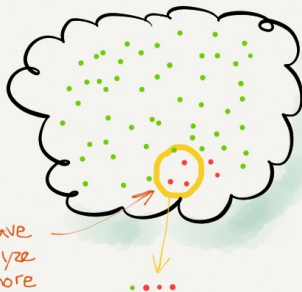
# SherlockDroid (Unbiased) Benefits

**WITHOUT**  
SHERLOCKDROID



WE WASTE OUR PRECIOUS  
TIME ON CLEAN SAMPLES  
(and usually don't have time  
to find nasty samples)

**WITH**  
SHERLOCKDROID



HIGHER CHANCES TO  
SPOT INTERESTING  
MALWARE  
(it can't be perfect, though)

we don't have  
time to analyze  
manually more  
than this



**It is not an AV scanner**  
because SherlockDroid does not handle known  
malware / minor variants



**It is not an AV scanner**  
because SherlockDroid does not handle known  
malware / minor variants

**We will miss some malware**  
we're not perfect :(

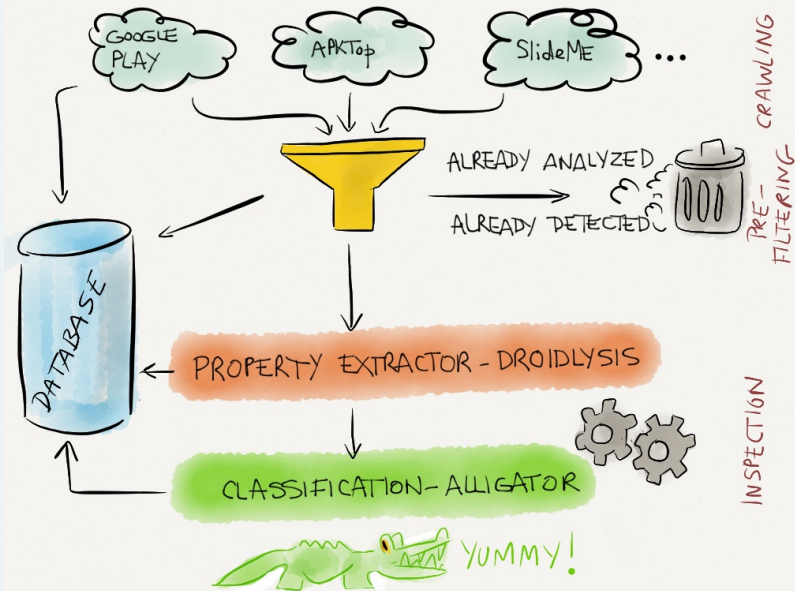
### **It is not an AV scanner**

because SherlockDroid does not handle known  
malware / minor variants

### **We will miss some malware**

we're not perfect :(  
*but we would have missed them without  
SherlockDroid too*

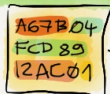
# SherlockDroid Architecture



## 7 CRAWLERS



## 289 PROPERTIES



ADAPT TO YOUR OWN INFRASTRUCTURE



We can share doc + this paper!  
+ a few examples

No release though because   
will follow/adapt  
→ more work for us!



-, ==, 00<  
alligator

→ <http://perso.telecom-paristech.fr/~apvrille/alligator.html>

**OPEN-SOURCE / FREE SOFTWARE**

SherlockDroid is currently in 'heavy testing' phase

SherlockDroid is currently in 'heavy testing' phase

**Crawled 140 K samples**

SherlockDroid is currently in 'heavy testing' phase

**Crawled 140 K samples**

**Extracted properties of 550 K + samples**



SherlockDroid is currently in 'heavy testing' phase

**Crawled 140 K samples**

**Extracted properties of 550 K + samples**

**Learning and classification: 480 K clusters!**  
**At 50 K, FP: 0.99%, FN: 3.3%**

## 8 new unknown malware and Potentially Unwanted Apps

## **8** new unknown malware and Potentially Unwanted Apps

Okay, we would have preferred only nasty malware



## 8 new unknown malware and Potentially Unwanted Apps

Okay, we would have preferred only nasty malware

*Do you know any other framework who identified  
real unknown malware?*

## 8 new unknown malware and Potentially Unwanted Apps

Okay, we would have preferred only nasty malware

*Do you know any other framework who identified  
real unknown malware?*

**Answer: DroidRanger: 2**

## 8 new unknown malware and Potentially Unwanted Apps

Okay, we would have preferred only nasty malware

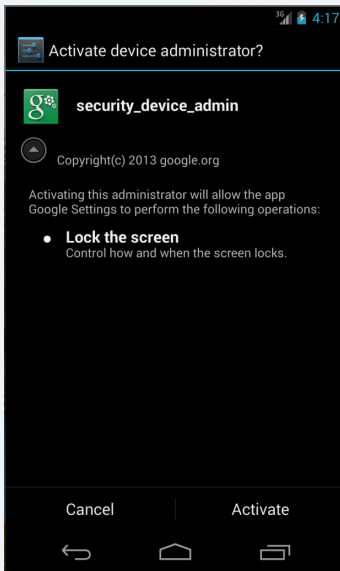
*Do you know any other framework who identified  
real unknown malware?*

**Answer: DroidRanger: 2**

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

- ▶ [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](#)

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



## Android/MisoSms.A!tr.spv

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



```
▼ 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/Geysers!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"



## 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>



**Permissions are good ... but insufficient!**



**Permissions are good ... but insufficient!**

**In Dalvik, every object points to a string**

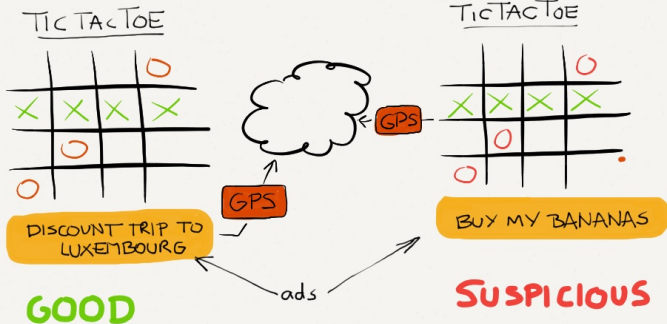


**Permissions are good ... but insufficient!**

**In Dalvik, every object points to a string**

**We also search assets and resources**

## DIFFICULT TO SPOT THE DIFF



The adkit requests  
INTERNET & GPS  
that's frequent  
(whether you like it or not)

The payload app  
requests INTERNET & GPS  
why ???



## Gather clusters for learning - **only once**

Test with 1,514 newer clean samples  
and 3,062 newer malware

Learning cluster size	Learning time	Classification time	FP	FN
50,000	2 hours 13 min	1 min 40 s	<b>0.99%</b>	<b>3.3%</b>

We can favour minimum False Positives



## Gather clusters for learning - **only once**

Test with 1,514 newer clean samples  
and 3,062 newer malware

Learning cluster size	Learning time	Classification time	FP	FN
50,000	2 hours 13 min	1 min 40 s	<b>0.99%</b>	<b>3.3%</b>
<b>480,000</b>	approx. 31 hours	9 min 21 s	1.8%	0.5%

It works with 480 K clusters !  
We can favour minimum False Positives





## Gather clusters for learning - **only once**

Test with 1,514 newer clean samples  
and 3,062 newer malware

Learning cluster size	Learning time	Classification time	FP	FN
50,000	2 hours 13 min	1 min 40 s	<b>0.99%</b>	<b>3.3%</b>
<b>480,000</b>	approx. 31 hours	9 min 21 s	1.8%	0.5%

SVM? Far worse! 50 K: **FP: 5.48%** FN: 0.65% !!!

It works with 480 K clusters !

We can favour minimum False Positives



Alligator unleashed!  
Wake up!

# DEMO - SherlockDroid GUI [Preview]

The screenshot shows a web browser window titled "SherlockDroid Web Front-End". The address bar shows "babylon/sherlockDroid/web/index.html". The browser's bookmark bar includes "Apps", "FortiCare", "Search Files", "FT CrossScan", "FortSignature", "Sample Tracker 2", "FortWiki", "Fortinet Google D...", and "Other bookmarks".

The main content area features a control panel with the following elements:

- A horizontal row of six cloud-shaped buttons: "GOOGLE PLAY Start", "SLIDESHOW Stop", "APKTOP Stop", "APPSMGR Start", "RECURSIVE CRAWLER Start", and "NDROID CRAWLER Start".
- A vertical stack of three green rectangular buttons: "INFINITE PRI Stop", "INFINITE INS Stop", and "LAUNCH ALL Stop".
- A text log at the bottom with the following content:

```
Sherlock's inspector is analyzing pre-filtered samples
Currently no more file to process
Sherlock's inspector is analyzing pre-filtered samples
Currently no more file to process
Sherlock's inspector is analyzing pre-filtered samples
Currently no more file to process
Sherlock's inspector is analyzing pre-filtered samples
Currently no more file to process
Sherlock's inspector is analyzing pre-filtered samples
Currently no more file to process
```

A vertical label on the left side of the browser window reads "SherlockDroid Web Front-End - Chromium".

# DEMO - SherlockDroid's Database

## Sample recently crawled, to pre-filter

```
115118|f8ef5f5306fb7...|net.mnprogram.mnagenda.apk|  
Google Play|0|0|toanalyze||2014/10/09-14:04|967041
```

## Known malware

```
114902|ae084007fab965f829ba3fc...|  
JJLord.30103.30000.visible.apk||0|0|  
|detected|Android/SMSreg.AK, SIGID: 49829716, VID: 5236396||0
```

## Unknown sample, to be inspected

```
115117|4dd15425c67b744125d7386...|  
com.apalusa.lavoz.AgendaVos.apk|Google Play|0|0|  
|toanalyze||2014/10/09-14:04|2342643
```

## Unknown sample probably clean

```
115072|be849297862a50d7116d7a6be0...|  
com.covertapps.joomlaadminmobilelite.apk|Google Play  
|248.974979321754|145.030471289058|done||2014/10/09-13:48|583248
```

## Example of suspicious samples

```
$ ./suspiciousApk.pl
suspiciousApk - show which samples are currently found suspicious by AI
Suspect: com.indvseng.indCENSORED.apk (f178c77d...
  origin: Google Play
  scoreRegular: 153.974979321754 scoreMalware: 161.923639714817
  difference: 7.94866039306393
-----
Suspect: floating-toucCENSORED.apk (3162b0c...
  origin: http://link.appsapk.com/downlo...
  scoreRegular: 153.974979321754 scoreMalware: 164.390159536531
  difference: 10.4151802147771
-----
Suspect: com.Ninjastrike456.ninjastrike.apk (65bb4...
  origin: Google Play
  scoreRegular: 153.621310611974 scoreMalware: 169.818181818182
  difference: 16.1968712062074
-----
Found 3 suspects
--- END
```



## Size of clusters

```
$ wc -l learn-malware.csv learn-clean.csv  
guess-malware.csv guess-clean.csv  
486890 learn-malware.csv  
12368 learn-clean.csv  
3062 guess-malware.csv  
1514 guess-clean.csv  
503834 total
```

## Size of clusters

```
105A663E.var,0.166667,0.000800,0,0.001930,0.000100,  
0,0,0,0,0,0,0,0,0,1,0,0.201400,1,1,unknown,unknown,  
0.000020,0,0.015000,0,0.000020,0.000010,0,0,0,0,0,1,  
0,0,0,0,1,0,0,0,1,1,0,1,0,0,0,unknown,0,0,1,0,0,0,  
0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,1,  
1,1,1,0,1,1,1,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,1,0,1,0,0,0,0,1,1,0,0,0,1,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,1,0,0,0,0,0.000010,0,0,0,0,0,0,0,0,0,  
...
```

- ▶ Mostly boolean values (0, 1) + 'unknown'
- ▶ Integer values have been normalized to fit in [0,1]

# DEMO: Example of Learning Script

## The Alligator Language

```
setprintintermediatescore

printClusterSummary regular
printClusterSummary malware
printClusterSummary guess

setMultiplierWeight regular 0-100,1
setMultiplierWeight malware 0-100,1
compute inversedeviation

setPropertyWeightsFromColumn 63 6
...
setMultiplierWeight regular 0-100,1
setMultiplierWeight malware 0-100,1
compute inverseweightdeviation

setMultiplierWeight regular 0-100,1
setMultiplierWeight malware 0-100,1
compute degressiveproximity 5

...
```



Work! Work!

```
Alligator Daemon: AnaLyzing maLware wIth partitioninG and probAbility-b  
Rithms Daemon
```

```
-,===,oo< Alligator: (C) Institut Mines Telecom / Telecom ParisTech L  
VRILLE, ludovic.apvrille@telecom-paristech.fr
```

```
-,===,oo< http://perso.telecom-paristech.fr/~apvrille/alligator.html
```

```
-,===,oo< Alligator is released under a CECILL License. See http://www.nfo/index.en.html
```

```
-,===,oo< Enjoy!!!
```

```
*** Your Alligator version is: 0.3-beta1 -- build: 1433 date: 2014/10/0  
4 CET ***
```

```
-,===,oo< 1/7 0% unknown | 85MB/1820MB
```

```
-,===,oo< 2/7 0% unknown | 97MB/1820MB
```

```
-,===,oo= 2/7 0% unknown | 97MB/1820MB
```

```
-,===,oo< 2/7 0% unknown | 97MB/1820MB
```

```
-,===,oo= 2/7 0% unknown | 97MB/1820MB
```

```
...
```

## Classifying samples

```
*** Overall report of guess ***
```

```
Classification time:468.121s
```

```
** Overall results **
```

```
regular - 11249 elements in cluster, nb of properties: 288
```

```
malware - 50000 elements in cluster, nb of properties: 288
```

```
guess - 3 elements in cluster, nb of properties: 288
```

```
Results summary:
```

```
2 regular(s) found, 1 malware(s) found in guess
```

```
Percentage of regular: 66.66666666666666
```

```
Percentage of malware: 33.33333333333333
```

```
regular: Light:2 (66.67%) Medium:0 (0.00%) Strong:0 (0.00%)
```

```
malware: Light:1 (33.33%) Medium:0 (0.00%) Strong:0 (0.00%)
```

```
105A663E.var: regular (regular:131.36352883261992, malware:121.909090
```

```
...
```

## Contact info

SherlockDroid: aapvrille at fortinet dot com

Alligator: ludovic dot apvrille at telecom minus paristech dot com

## Downloads

### Alligator Release

L. Apvrille, A. Apvrille, *Pre-filtering Mobile Malware with Heuristic Techniques*, GreHack 2013

A. Apvrille, T. Strazzere, *Reducing the Window of Opportunity for Android Malware*, EICAR 2012

Powerpoint slides? No way! This is L<sup>A</sup>T<sub>E</sub>X- Beamer !