On Mobile Malware Infections

N. Asokan

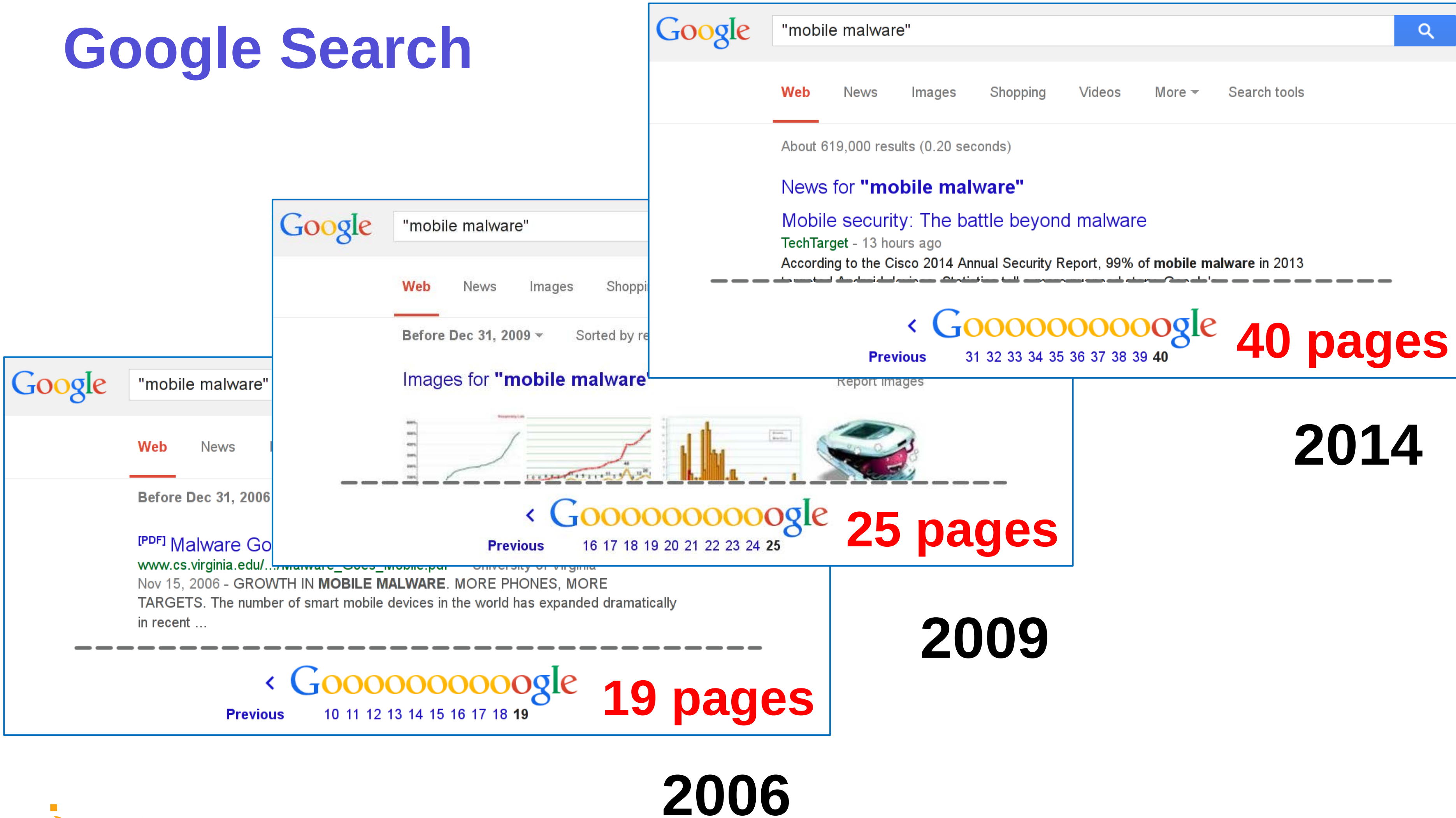
(joint work with Hien Thi Thu Truong, Eemil Lagerspetz, Petteri Nurmi, Adam J. Oliner, Sasu Tarkoma, Sourav Bhattacharya)

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Aalto University

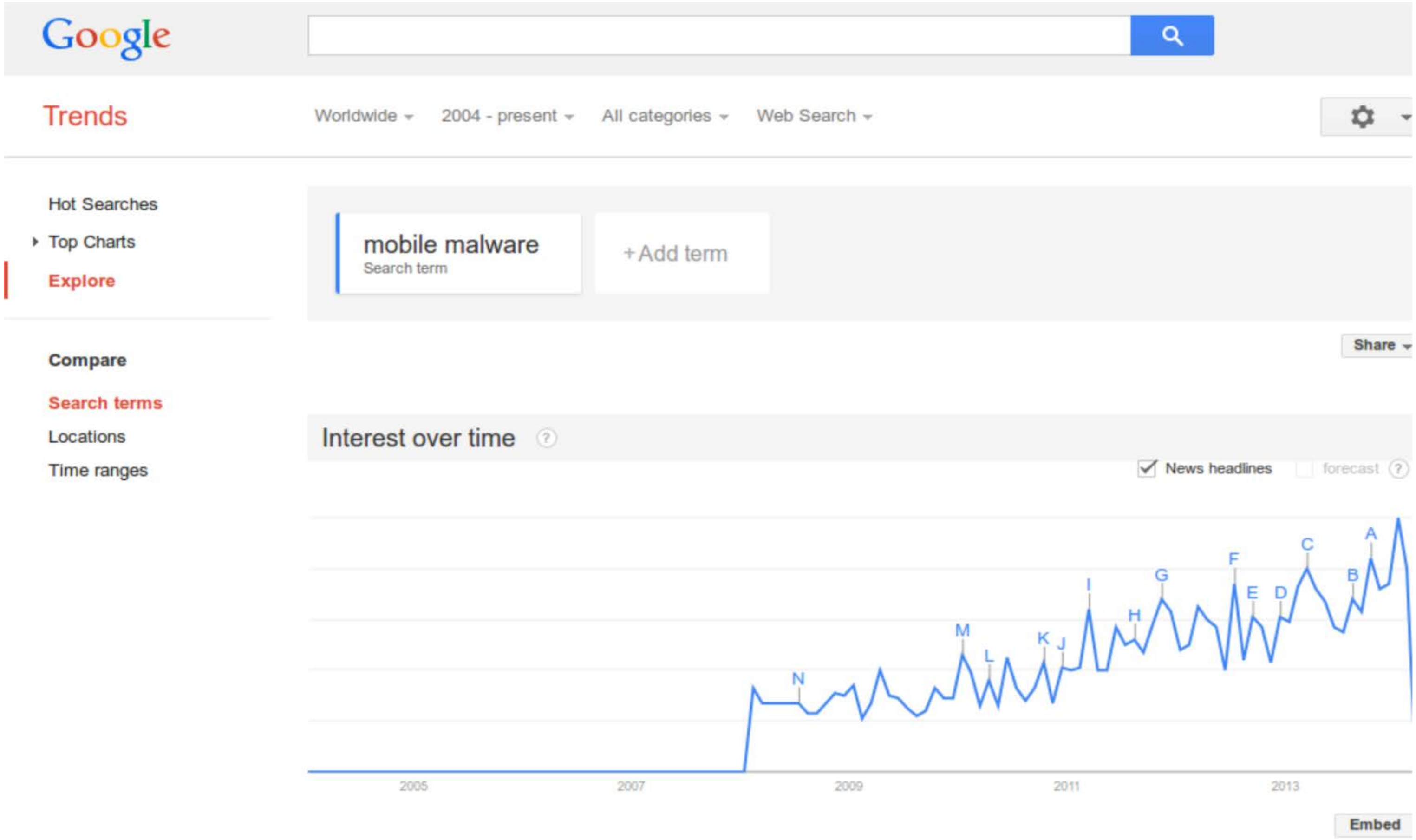
Mobile malware alarm bells

Google Search



Mobile malware alarm bells

Google Trends



Research focus: analysis of malware

Google Scholar

100 pages

How prevalent is mobile malware?

domains. We make several important observations. The mobile malware found by the research community thus far appears in a minuscule number of devices in the network: 3,492 out of over 380 million (less than 0.0009%) observed during the course of our analysis.

The Core of the Matter: Analyzing Malicious Traffic in Cellular Carriers

Charles Lever
Georgia Institute of Technology
chazlever@gatech.edu

Manos Antonakakis
Damballa
manos@damballa.com

Brad Reaves
Georgia Institute of Technology
brad.reaves@gatech.edu

Patrick Traynor
Georgia Institute of Technology
traynor@cc.gatech.edu

Wenke Lee
Georgia Institute of Technology
wenke@cc.gatech.edu

NDSS 2013

?

Lookout

TIME Tech

SECTIONS

SECURITY

Study: 32.8 Million Android Phones Infected with Malware

By Techlicious / Fox Van Allen | April 17, 2013 | 9 Comments

Do you have an anti-virus app on your Android phone yet? If not, a new study conducted by security firm NQ Mobile suggests you're playing with fire: The number of malware threats to your Android phone has increased 163% over the past year alone.

The study, which looked at over 5.3 million apps available in 406 different online stores, identified 65,227 different pieces of potentially dangerous malware last year. A quick look at the trend suggests that malware is growing at an exponential rate – there were only 1,649 such malware discoveries in 2009.

In total, 32.8 million Android phones were infected with malware in 2012 – more than triple the number of the year before. The majority of these infections involve spyware or adware, while about a quarter are designed to steal and profit off of your personal data. A smaller minority is designed to make your phone permanently unusable, something we'd all no doubt like to

Email Print Share

Follow @techland

Outline

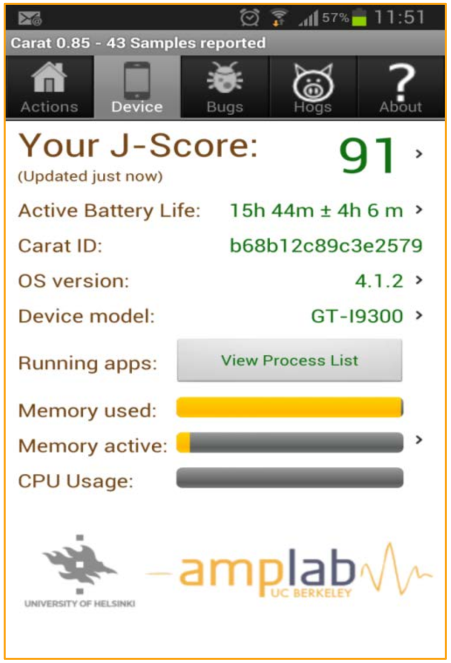
- Gather data directly from devices
- Accurately estimate malware infection rate
- Identify risk factors, cheaply

Outline

- **Gather data directly from devices**
- Accurately estimate malware infection rate
- Identify risk factors, cheaply

Gather data directly from devices

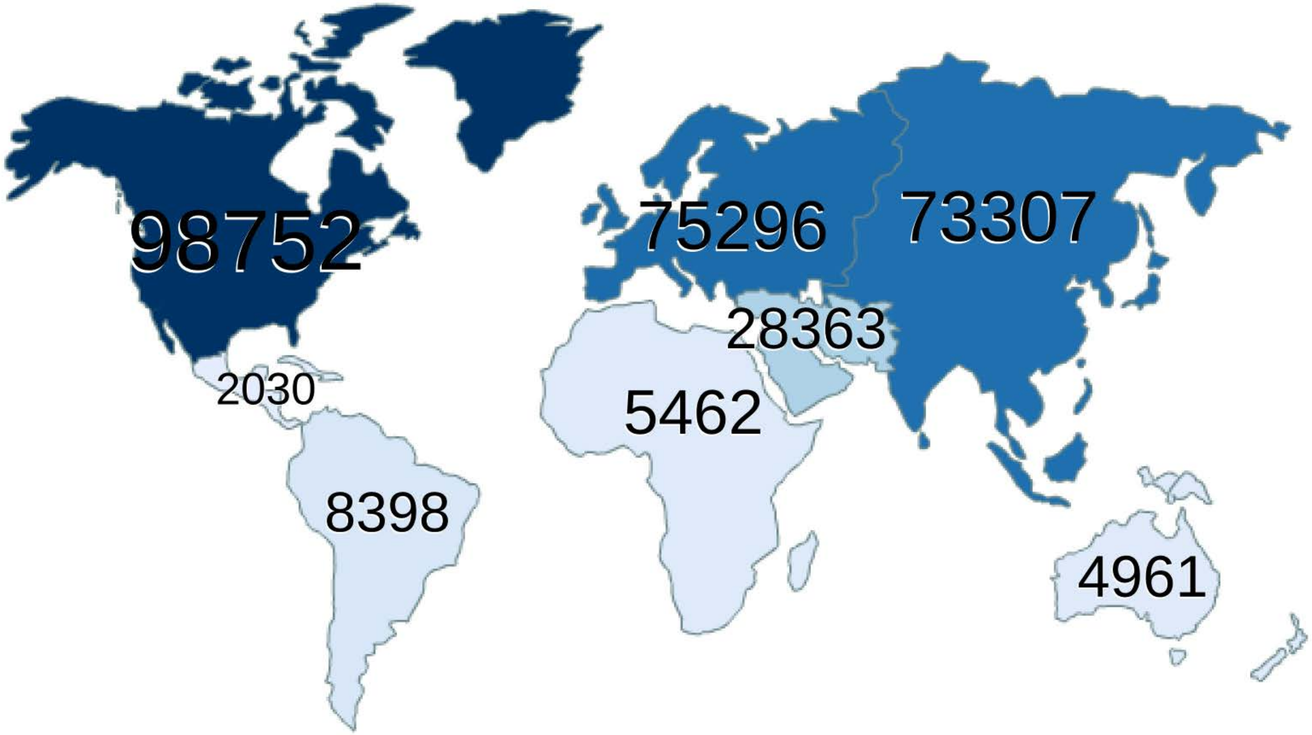
Piggyback on a popular package



Need to be lightweight and unobtrusive

http://carat.cs.berkeley.edu

Carat (devices by continents)



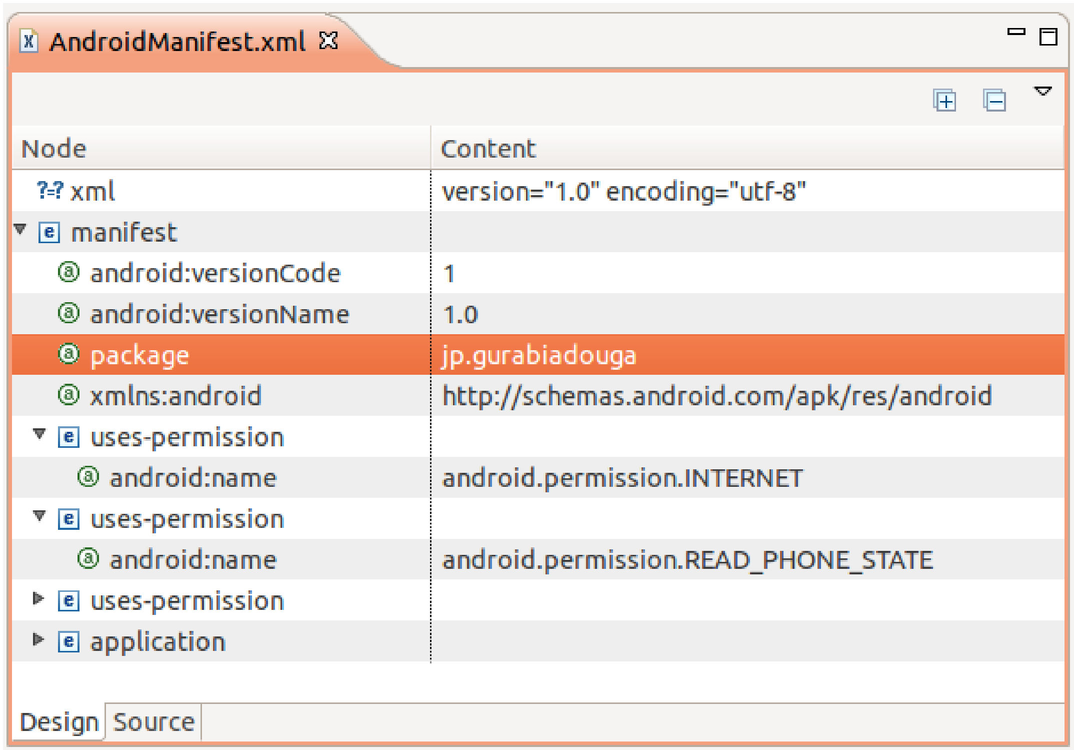
Android devices: geography distribution, (April 2, 2014)

http://carat.cs.berkeley.edu

What kind of data?

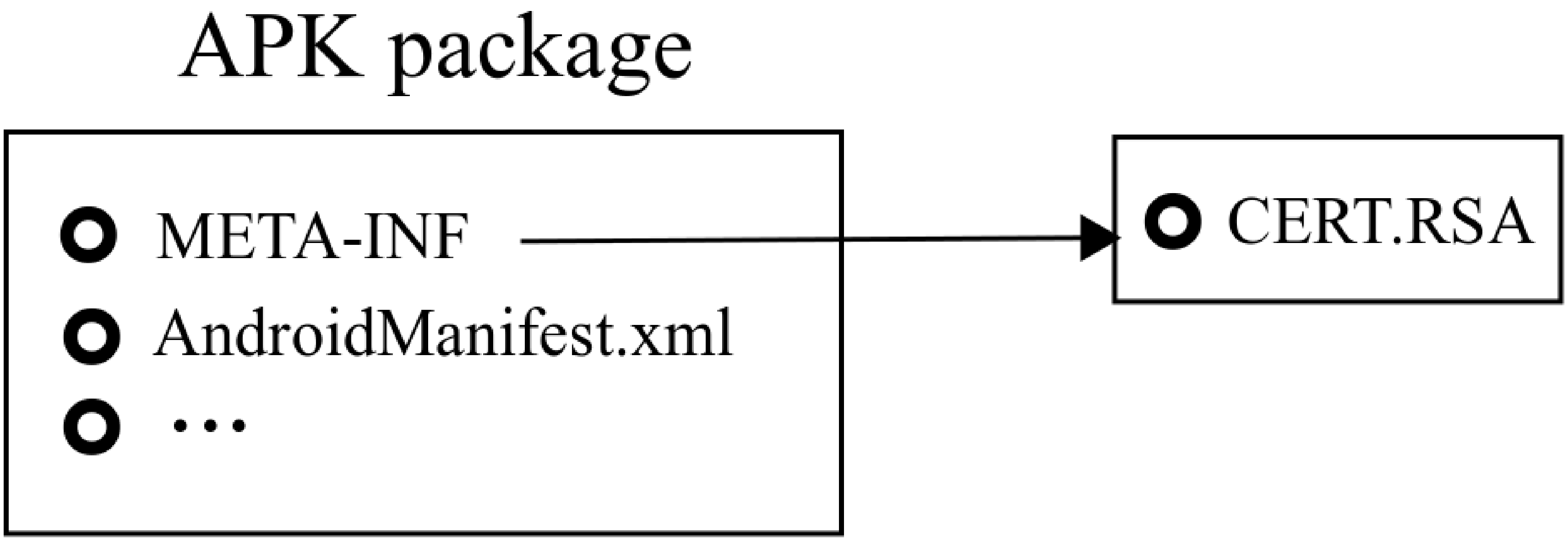
- How to estimate infection rate?
  - Identify a package on device; check for match with known malware
- How to identify an Android package?

Structure of an Android Package



<package, versionCode> tuples (<p,v>) should be unique but not enforced

Structure of an Android Package



Packages are (self-)signed by developers.
Developer certs (dc) are statistically unique.

Identifying a (malicious) package

- **Coarse-grained:**

  Use <developerCert> only

  - <dc> for short
  - upper bound for infections

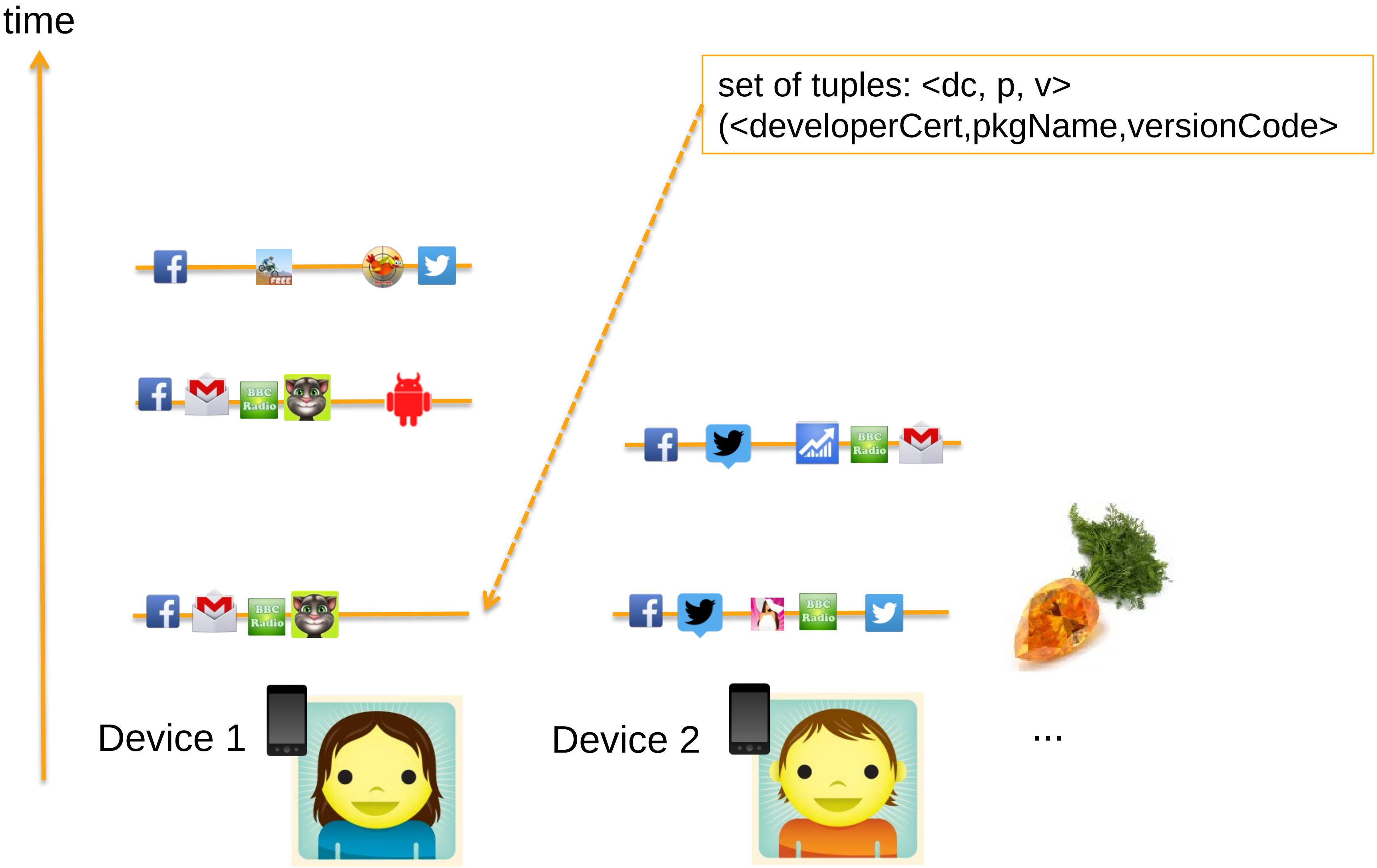
- **Fine-grained:**

  Use <developerCert, package, versionCode>

  - <dc, p, v> for short
  - lower bound for infections

Data

Carat dataset



Carat dataset

Mar 2013 – May 2014

Type	Count
Distinct devices	**99,414**
Unique developer certificates <dc>	**108,482**
Unique <dc, p, v> tuples	**512,342**

Malware datasets

Type	Mobile Sandbox	McAfee	Malware Genome	Total
Unique devcerts <dc>	3,879	1,456	136	**4,809**
Unique packages <dc, p, v>	16,743	3,182	1039	**19,094**
Unique package (.apk) files	96,500	5,935	1260	**103,695**

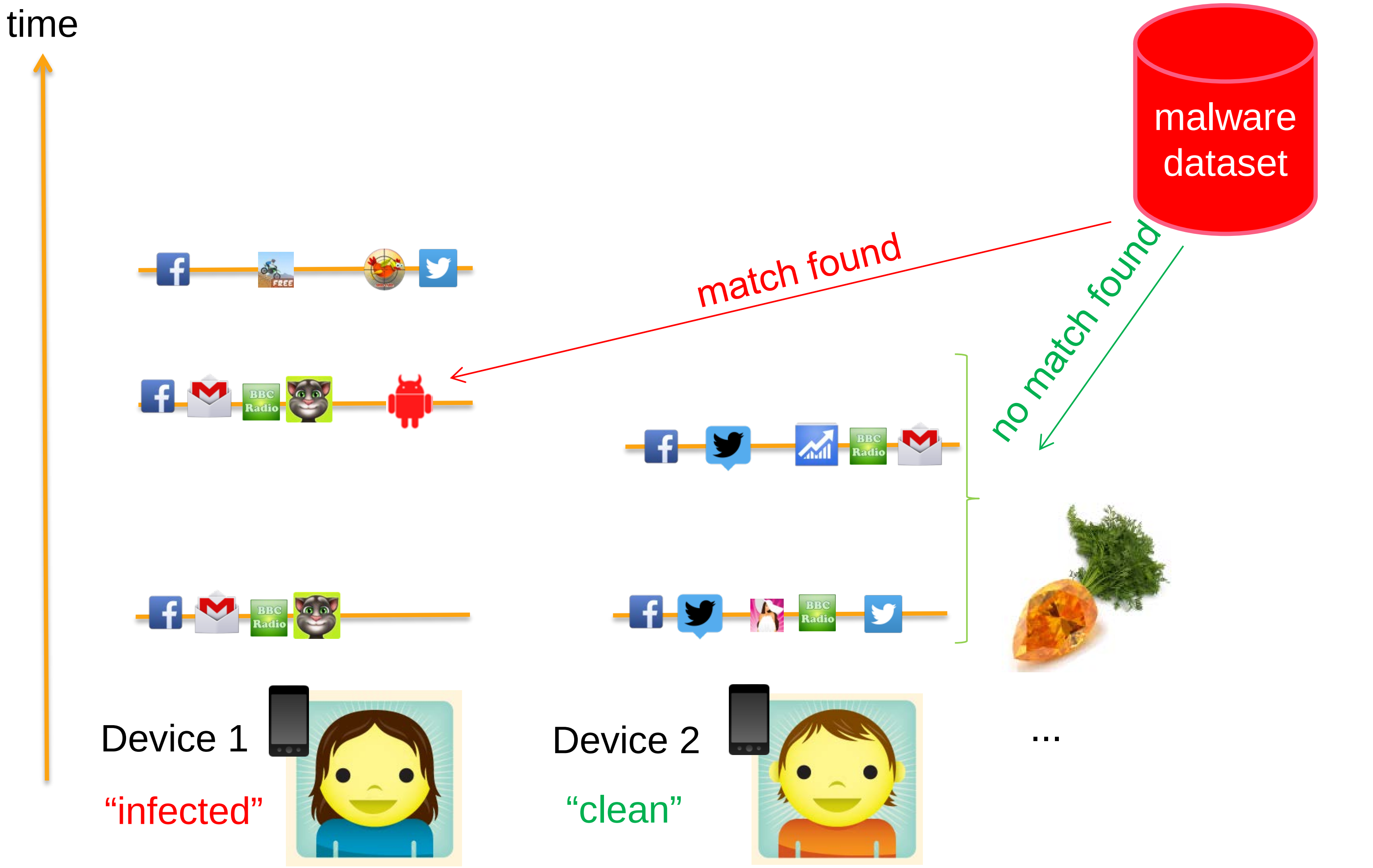
http://mobilesandbox.org/
http://mcafee.com
http://www.malgenomeproject.org/

Outline

- Gather data directly from devices
- **Accurately estimate malware infection rate**
- Identify risk factors, cheaply

Estimates

Carat dataset: identifying infection



Estimates

Incidence of infection

Mar 2013 – May 2014

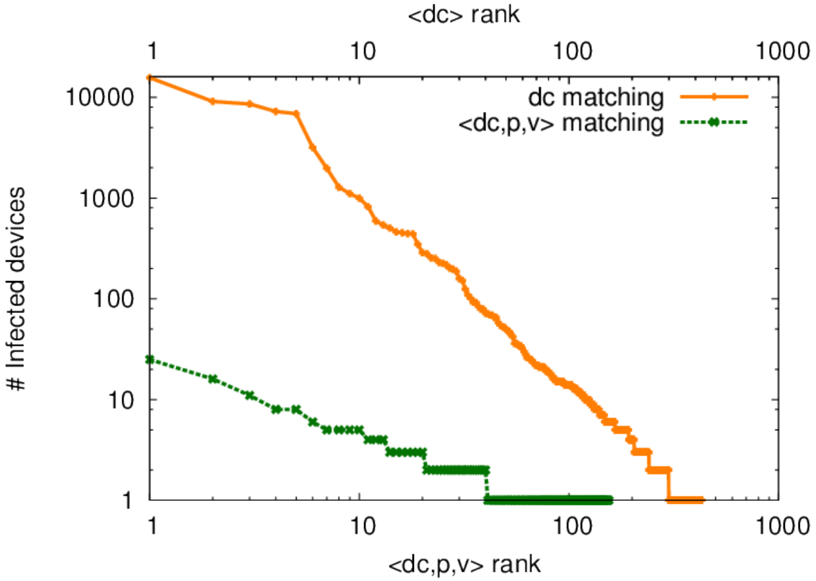
# Infected Devices	Mobile Sandbox	McAfee	Union
coarse-grained: dc match	37,355 (38%)	32,323 (33%)	40,334 (40%)
fine-grained: <dc,p,v> match	263 (0.26%)	255 (0.26%)	477 (0.48%)

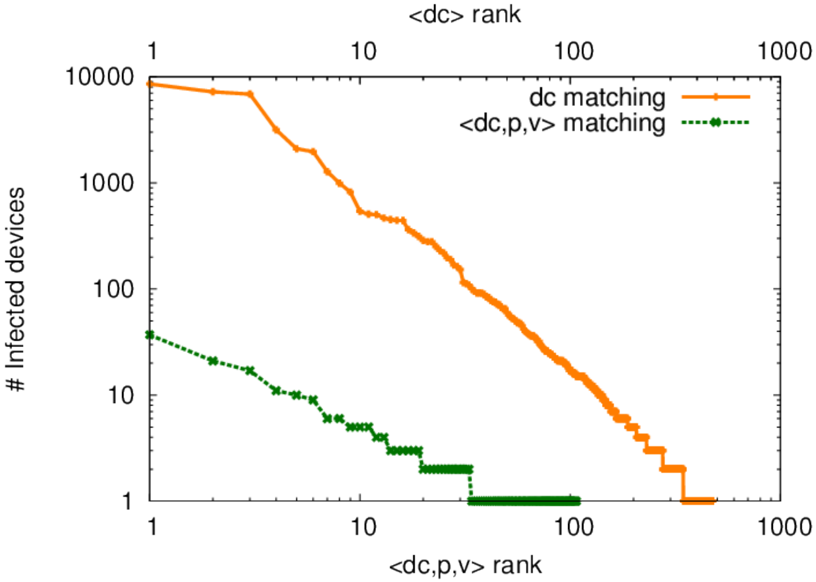
Data collected from 99414 devices over one year

Estimates

Coarse- vs. fine-grained

Mar 2013 – May 2014





Mobile Sandbox

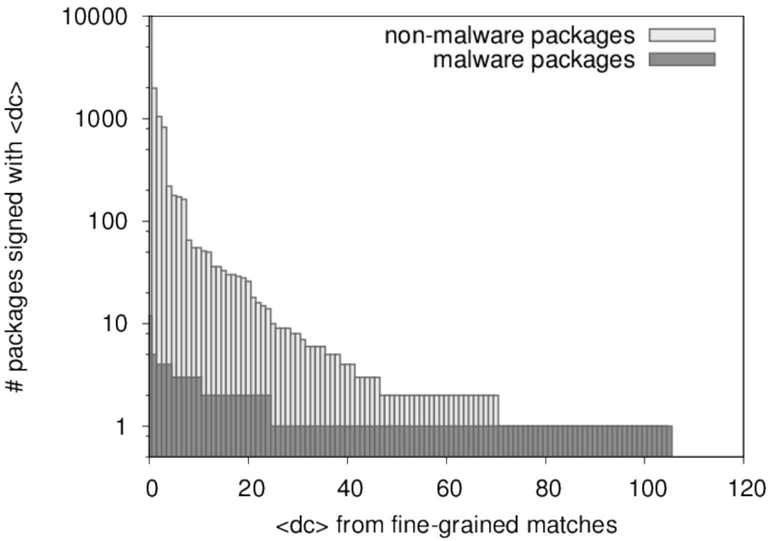
McAfee

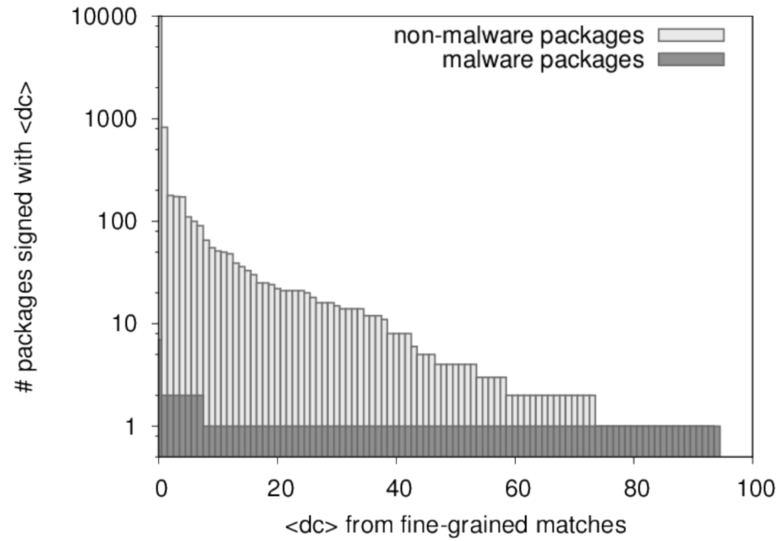
Coarse-grained: ***<dc>*** *matching*
Fine-grained: ***<dc,p,v>*** *matching*
Discrepancy is several orders of magnitude

Estimates

Re-use of signing keys

Mar 2013 – May 2014





Mobile Sandbox

McAfee

Widespread (ab)use of test keys: 544 malwares, 1948 innocuous packages signed with Android Open Source Project (AOSP) test key[1]

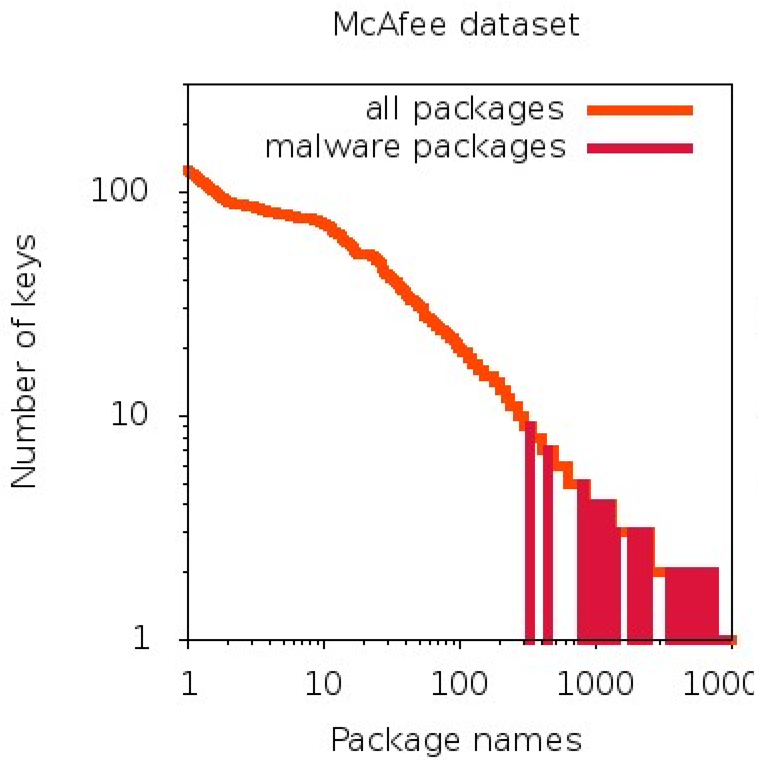
Same key signing malware and non-malware: Brightest Flashlight Free v17 is malware[2], other versions are not.

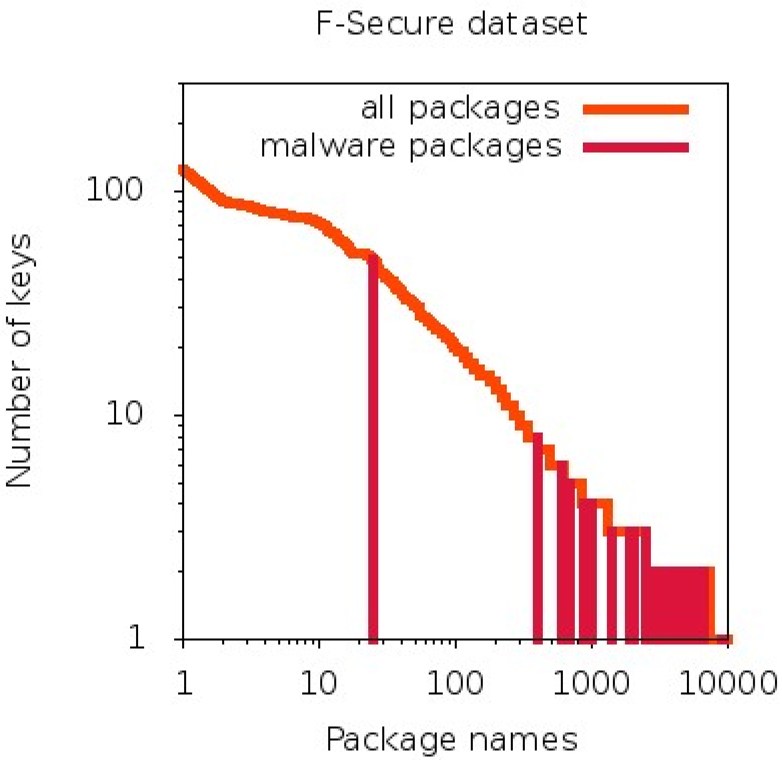
Use fine-grained (<dc,p,v>) matching from now on

1. https://androidobservatory.org/cert/61ED377E85D386A8DFEE6B864BD85B0BFAA5AF81
2. https://androidobservatory.org/cert/27DDACF8860D2857AFC62638C2E5944EA15172D2

Estimates

Rarity of signing keys

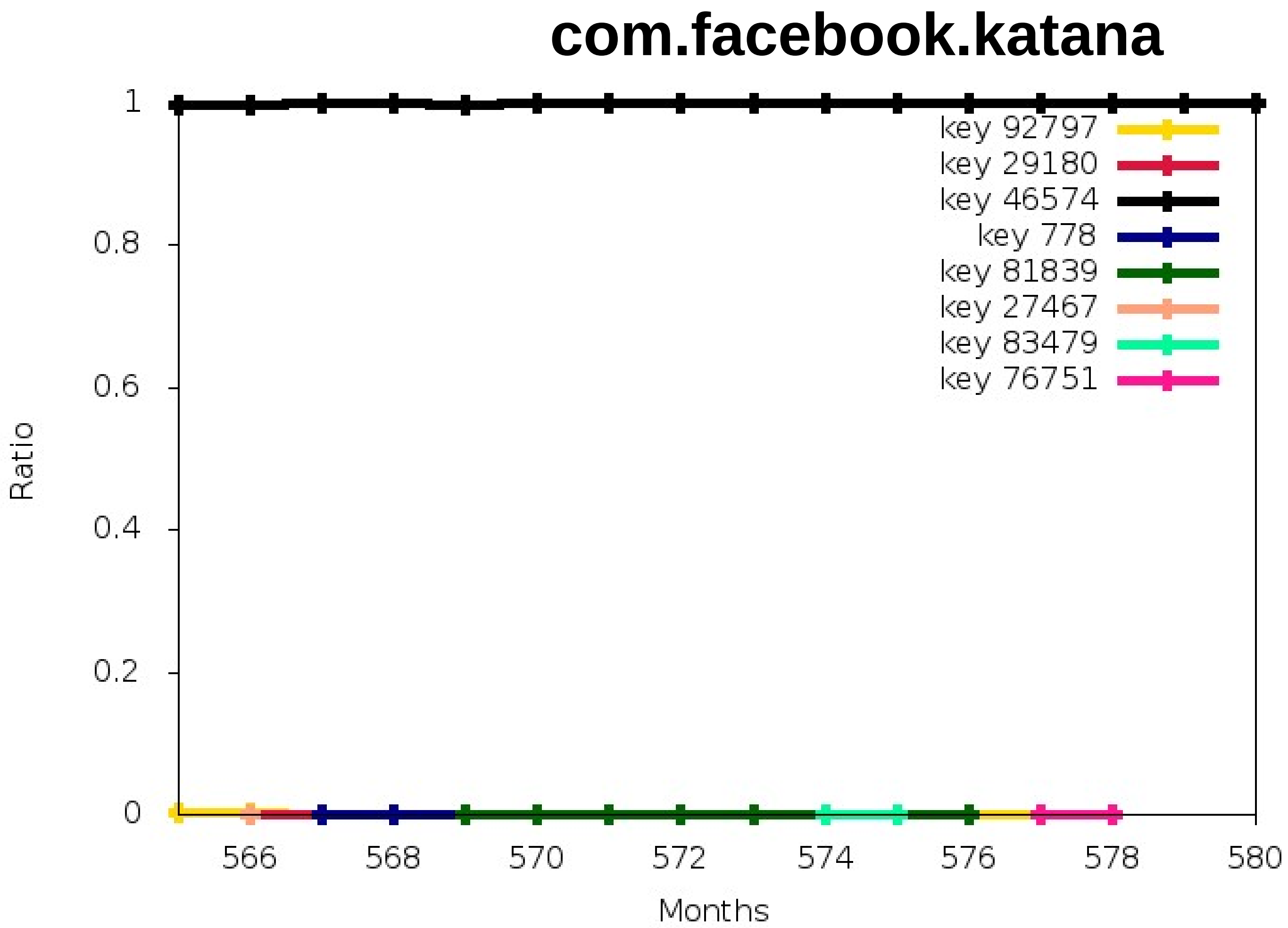




malware <10 keys

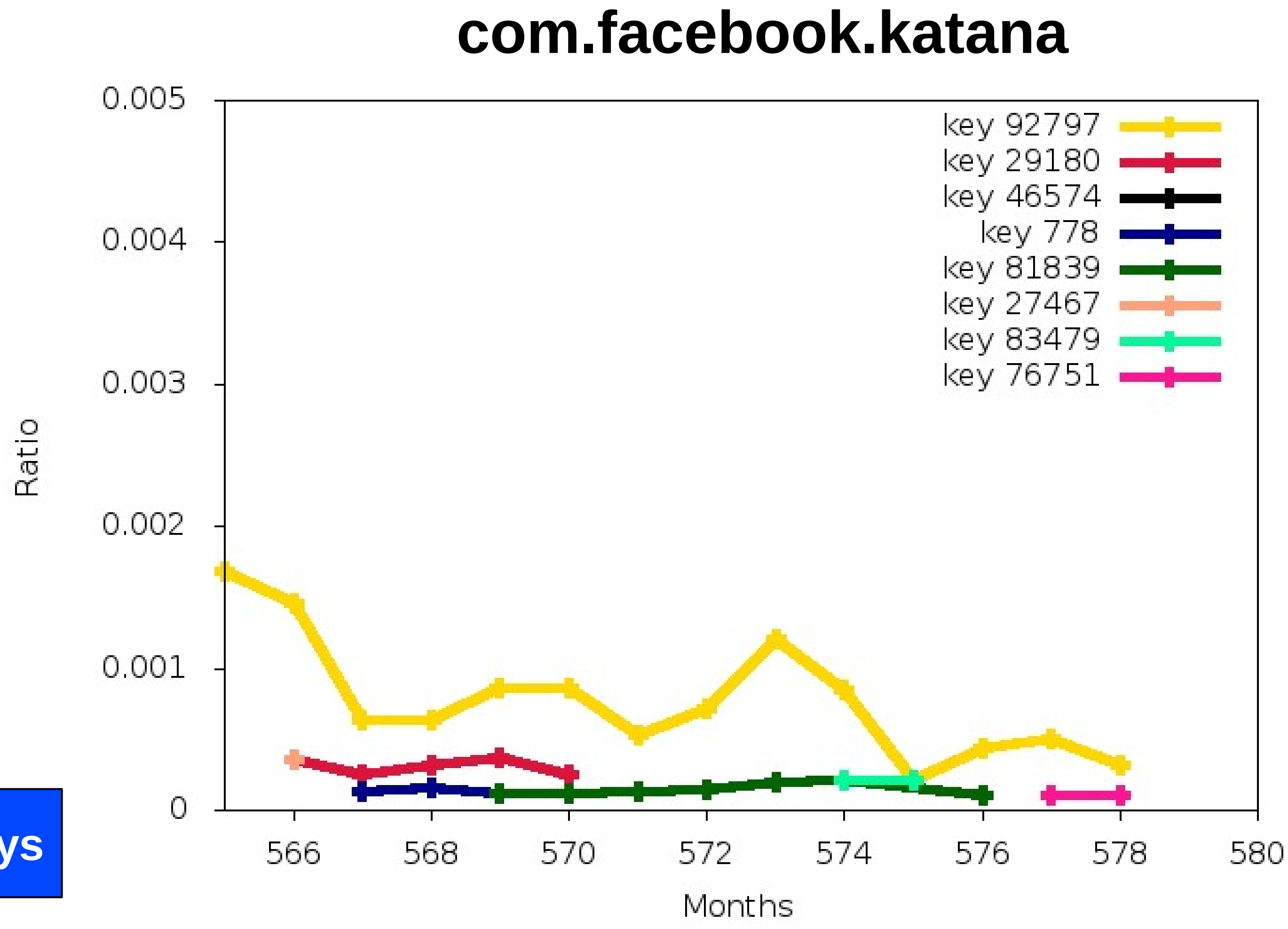
Estimates

Rarity of signing keys: Facebook



Estimates

Rarity of signing keys: Facebook

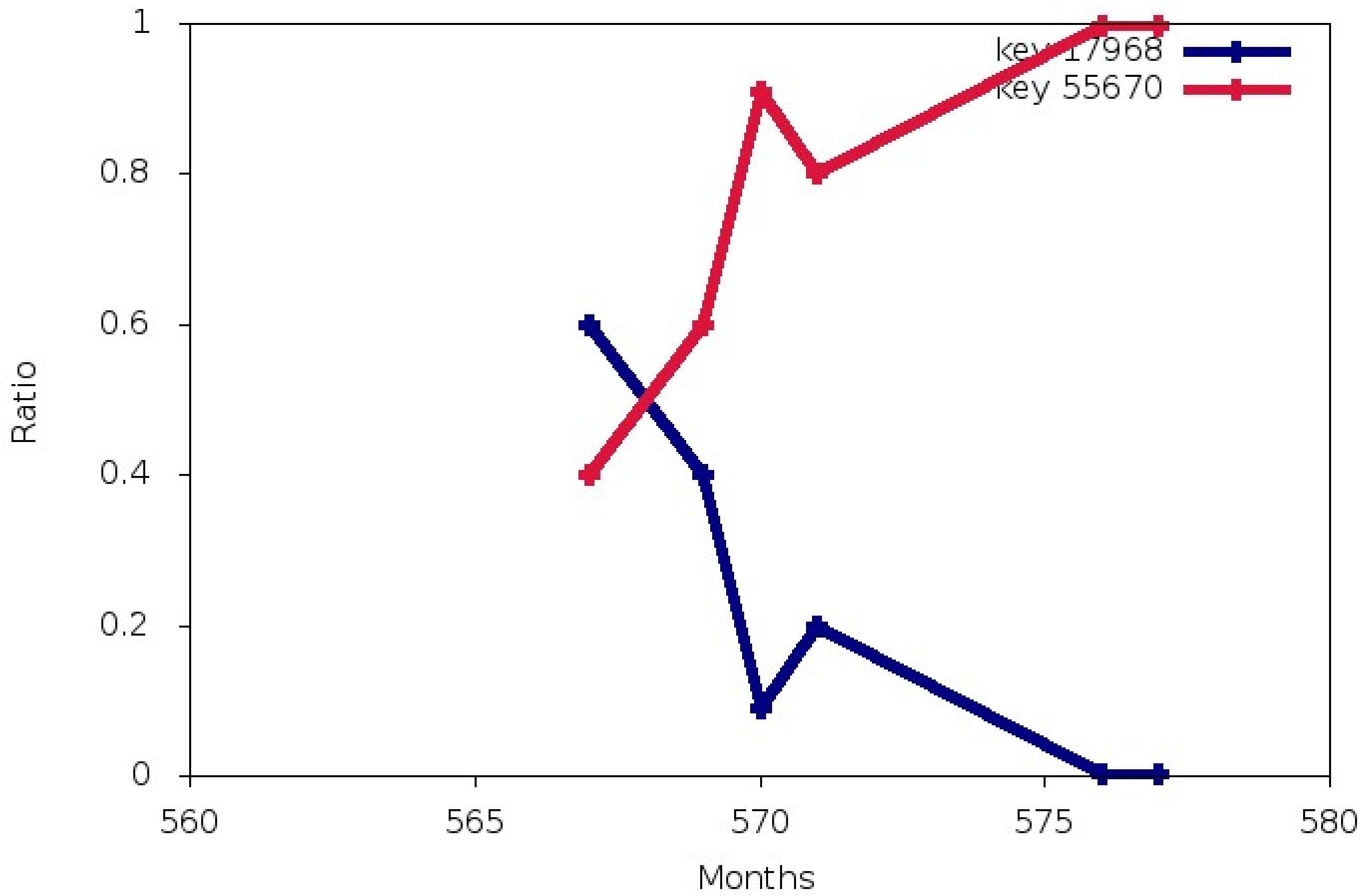


minor keys

Example: package with 2 keys

com.sony.smallapp.app.widget

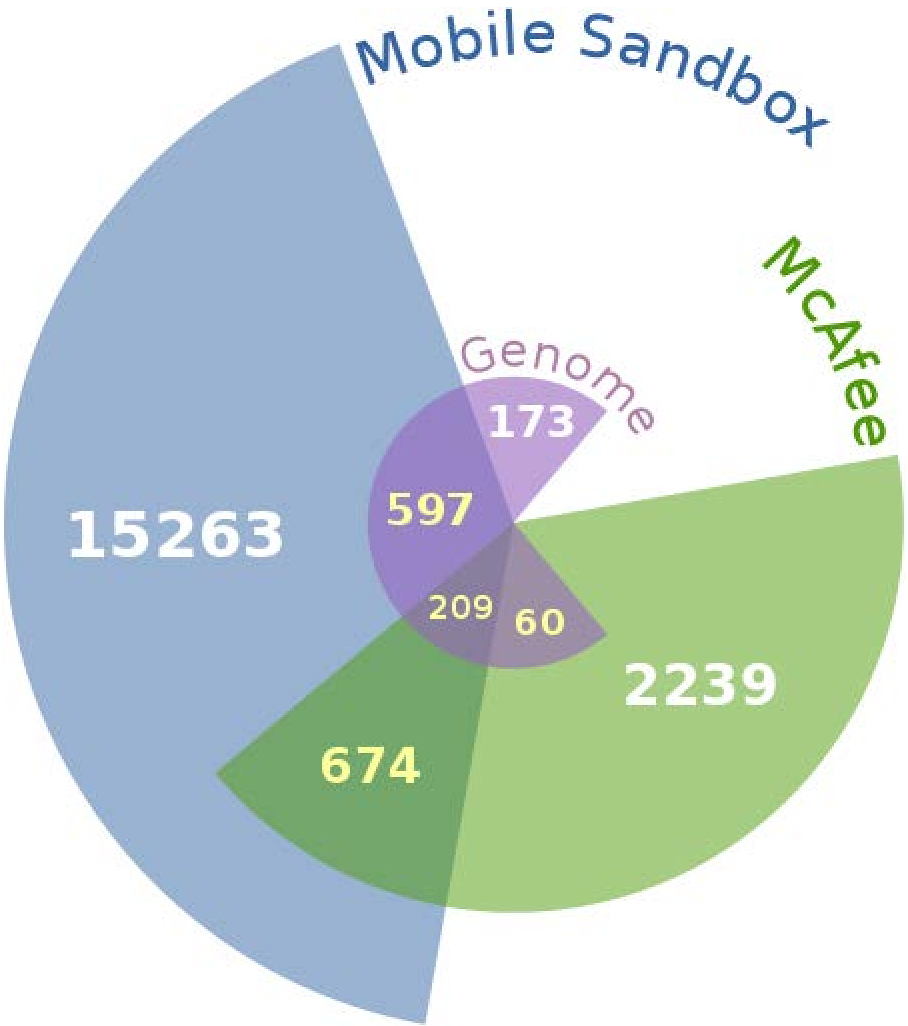
Key change



On-going work: can we use key rarity to identify malware?

Estimates

Malware datasets revisited



Estimates

What is malware?

Number of AV tools flagging this package as malware (Total ~50 AV tools)

lay 2014

Package name	No. Infected devices	Flagged by	Description	Source
it.evilsocket.dsploit	23	22	Monitoring	MC
com.noshufou.android.su		17	Rooting	MC
ty.com.android.SmsService		29	Trojan	MB
com.mixzing.basic		19	Adware	MC
pl.thalion.mobile.battery	10	12	Adware	MC
com.bslapps1.gbc	21	17	Adware	MC
com.android.antidroidtheft	16	17	Monitoring	MB
com.androidlab.gpsfix	7	9	Adware	MC
com.adhapps.QesasElanbiaa	7	18	Adware	MC
download.youtube.downloader.pro7	5	29	Adware	MB
com.android.settings.mt	5	12	Monitoring	MC

Reasons for classification as “malware”

Treat each dataset separately

MC: McAfee
MB: Mobile Sandbox

What is malware?

Curiously, AV vendors do take labeling by other vendors into account!

- Sometimes leads to false positives propagating
- ... and staying undetected!

Estimates

Propagation of False Positives

virustotal

"Google Security Tool", signed by test key

SHA256: 02a51d848315513c27cf6347aeb80f8afe51066a3e80fea95a9bb7eeeeeea431

File name: 02A51D848315513C27CF6347AEB80F8AFE5

Detection ratio: 20 / 46

Analysis date: 2013-02-01 04:13:35 UTC (1 year, 5 months

Analysis | File detail | Additional information

Antivirus	Result
Avast	Android
BitDefender	Android
CAT-QuickHeal	Android
Commtouch	Android
Comodo	Unclass
DrWeb	Android
ESET-NOD32	a variant
Emsisoft	Android

http://bit.ly/IDjR3W

virustotal

Google Security Tool, signed by legitimate Google key

SHA256: 064dcb4d096d9e70db526e342d750169b9d5248d4442e27fd287db904ba3c31f

File name: af311eaa2af1342ba0e39028cf4f99fb.apk

Detection ratio: 13 / 51

Analysis date: 2014-04-17 00:41:32 UTC (2 months, 4 weeks ago)

0 0

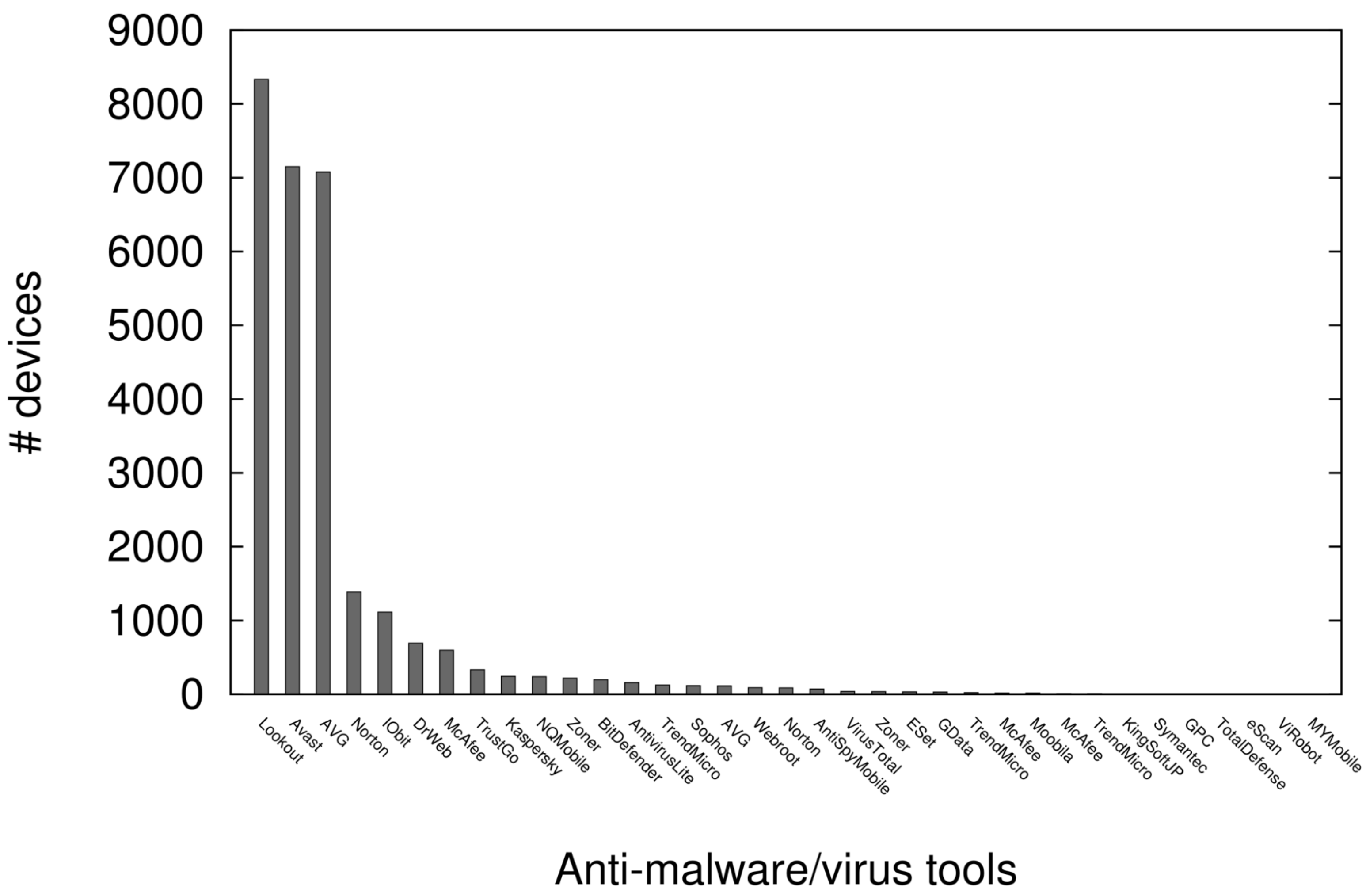
Analysis | File detail | Additional information | Comments 1 | Votes

Antivirus	Result	Update
AegisLab	Suspicious	20140417
AntiVir	Android/BgServ.C	20140416
Avast	Android:BGServ-D [Trj]	20140417
Baidu-International	Trojan.Android.Bgserv.C	20140416
CAT-QuickHeal	Android.Bgserv.C1111	20140416
Commtouch	AndroidOS/GenBl.AF311EAA!Olympus	20140417
Comodo	UnclassifiedMalware	20140416
ESET-NOD32	a variant of Android/Bgserv.C	20140416

http://bit.ly/IDjWos

Estimates

Deployment of AV tools *Mar 2013 – May 2014*



AV tools vs. infection

Mar 2013 – May 2014

25215 devices have some AV tool installed (25.3%)

None are infected according to any of our malware datasets

Information revealed by set of apps

Package names can be revealing:

language of device user

Can also reveal user traits:

Predicting User Traits From a Snapshot of Apps Installed on a Smartphone

Suranga Seneviratne[a,b] *suranga.seneviratne@nicta.com.au*
Aruna Seneviratne[a,b] *aruna.seneviratne@nicta.com.au*
Prasant Mohapatra[c] *prasant@cs.ucdavis.edu*
Anirban Mahanti[b] *anirban.mahanti@nicta.com.au*

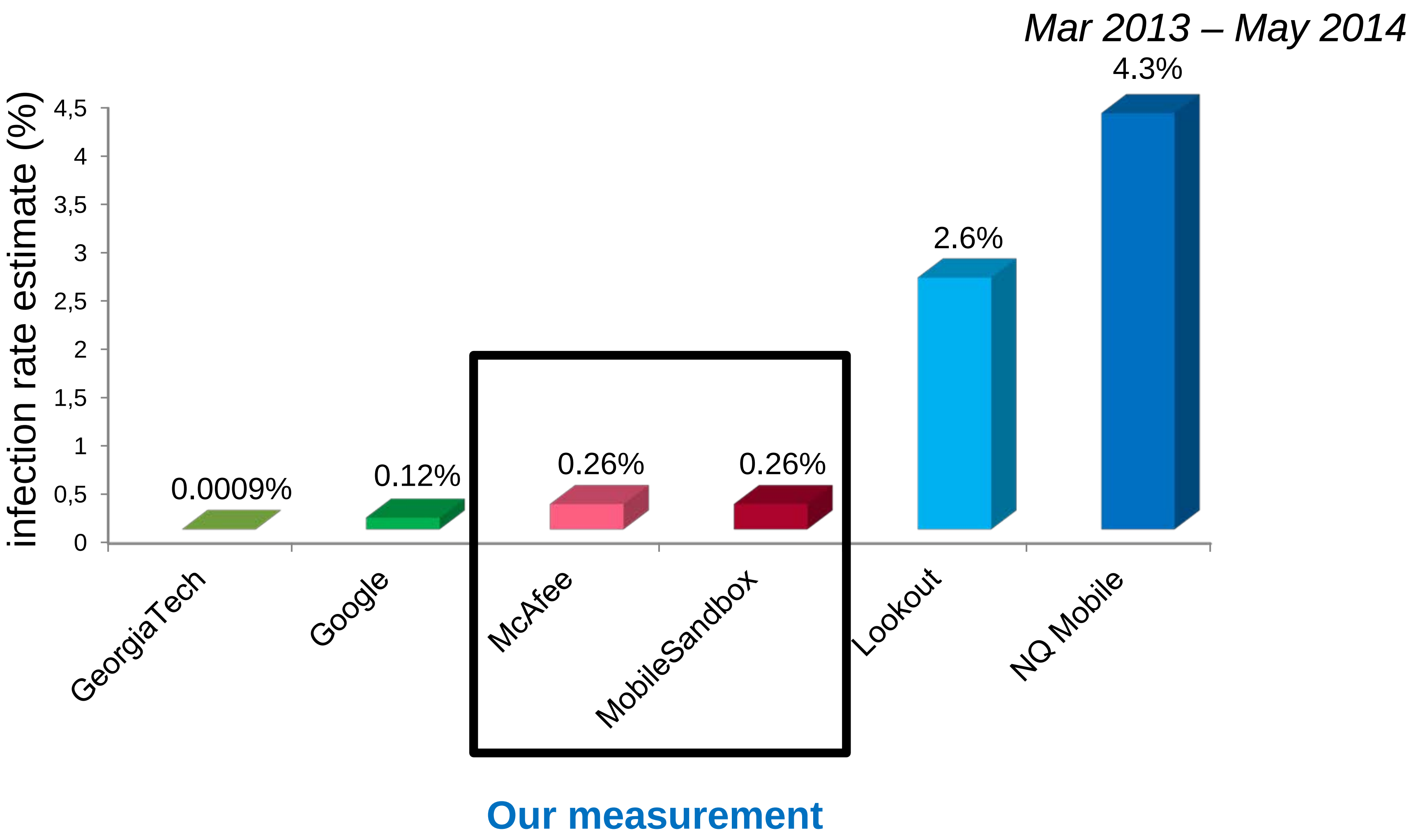
[a]School of EET, University of New South Wales, Australia
[b]NICTA, Australia
[c]Department of Computer Science, University of California, Davis

http://dx.doi.org/10.1145/2636242.2636244

Indicative of user behaviour?

Estimates

Summary: infection rate estimates

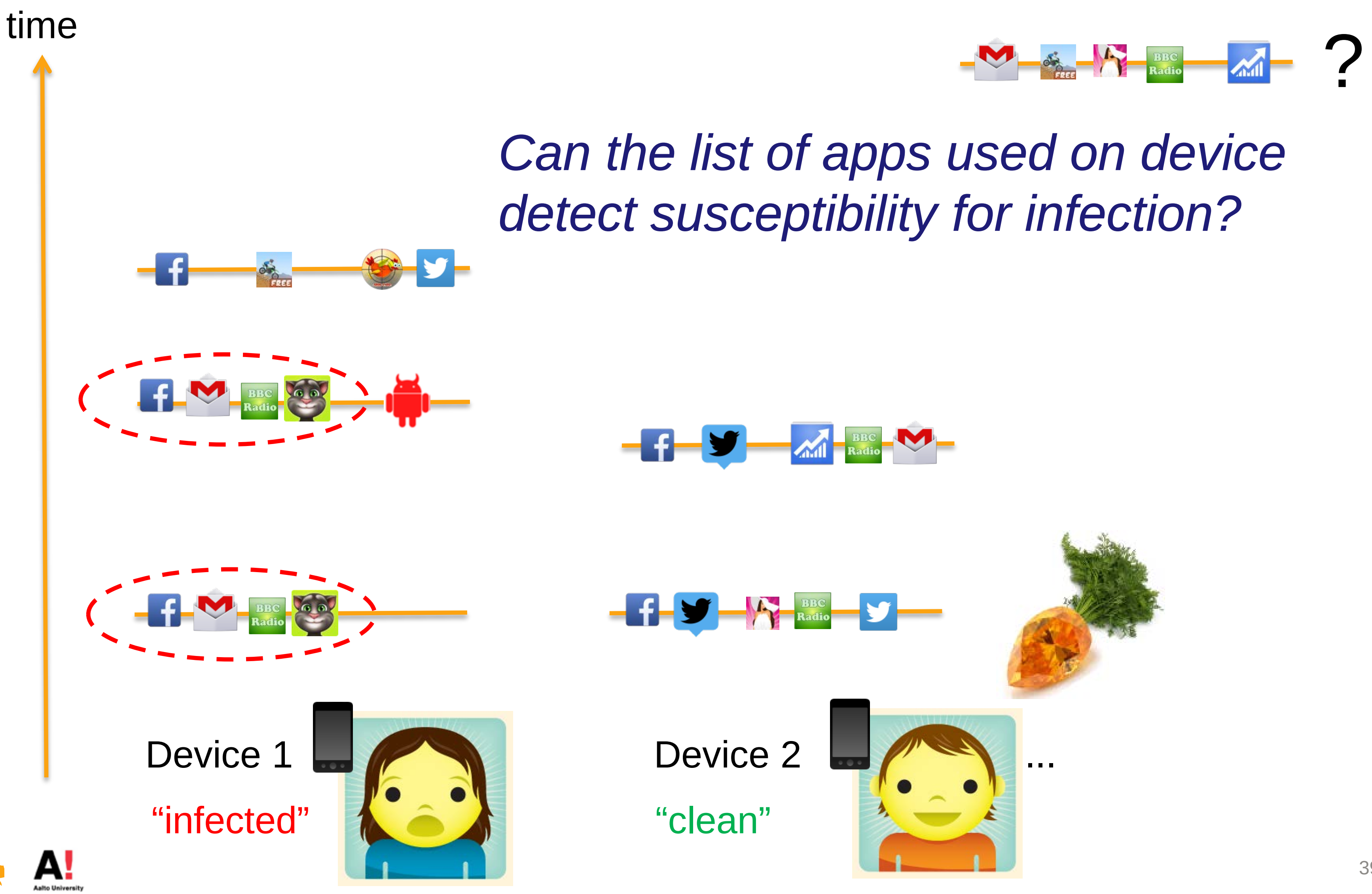


Outline

- Gather data directly from devices
- Accurately estimate malware infection rate
- **Identify risk factors, cheaply**
  - *Separately for each malware dataset*
  - See if we can detect susceptibility for infection!

“The Company You Keep”

Can the list of apps used on device detect susceptibility for infection?



Classifying based on set of apps

- Identifying new malware requires extensive analysis of candidates
- Baseline: random sampling
  - Low infection rates imply that baseline is costly
- Using set of apps to detect susceptibility for infection is cheap
  - Lightweight instrumentation: at virtually no cost

Application: Help anti-malware vendors in the search for new malware

Classifying based on set of apps

Mar 2013 – May 2014

Datasets	Precision	Baseline	Improvement
Detecting infection (new malware)			
McAfee	1.2%	0.26%	**4.5X**
Mobile Sandbox	0.9%	0.25%	**3.5X**
Detecting infection (undetected malware)			
McAfee	0.16%	0.05%	**3.5X**
Mobile Sandbox	0.12%	0.05%	**2.6X**

Detecting infection: the "Real-life" case

Mar 2013 – May 2014

"Original" malware set used for training;

- **Training set** labeled using "Original" malware set only

"New" set used for testing

- See how well we can detect infection by "New" malware set

Datasets	Precision	Baseline	Improvement
McAfee	0.7%	0.19%	**3.5X**
Mobile Sandbox	0.3%	0.08%	**4X**

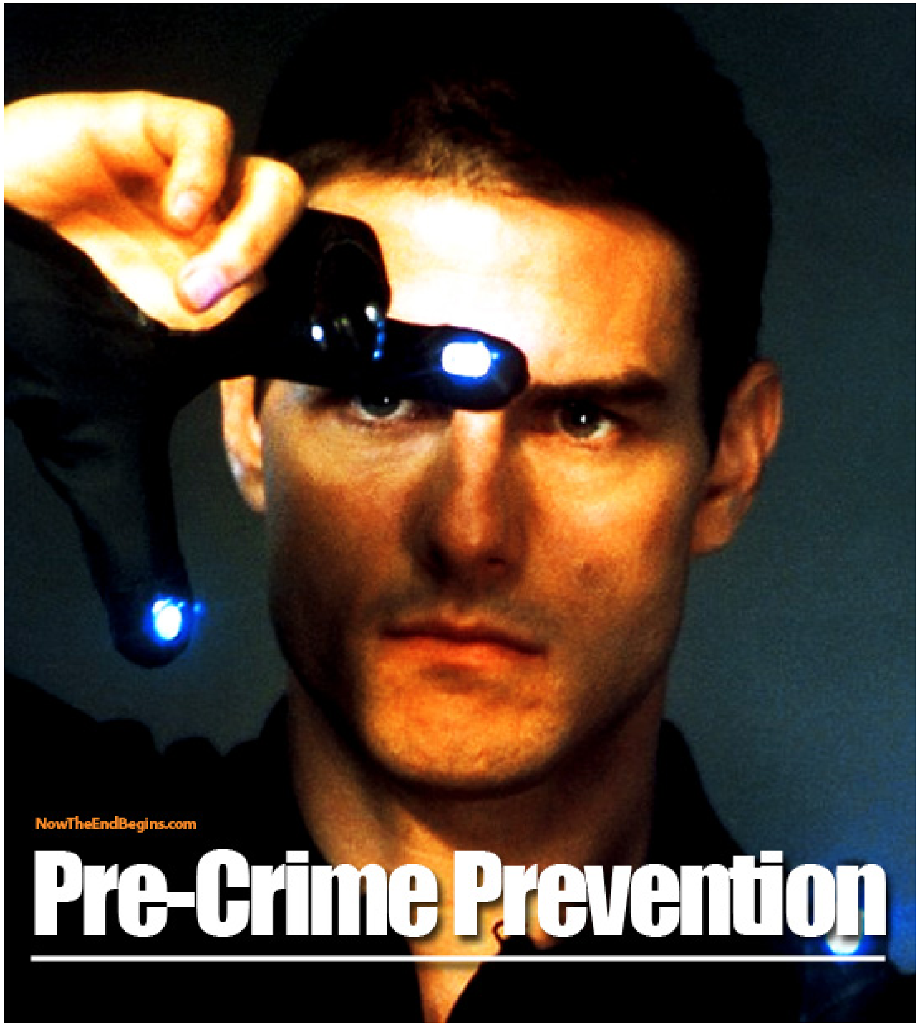
Multinomial Naïve Bayes

Taking timestamps into account

Mar 2013 – Oct 2013

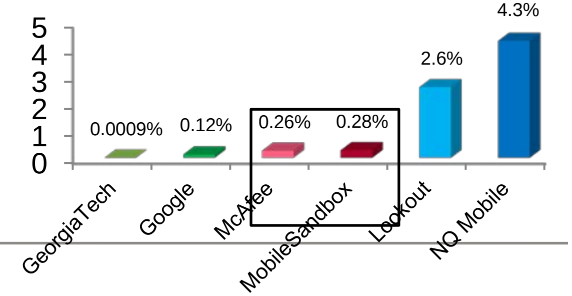
- Carat records have timestamps
  - At least 155 devices changed state from clean to infected during data collection period
  - can we predict likelihood of ***eventual*** infection?

Identify vulnerable devices before they are infected?



Application: Help enterprise IT admin identify users for training

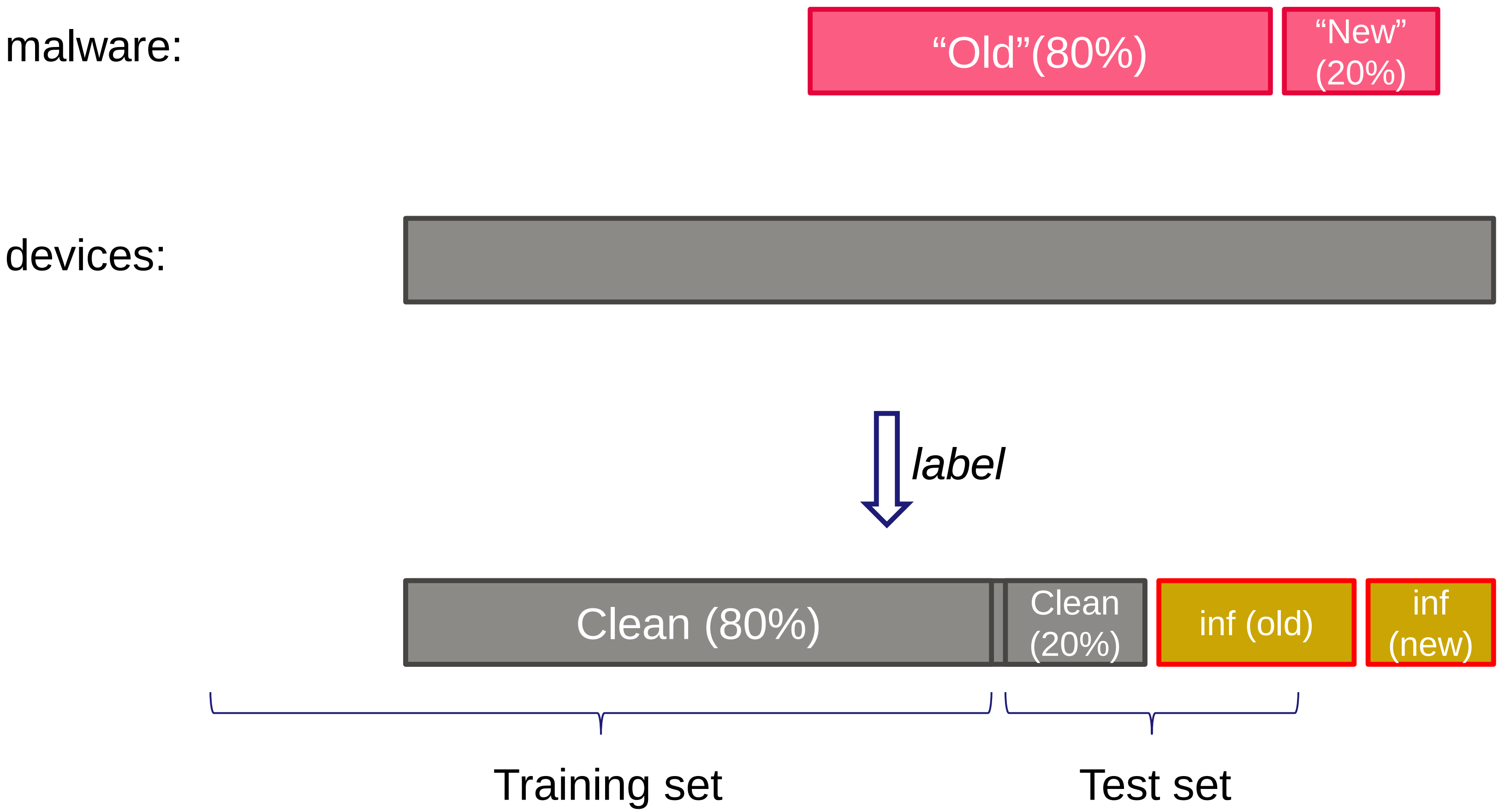
Summary



- Measure Android malware infection rates directly
  - No common agreement of what is malware
  - False positives and re-classifications are common
- Identify inexpensive risk factors
  - can aid in search for new malware
  - set of apps indicative of user behaviour/traits'

http://se-sy.org/projects/malware/

Detecting infection (new)



Detecting infection (unknown)