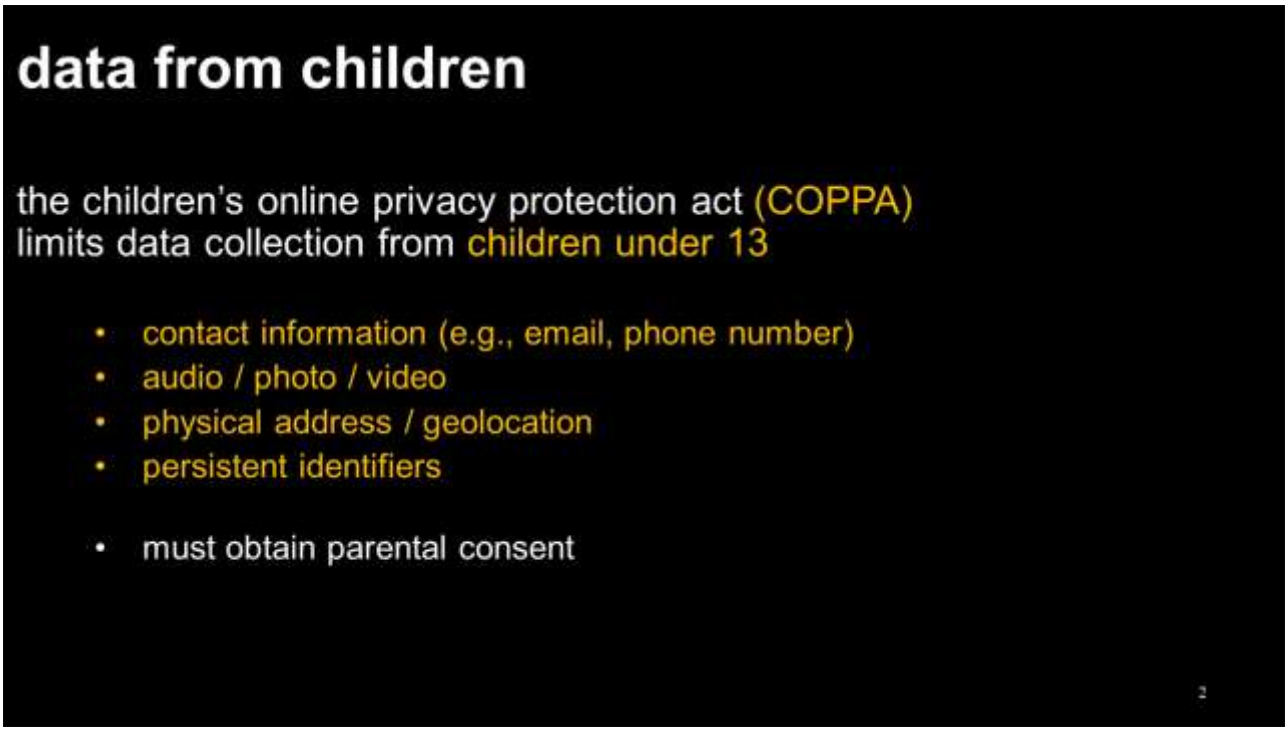
“is our children’s apps learning?”

automatically detecting COPPA violations

Irwin Reyes
Primal Wijesekera
Abbas Razaghpanah
Joel Reardon
Narseo Vallina-Rodriguez
Serge Egelman
Christian Kreibich

My name is Irwin Reyes from the International Computer Science Institute. I’ll be speaking to you about our ongoing work to automatically identify privacy issues in mobile apps for kids.

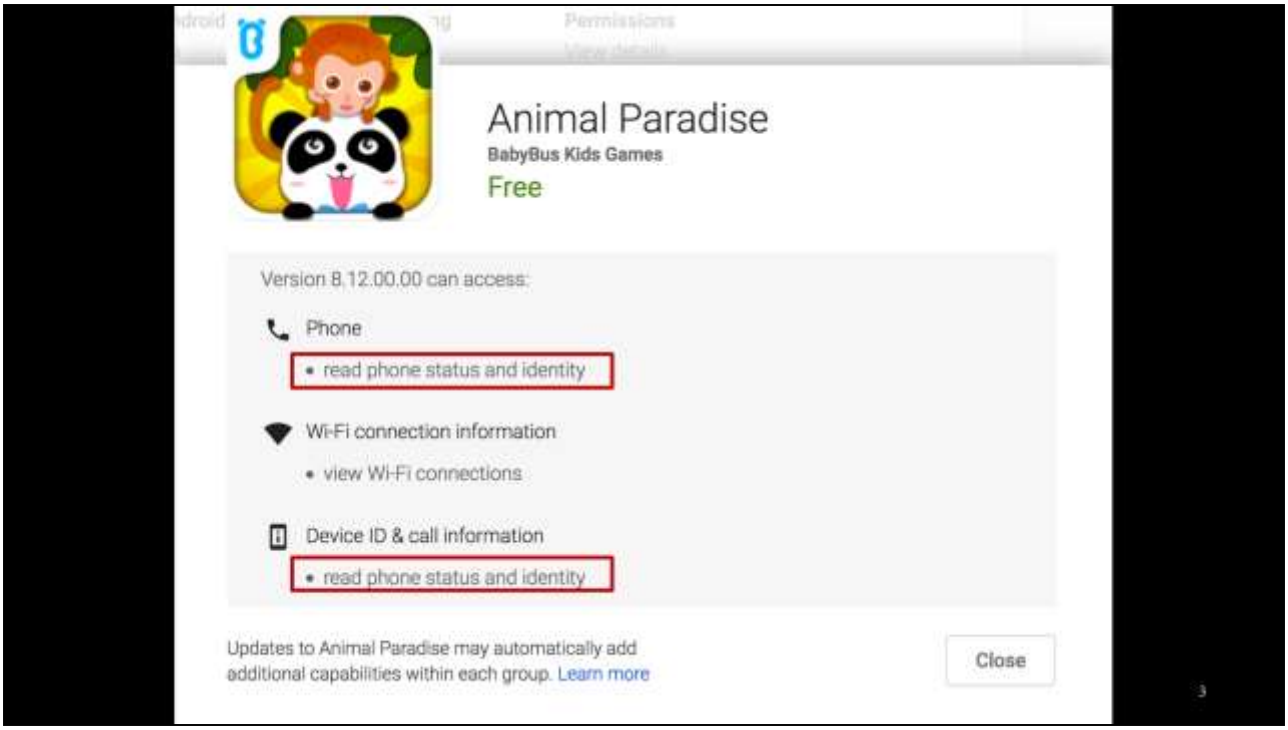


First, a bit of background. In the United States, we have the Children's Online Privacy Protection Act – COPPA, for short. COPPA governs how online services such as mobile apps collect data from users under 13 years of age.

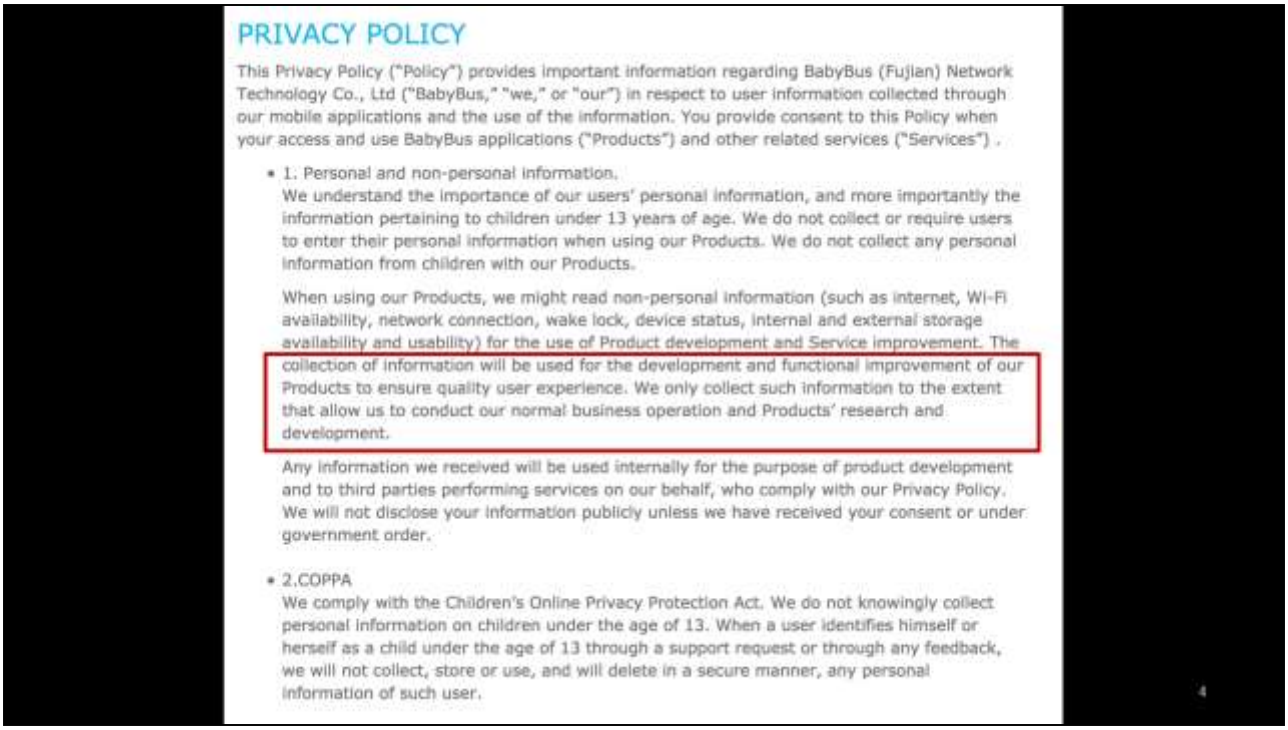
This regulation limits the gathering personal information like email addresses and phone numbers, audio/visual captures, and the fine geographic location of children.

Additionally, COPPA restricts the collection and use of persistent identifiers that can be used to uniquely identify individuals over time and across different services. Persistent identifiers include hardware serial numbers and software fingerprints. Using these for behavioral advertising is prohibited.

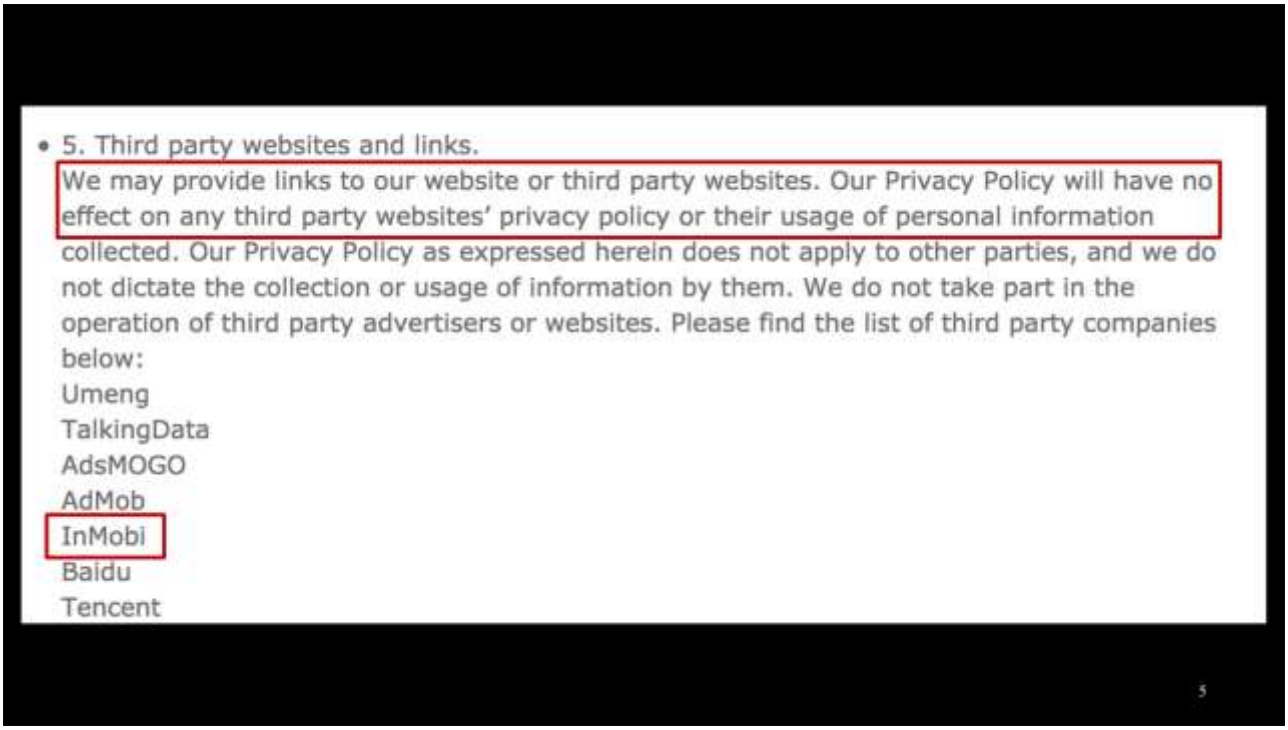
In cases where a mobile app operator wishes to collect restricted data for internal operational purposes (e.g., infrastructure maintenance, high score lists), they must notify parents of this and obtain parental consent prior to collection.



Let's step through an example of how notice and consent is done. Here, the developer BabyBus declares that their Animal Paradise game is able to access device identifiers. Note that these declarations just mean the app *can* use those privileges, not that they actually do so. Still, parents can review this list prior to installing an app and decide if this is appropriate.



Developers also provide privacy policies to further elaborate on their data collection practices. In BabyBus's privacy policy, the developer states that any data collection is strictly for product maintenance purposes.

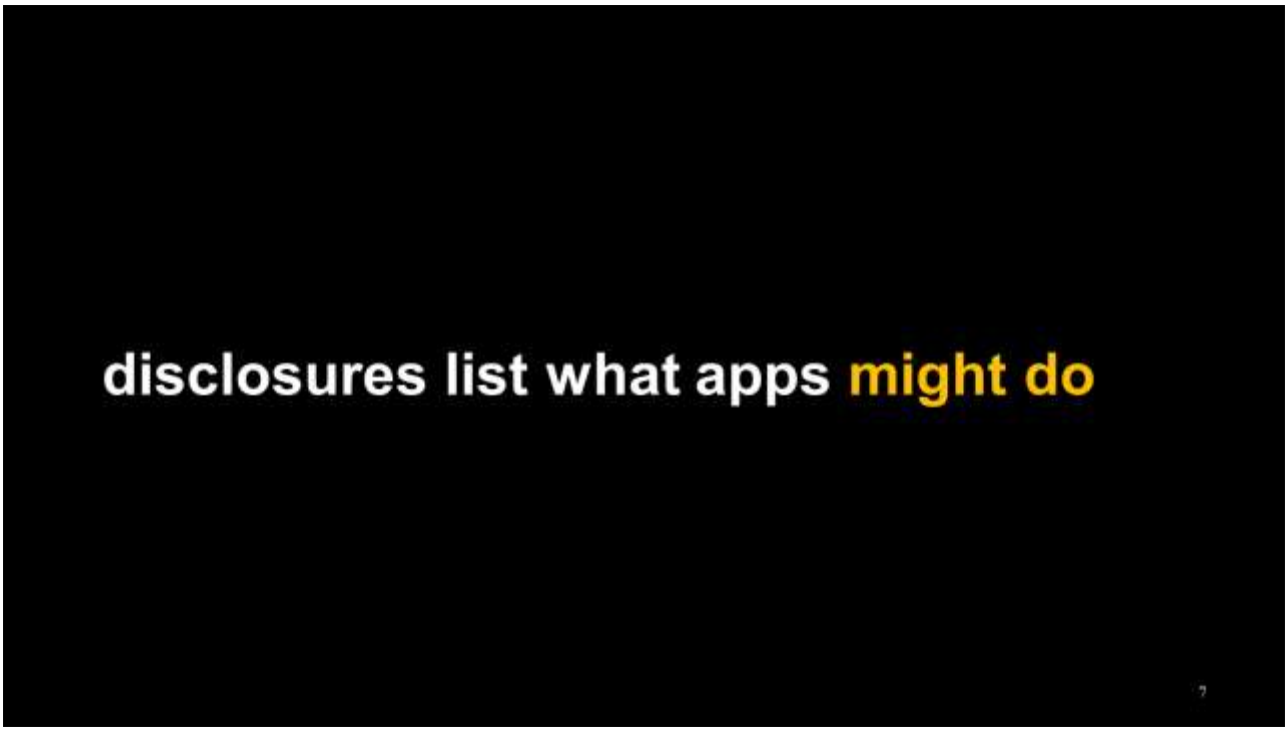


However, disclosures don't always tell the full story. Further down in the privacy policy for the Baby Panda Care game, the developer lists third-party websites that the developer might link to, and that this policy does not apply to them.

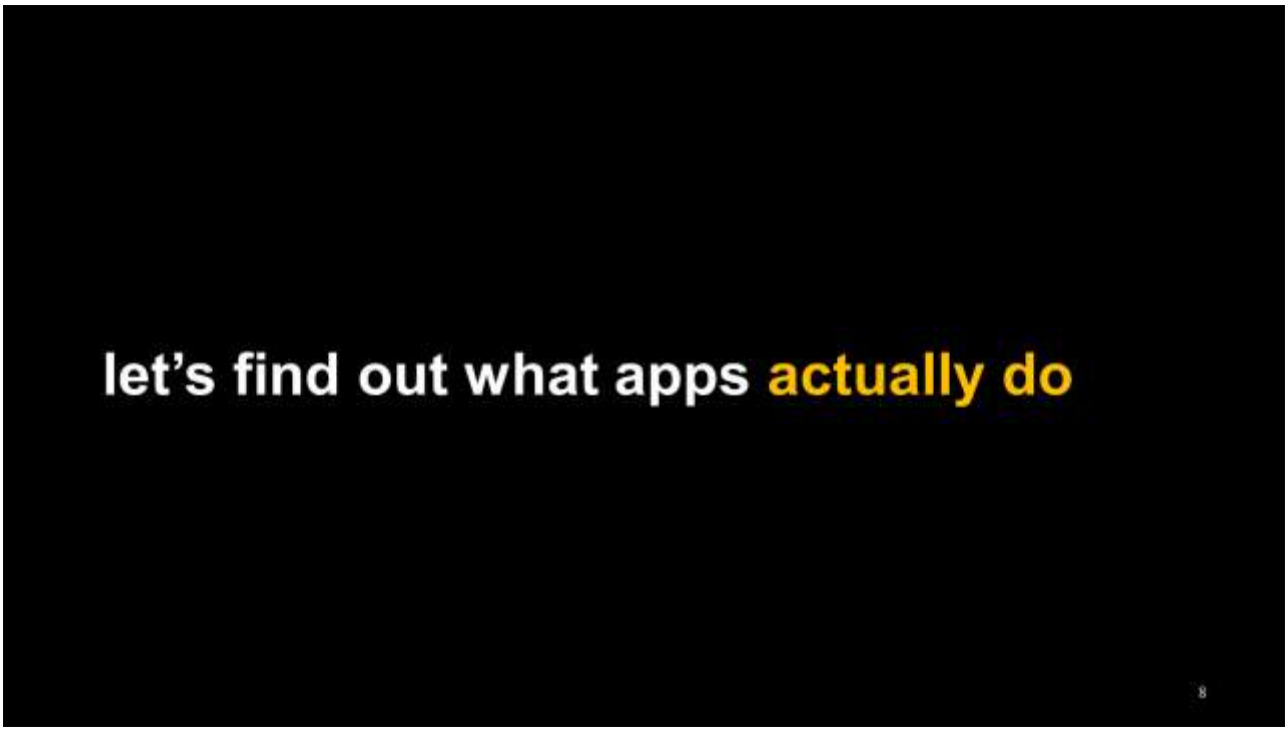
They fail to mention a number of important details:

- First, although the policy doesn't explain how these are integrated into products, if at all
- Second, that a number of these are not just websites, but advertising libraries that may be bundled with apps to target apps to particular users
- And finally, that one of those advertisers, InMobi…

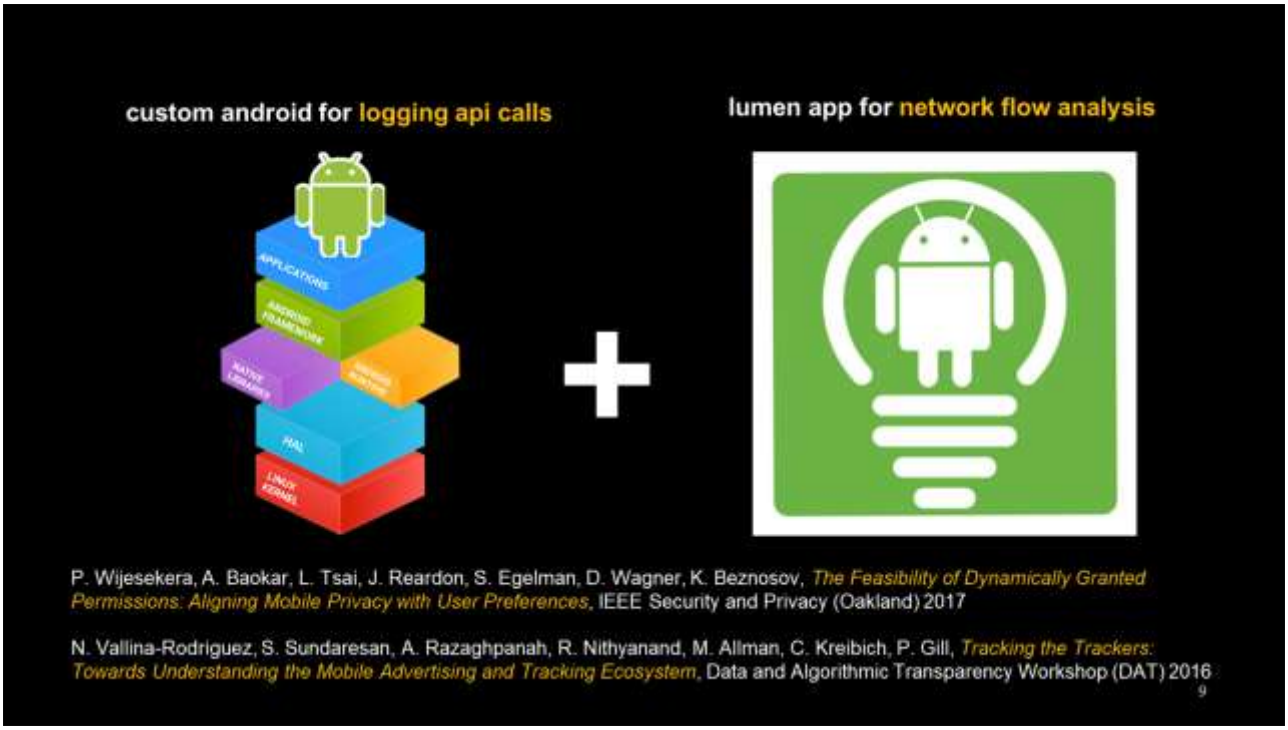
last year settled with the FTC for the silent collection of sensitive geolocation data from children.



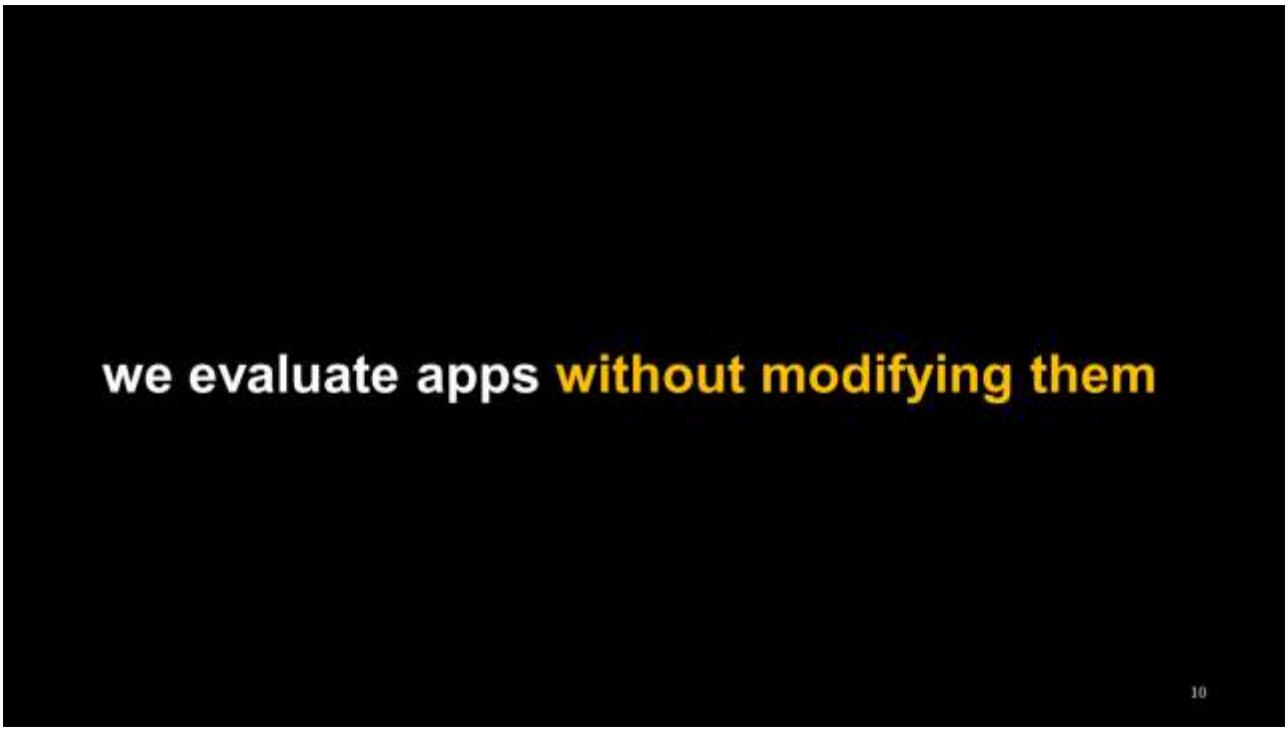
At best, disclosures list what apps might do: they might access a piece of identifying information, or might share it with this third-party company or that.



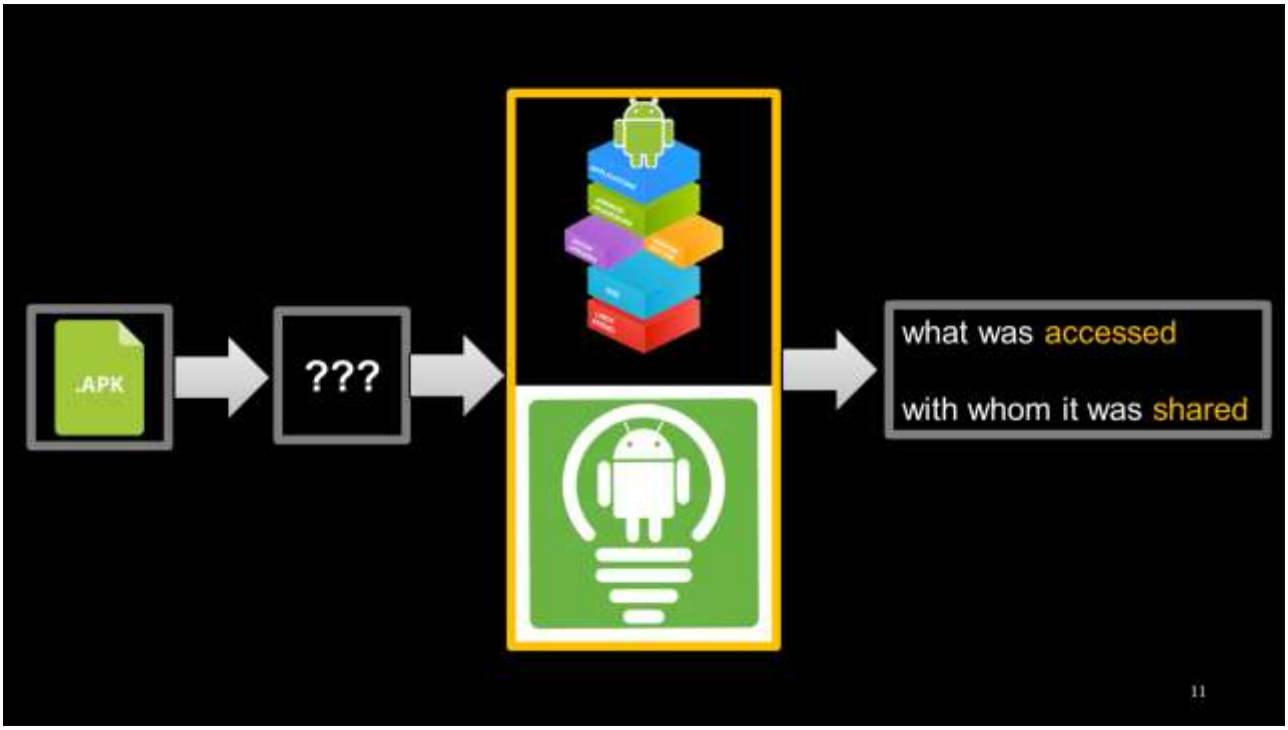
What we want is something more actionable. Let’s find out what apps actually do.



To that end, we propose a method of dynamic analysis to characterize how apps access and share sensitive data. We combine two of our existing tools: First, an instrumented version of the Android platform to determine which sensitive API calls are being invoked. This is part of the work on permission systems my colleague Primal Wijesekra presented in yesterday's session. And second, the Lumen network flow tool to examine what data is leaving the device and to whom it's going, even over TLS-protected communications.

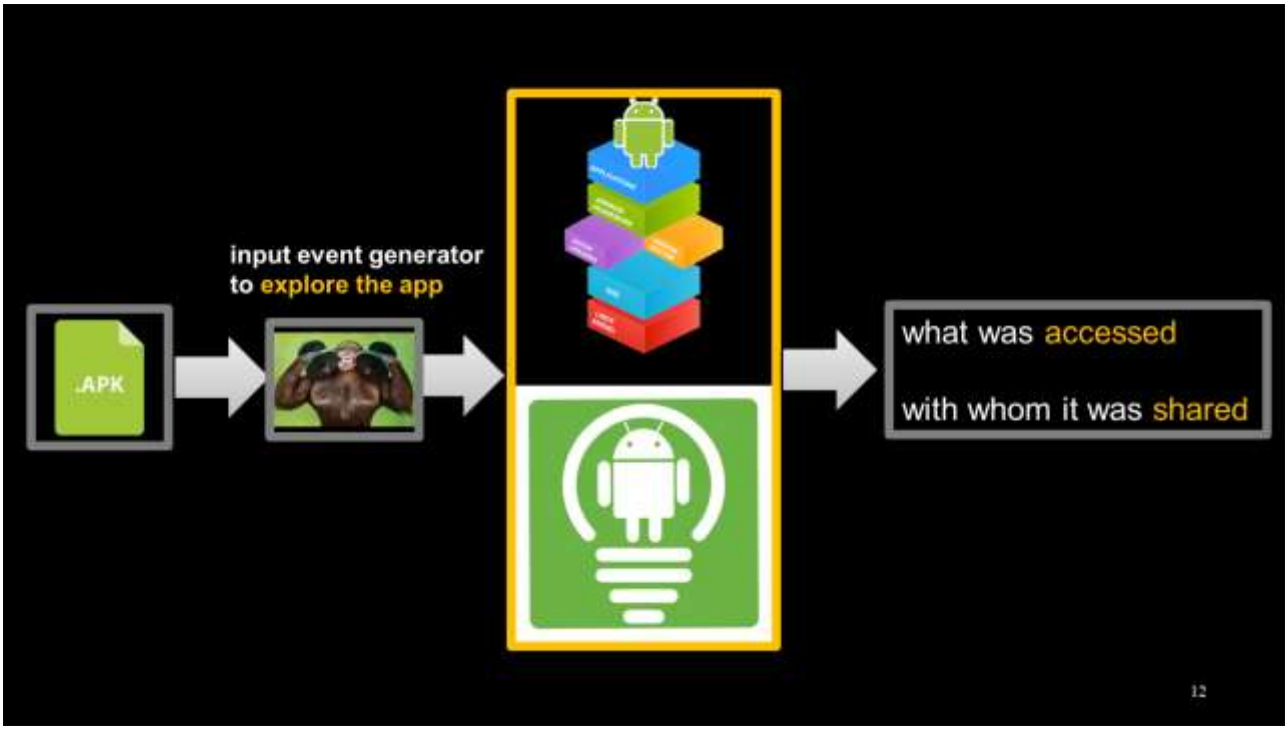


This dynamic analysis approach allows us to examine apps without having to modify them or otherwise examine their internals. We can download apps and run them as-is, exactly how a consumer would use them.

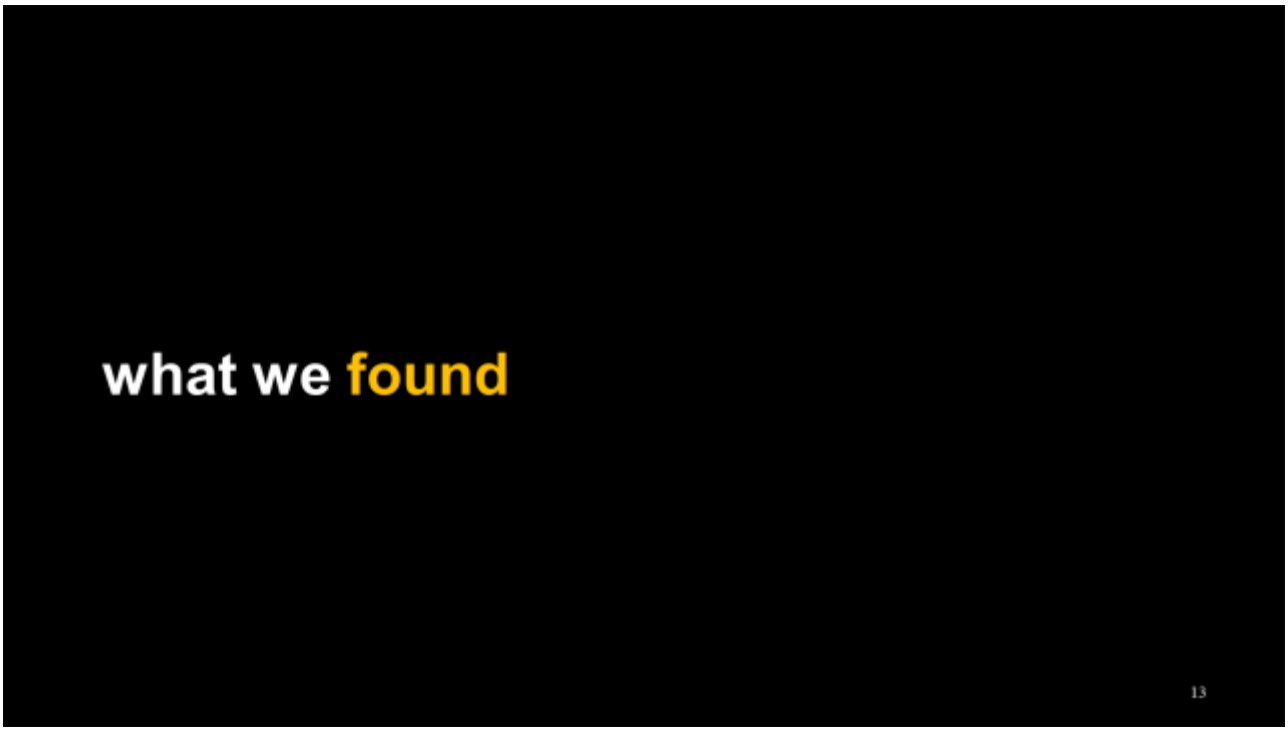


Dynamic analysis does present a technical challenge, in that it's insufficient to just run the application on a testbed with our tools.

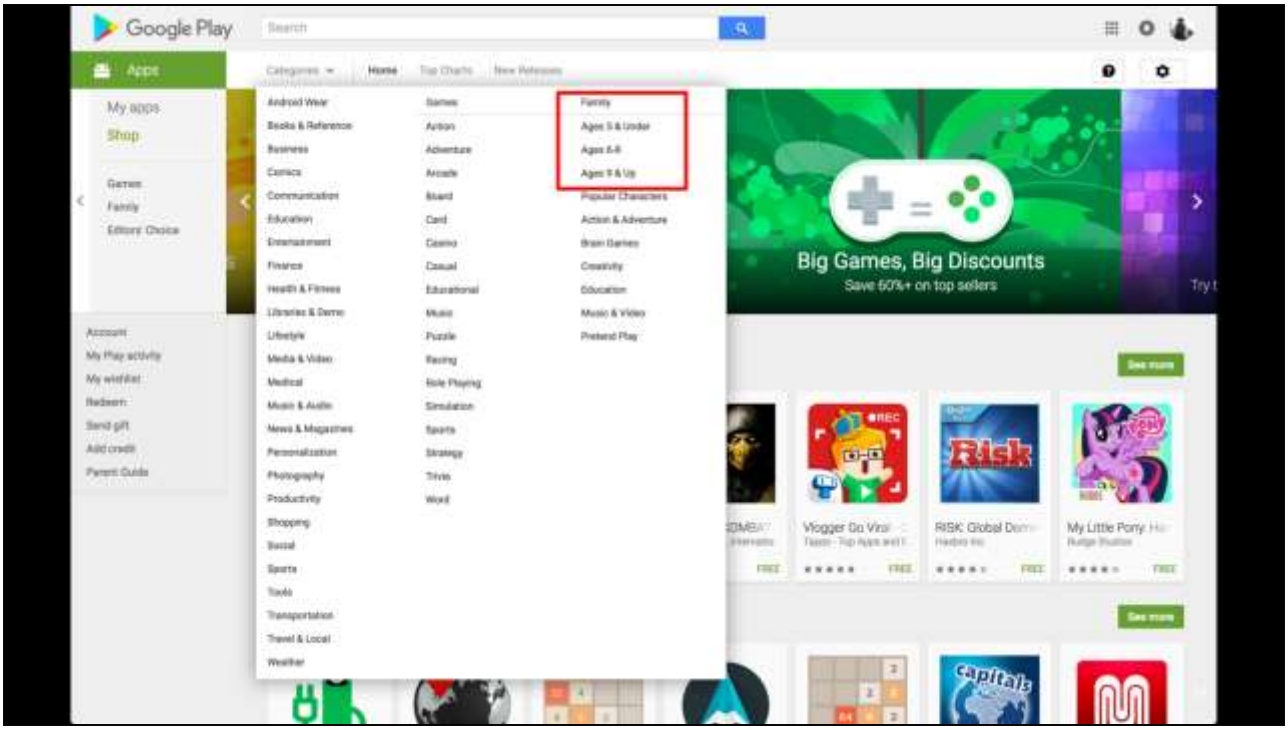
We need someone or something to actually interact with the app being tested. Otherwise, we would only observe the app's behavior as it sits on its initial screen. We'd like to be able to examine apps at scale.



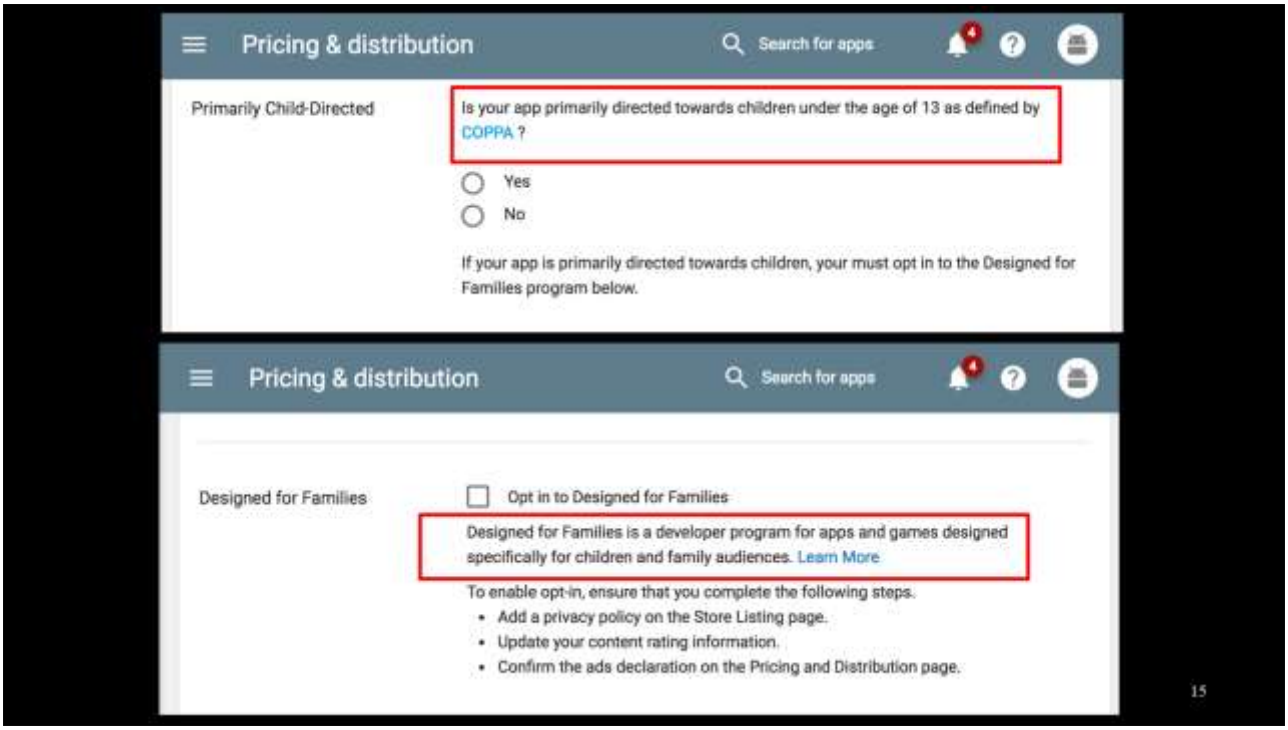
For that, we turn to the Android Exerciser Monkey, a tool included in the Android SDK that generates a pseudorandom stream of input events – taps, swipes, and screen transitions. In our internal testing, we found that Monkey had comparable coverage to that of an undergraduate tester in a majority of the children's apps evaluated – no surprise, given the simple nature of apps meant for toddlers and young children.



This system has been built and been used to test a number of children's apps.



We collected free apps from the Google Play Store listed under child-specific categories.



Developers who list their apps in those child-specific categories must opt-in to Google Play Store

All in all, we built a corpus of 735 free apps explicitly targeted to kids.

This was done in two stages as we developed our test pipeline.

out of 735 apps tested

- 41% share at least one type of persistent hardware identifiers with third-parties
  - 36% hardware serial number
  - 13% imei
  - 8% wi-fi mac address

17

The Washington Post

Democracy Dies in Darkness

Innovations

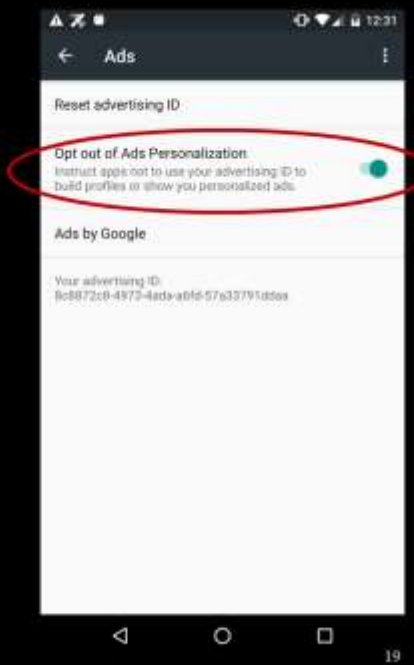
Uber's 'fingerprinting' of iPhones after users delete app has sparked an FTC complaint

By Steven Overly April 27

18

out of 735 apps tested

- 51% share the google advertising identifier despite having opted out
- ideal behavior: don't share at all, or send non-unique fixed value



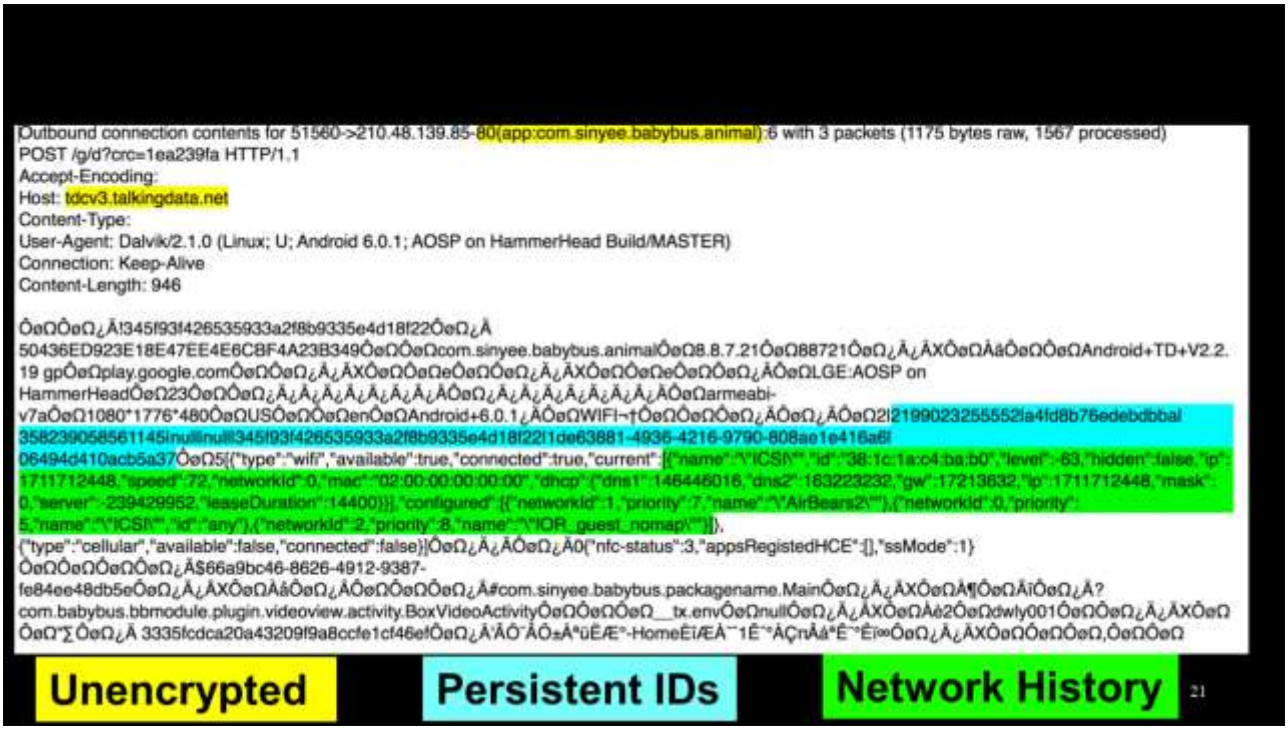
19

out of 735 apps tested

- 29 apps use device geolocation functionality
- 6 transmit geolocation data to third-parties
- 3 do so using unencrypted HTTP

- 5 apps read the registered google account
- 3 transmit the email address to third parties
- all use TLS

20


Outbound connection contents for 51560->210.48.139.85-80(app:com.sinyee.babybus.animal):6 with 3 packets (1175 bytes raw, 1567 processed)
POST /g/d?crc=1ea239fa HTTP/1.1
Accept-Encoding:
Host: tdcv3.talkingdata.net
Content-Type:
User-Agent: Dalvik/2.1.0 (Linux; U; Android 6.0.1; AOSP on HammerHead Build/MASTER)
Connection: Keep-Alive
Content-Length: 946
ÔøΩÔøΩ¿Ã!345f93f426535933a2f8b9335e4d18f22ÔøΩ¿Ã
50436ED923E18E47EE4E6CBF4A23B349ÔøΩÔøΩcom.sinyee.babybus.animalÔøΩ8.8.7.21ÔøΩ88721ÔøΩ¿Ã¿ÃXÔøΩÀåÔøΩÔøΩAndroid+TD+V2.2.
19 gpÔøΩplay.google.comÔøΩÔøΩ¿Ã¿ÃXÔøΩÔøΩeÔøΩÔøΩ¿Ã¿ÃXÔøΩÔøΩeÔøΩÔøΩ¿ÃÔøΩLGE:AOSP on
HammerHeadÔøΩ23ÔøΩÔøΩ¿Ã¿Ã¿Ã¿Ã¿Ã¿Ã¿ÃÔøΩ¿Ã¿Ã¿Ã¿Ã¿Ã¿Ã¿ÃÔøΩarmeabi-
v7aÔøΩ1080*1776*480ÔøΩUSÔøΩÔøΩenÔøΩAndroid+6.0.1¿ÃÔøΩWIFI¬†ÔøΩÔøΩÔøΩ¿ÃÔøΩ¿ÃÔøΩ2l2199023255552la4fd8b76edebdbbal
358239058561145inull|null|345f93f426535933a2f8b9335e4d18f22l1de63881-4936-4216-9790-808ae1e416a6l
06494d410acb5a37ÔøΩ5{{"type":"wifi","available":true,"connected":true,"current":[{"name":"\"ICSI\"","id":"38:1c:1a:c4:ba:b0","level":-63,"hidden":false,"ip":
1711712448,"speed":72,"networkId":0,"mac":"02:00:00:00:00:00","dhcp":{"dns1":146446016,"dns2":183223232,"gw":17213632,"ip":1711712448,"mask":
0,"server":-239429952,"leaseDuration":14400}}],"configured":[{"networkId":1,"priority":7,"name":"\"AirBears2\""},{"networkId":0,"priority":
5,"name":"\"ICSI\"","id":"any"},{"networkId":2,"priority":8,"name":"\"IOR_guest_nomap\""}]},
{"type":"cellular","available":false,"connected":false}]ÔøΩ¿Ã¿ÃÔøΩ¿Ã0{"nfc-status":3,"appsRegistedHCE":[],"ssMode":1}
ÔøΩÔøΩÔøΩÔøΩÔøΩ¿Ã$66a9bc46-8626-4912-9387-
fe84ee48db5eÔøΩ¿Ã¿ÃXÔøΩÀåÔøΩ¿ÃÔøΩÔøΩÔøΩÔøΩ¿Ã#com.sinyee.babybus.packagename.MainÔøΩ¿Ã¿ÃXÔøΩÀ¶ÔøΩÃïÔøΩ¿Ã?
com.babybus.bbmodule.plugin.videoview.activity.BoxVideoActivityÔøΩÔøΩÔøΩÔøΩ__tx.envÔøΩnullÔøΩ¿Ã¿ÃXÔøΩÀé2ÔøΩdwly001ÔøΩÔøΩ¿Ã¿ÃXÔøΩ
ÔøΩ"∑ÔøΩ¿Ã 3335fcdca20a43209f9a8ccfe1cf46efÔøΩ¿Ã'ÃÕ¯ÃÕ±Ã°üÊÆ°-HomeÊïÆÀ¯¯1Ê¯°ÀÇπÅá°Ê¯°Êï∞ÔøΩ¿Ã¿ÃXÔøΩÔøΩÔøΩÔøΩ,ÔøΩÔøΩ
Unencrypted
Persistent IDs
Network History
21


Home
About
Contact
AppCensus
FreeN'Throw
HD WallPapers
Learning
Live for EuroLeague
Sniper
Training Workout
Best Shot
Big Finger
Double Free
Flying Slam Dunks
Guess the Player
22


AppCensus
Search Game App on Google Play
By Category
About
Contact
Personal Information
We searched for the following data types:
Location
Phone
Email
Location
Device Identifiers
We searched for the following data types:
IMEI
Wifi Mac
AAID
AAID
SERIAL #
Hardware Identifiers
Multimedia
We searched for the following data types:
Photo
Video
No personal data detected leaving the device during our testing.
23

what's next

- virtualized testbed: scalable tests on commodity servers
- crowdsourcing to help "stranded" monkeys
- expanded analysis to all apps
- trends over time

THANK YOU!

25