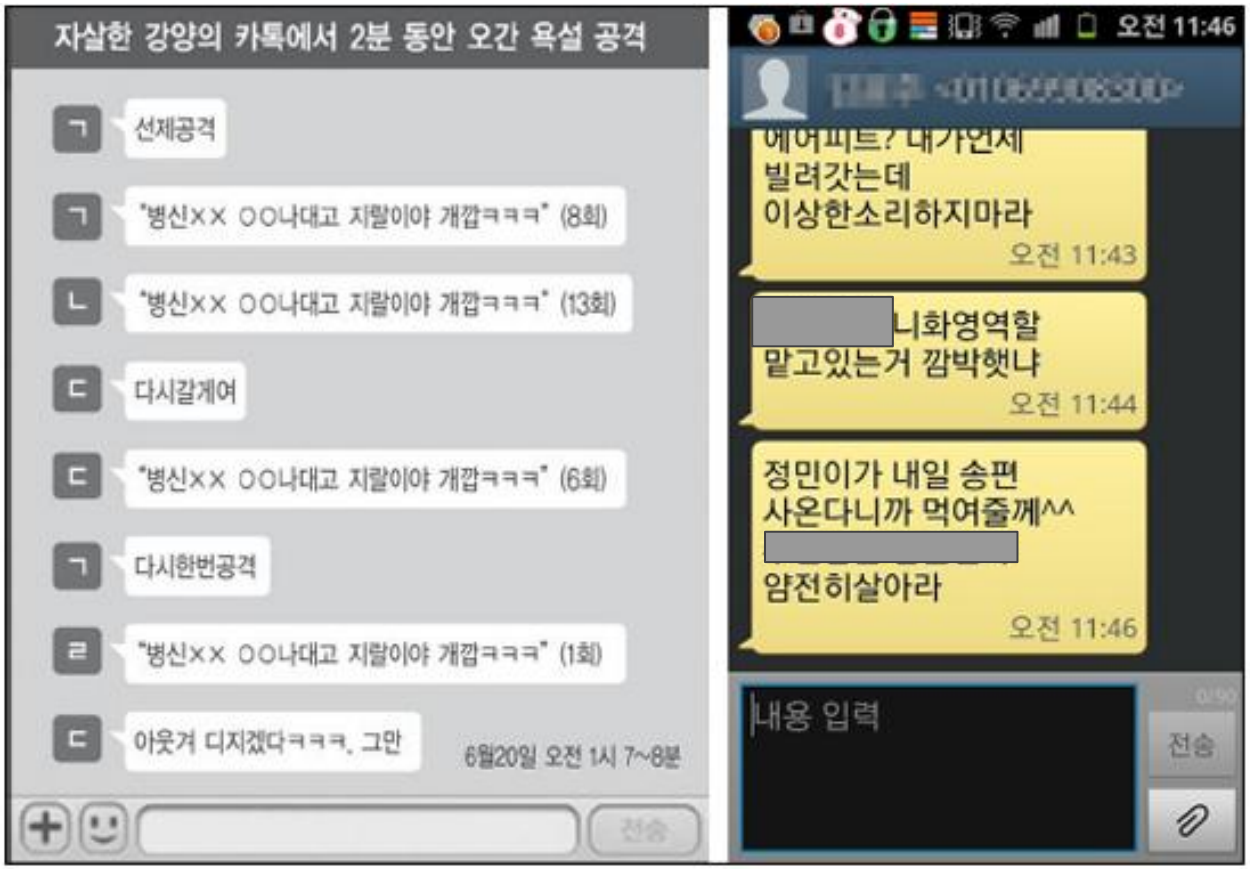
PinkDroid
A System Level Contents Blocker

Wonsup Yoon, Bilgehan Bingol, Nak Hyun Choi


Children are important!

But in their smartphones...



Bad words

But in their smartphones...

– – –

Adult contents

But in their smartphones...

– – –

Violence

But in their smartphones...

Medical

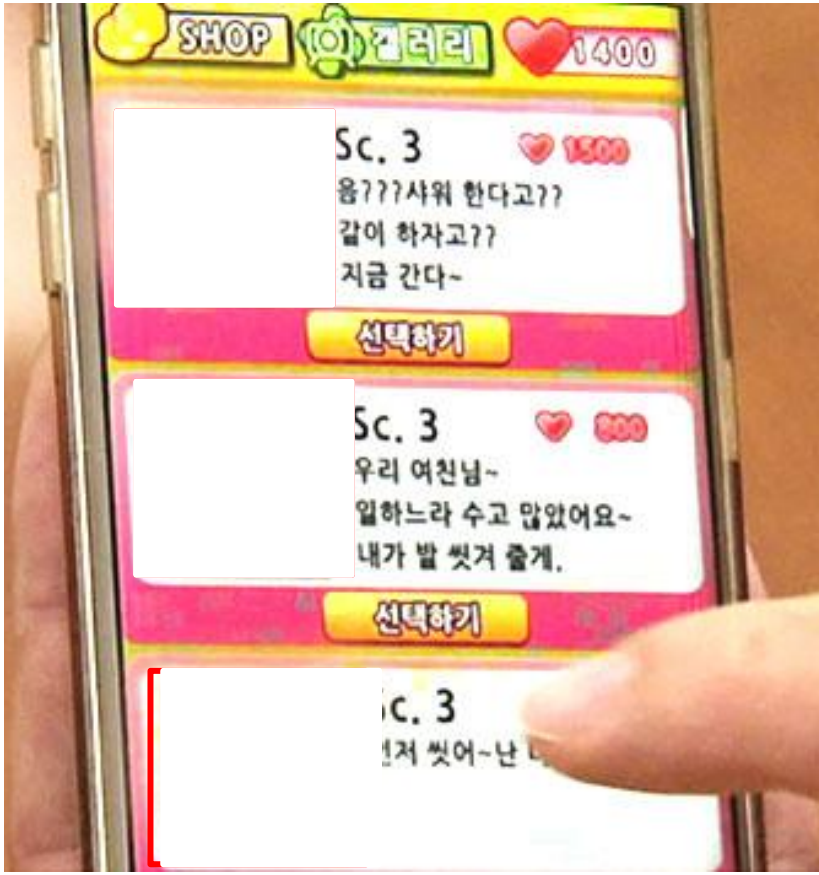
We must protect our children from the inappropriate contents!

Previous solution

- Only blocks applications (block or permit)
- Is is worthwhile to block KakaoTalk or Facebook?
- Even in the permitted applications, there can be inappropriate contents

Proposed Solution

- Block inappropriate contents at the UI rendering
- Detect which contents(text, images) are inappropriate
- Block that contents at runtime



Proposed Solution

- Anti-Social Dictionary
- Content Trust Model
- External API

Anti-Social Dictionary

- Target contents: Texts
- Using predefined text set
- Simple string matching
- But, it is not suitable for photos.

Content Trust Model

- Target contents: Photos
- Tracking all photos in a smartphone
- Determine drawable photos with their sources or characteristics.

Content Trust Model

Trusted	Untrusted
• Built with firmware image • Built with trusted apps * • User generated contents ** (ex. camera) • Contents generated from trusted contents *** • Small images **** • ...	• Downloaded from the Internet • Built with untrusted apps • Sended from outside • Contents generated from untrusted contents *** • ...

* If Google Play testers work very hard and don't make mistakes
** If your children are not anti-social
*** Trusts can be inherited
**** If a photo is small enough, it is unable to store much information. It should be used for UI.

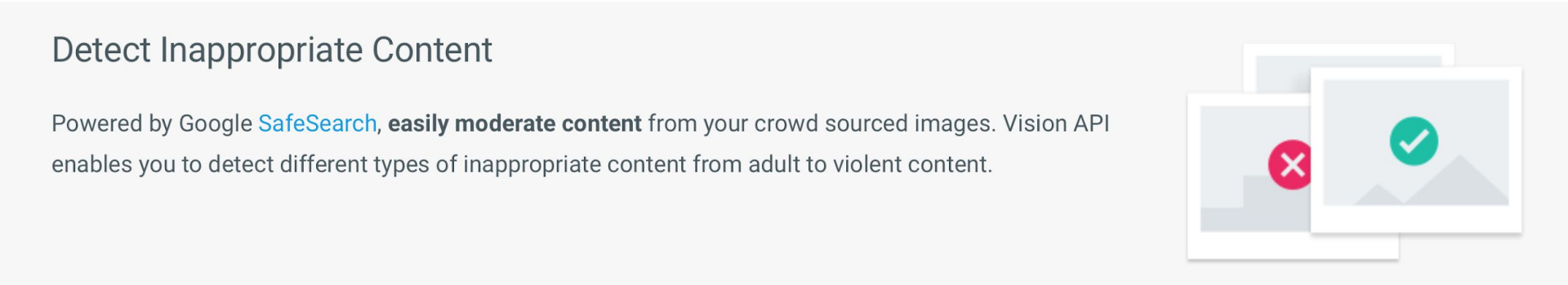
Content Trust Model

Trusted	Untrusted
• **Built with firmware image** • **Built with trusted apps *** • User generated contents ** (ex. camera) • **Contents generated from trusted contents ***** • **Small images ****** • ...	• Downloaded from the Internet • Built with untrusted apps • Sended from outside • Contents generated from untrusted contents *** • ...

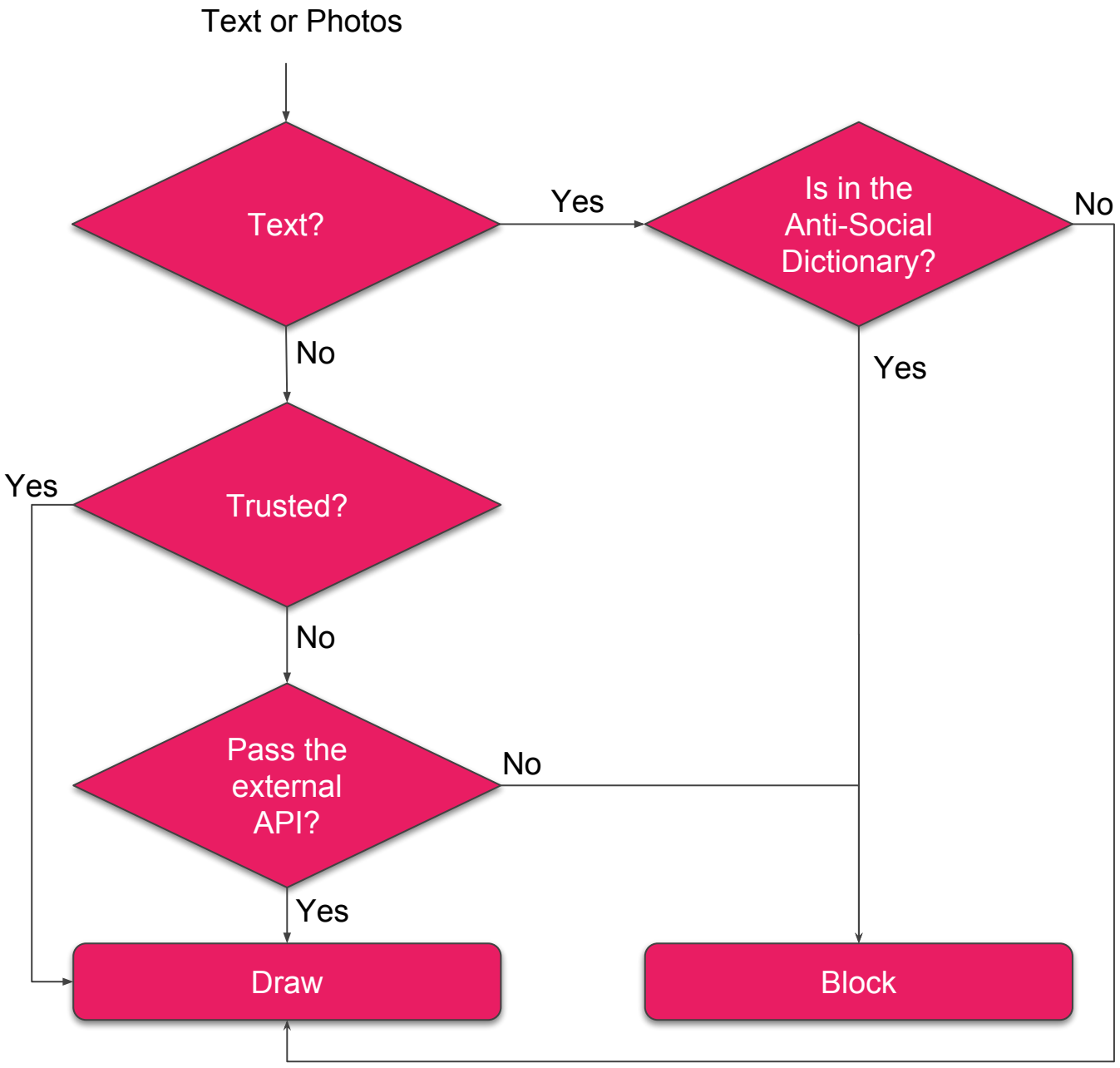
* If Google Play testers work very hard and don't make mistakes
** If your children are not anti-social
*** Trusts can be inherited
**** If a photo is small enough, it is unable to store much information. It should be used for UI.

External API

- Target contents: Untrusted photos
- Google offers Cloud Vision API
  - It can detect spoof, medical, adult, and violence contents.
- It can be used via web API.
- For performance reasons results should be cashed.
- It can be replaced by any other Deep learning frameworks.



* https://cloud.google.com/vision/


Text or Photos
Text?
Yes
Is in the Anti-Social Dictionary?
No
No
Yes
Yes
Trusted?
No
Pass the external API?
No
Yes
Draw
Block

Evaluation

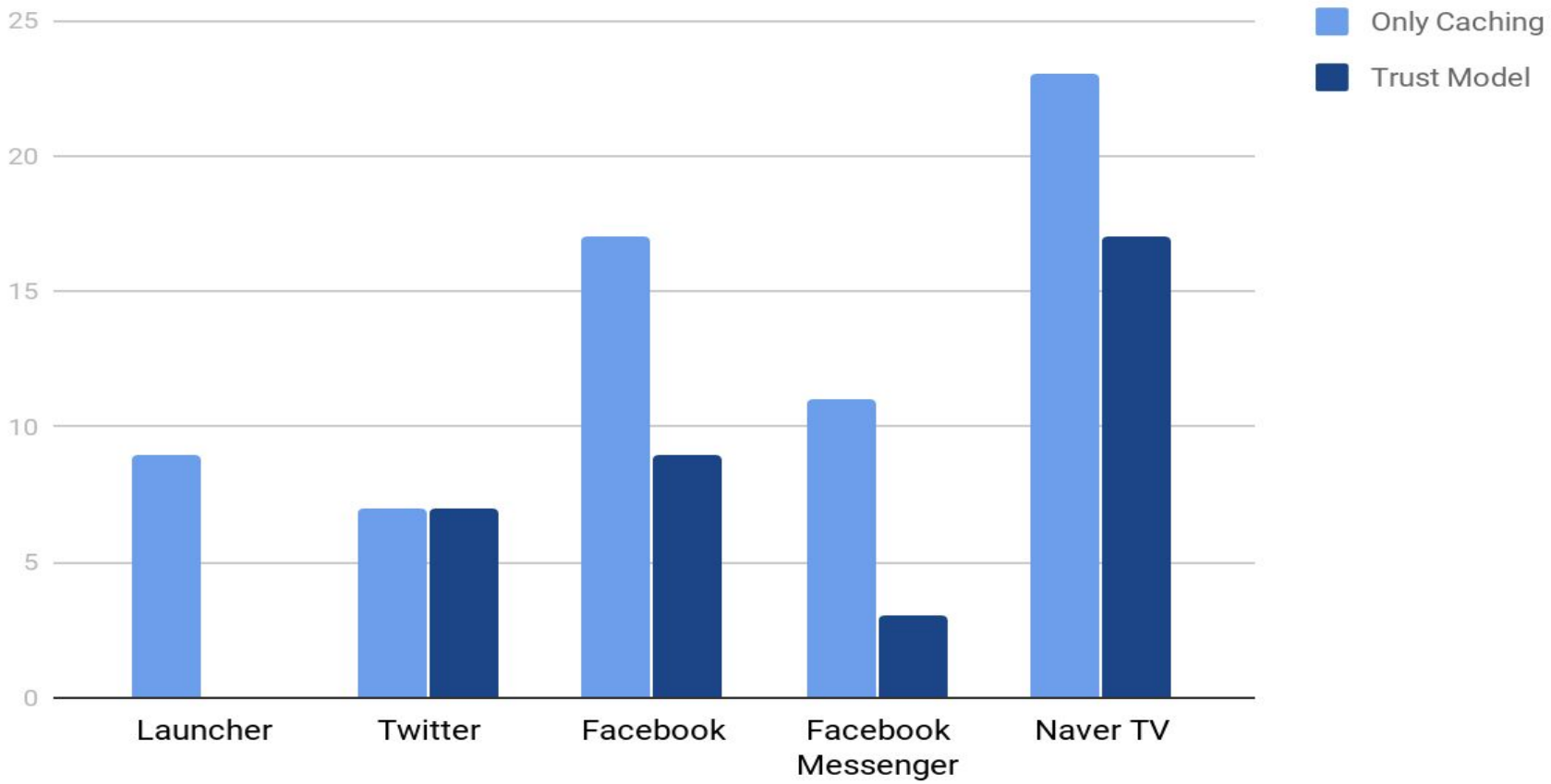
Evaluation

- Evaluation
  - Performance (Cost): # of images sent to Google
    - $1.50 for 1,000 images
  - Safety: # of images filtered

- Environment
  - Android 7.1.1
  - aosp_x86-eng
  - Testing Apps
    - Launcher
    - Twitter
    - Facebook
    - Facebook Messenger
    - Naver TV

Performance evaluation (Only large images)

of requests

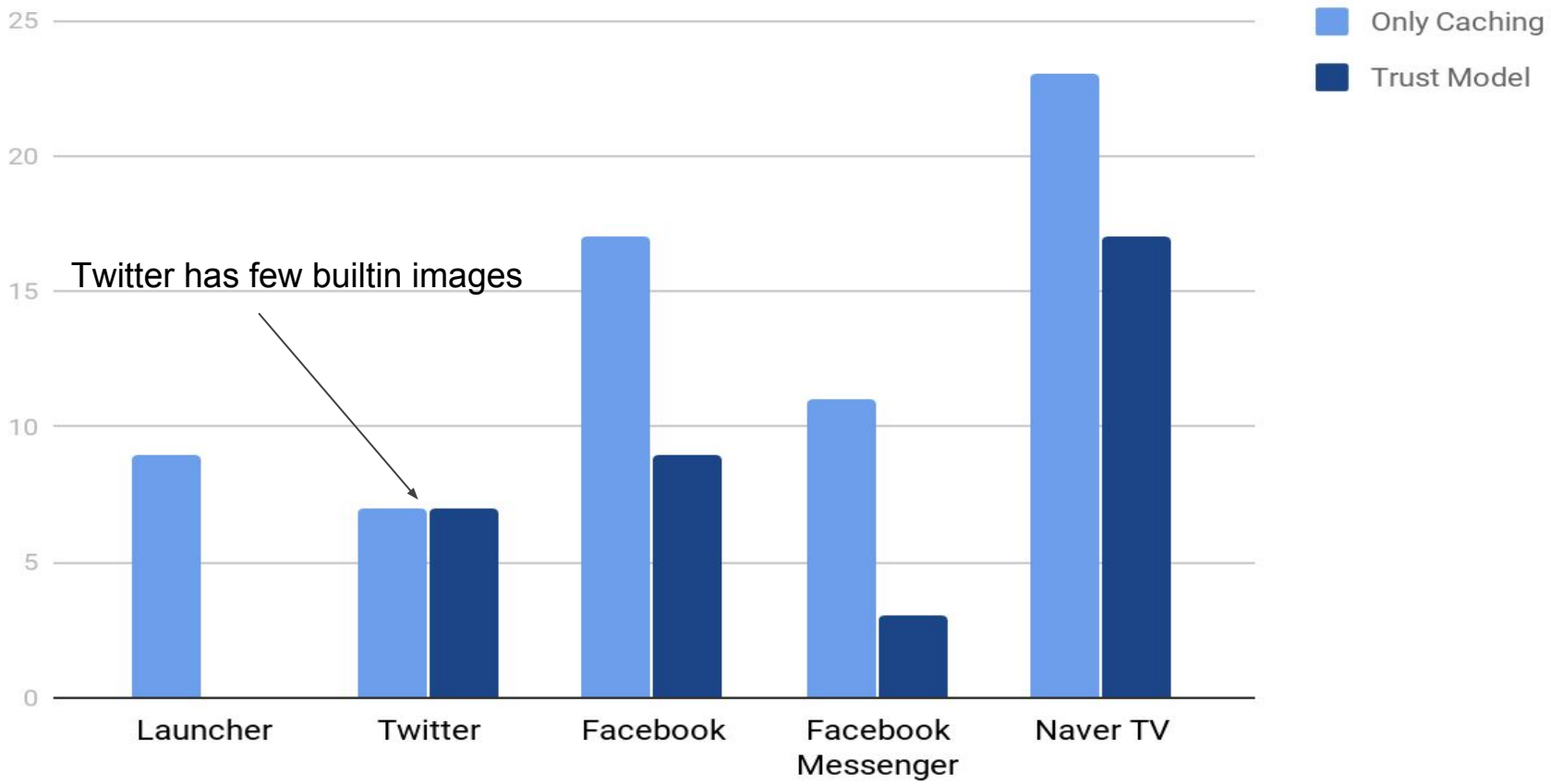


Lower is Better

* Workloads are not exactly same.

Performance evaluation (Only large images)

of requests

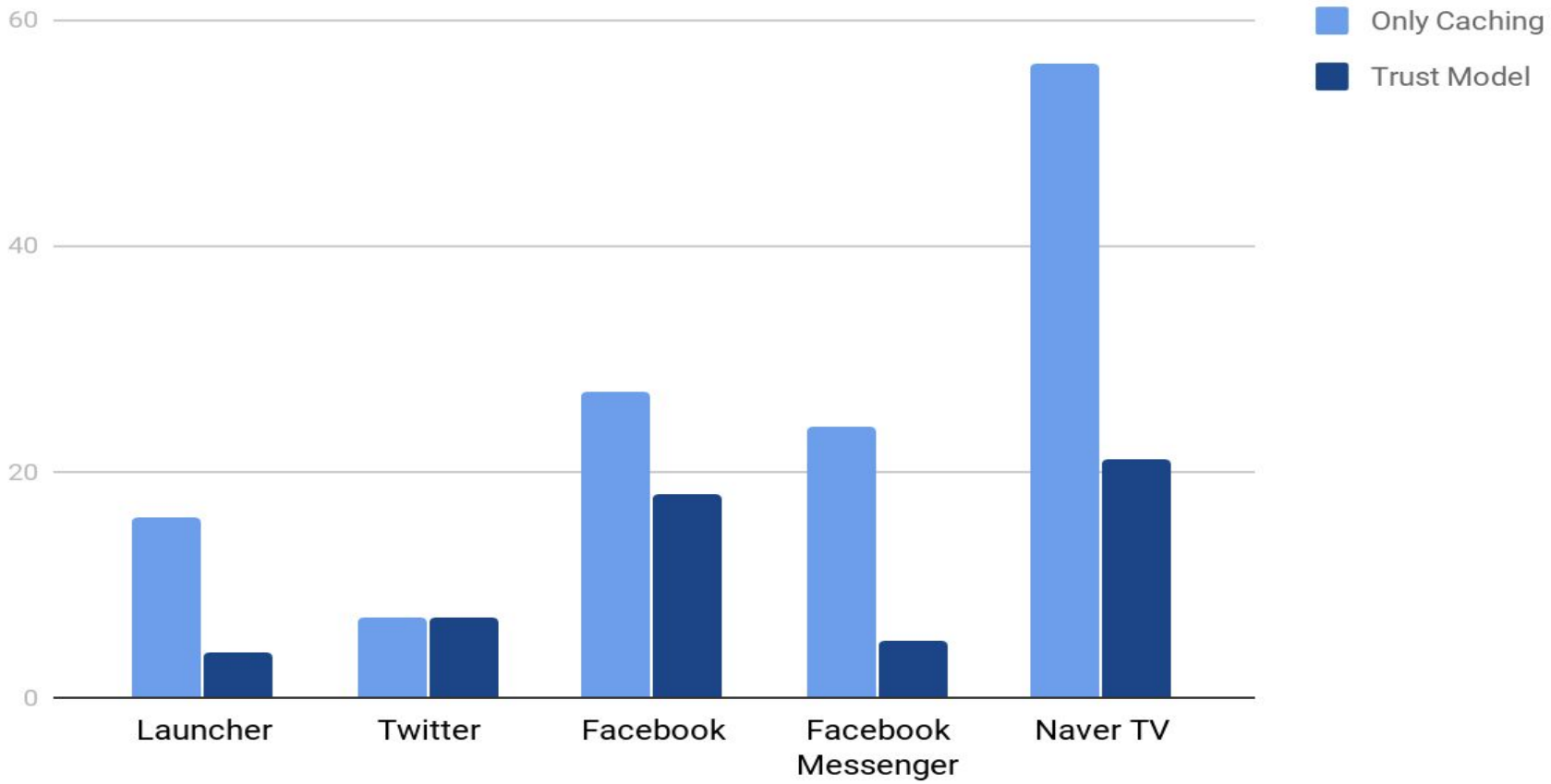


Lower is Better

* Workloads are not exactly same.

Performance evaluation (All images)

of requests

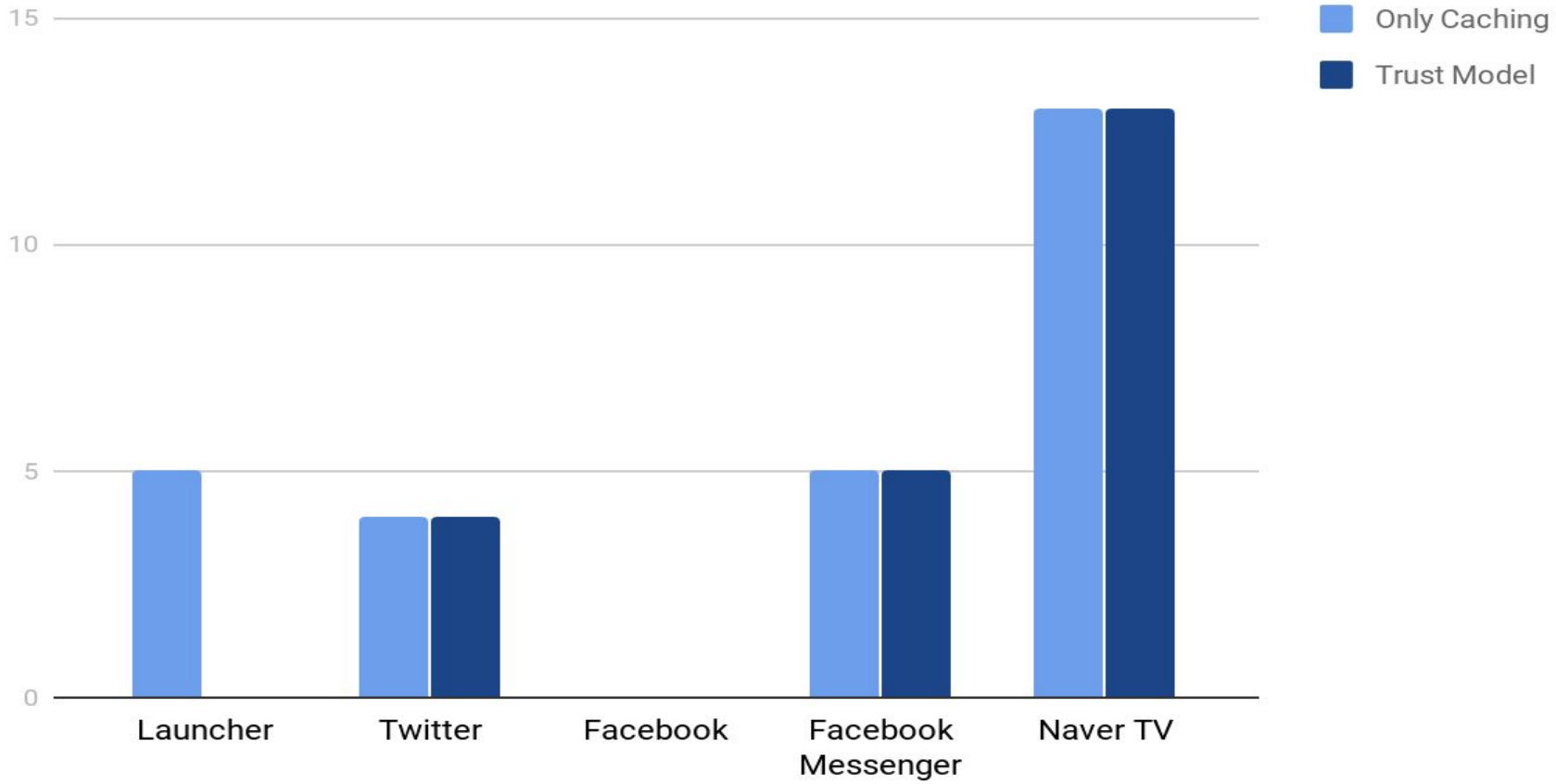


Lower is Better

* Workloads are not exactly same.

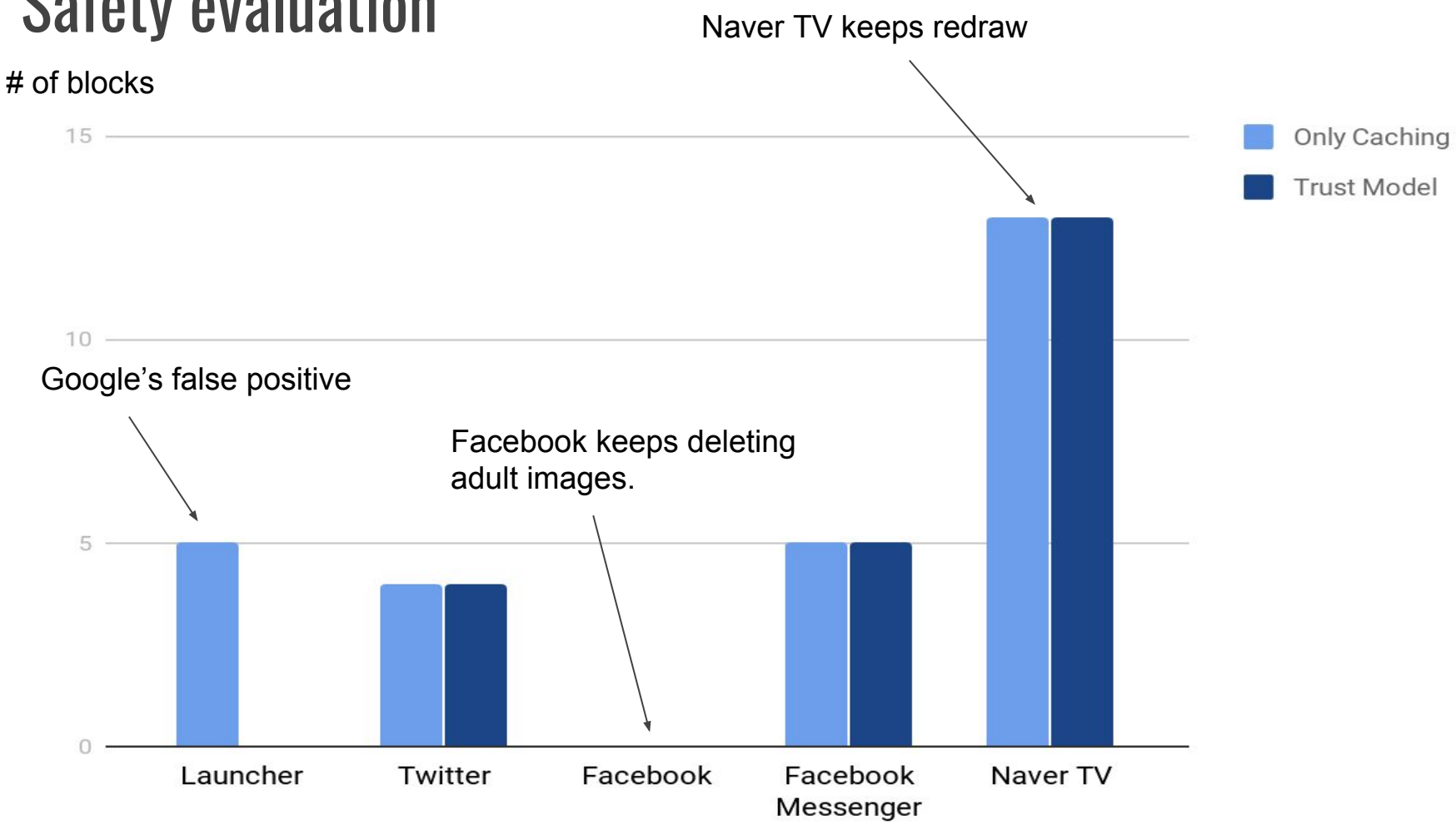
Safety evaluation

of blocks



* Workloads are not exactly same.

Safety evaluation



* Workloads are not exactly same.

Demo

Facebook & Facebook Messenger

Actors

Wonsup Yoon

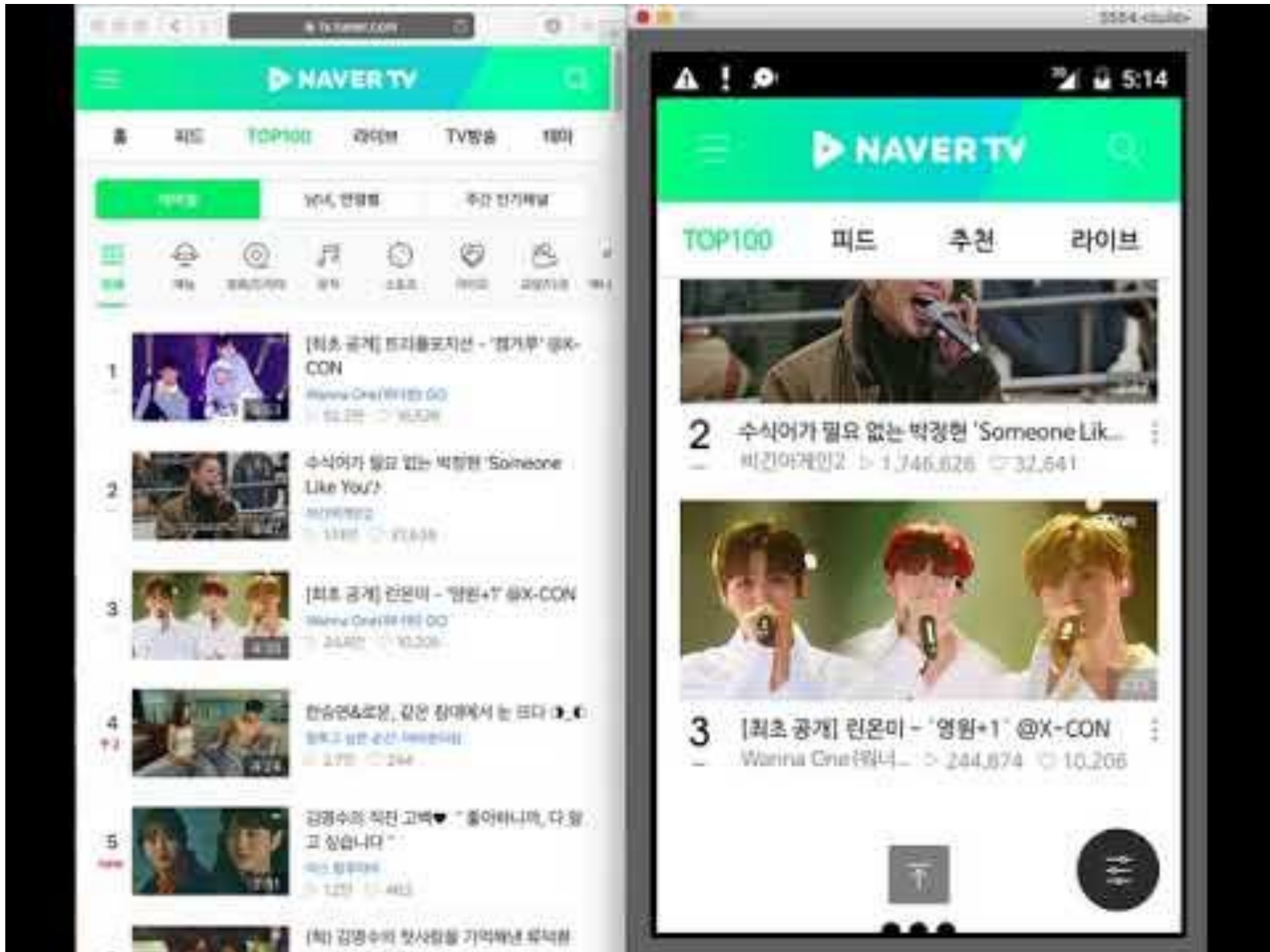
A Tiny Little Boy

Mark Kim

Creepy Pedophile
Sex Offender


Mark Kim
4:34
Google
Camera

Naver TV


NAVER TV
TOP100
피드
추천
라이브
2
수식어가 필요 없는 박정현 'Someone Lik...
3
[최초 공개] 린온미 - '영원+1' @X-CON

Thank You!