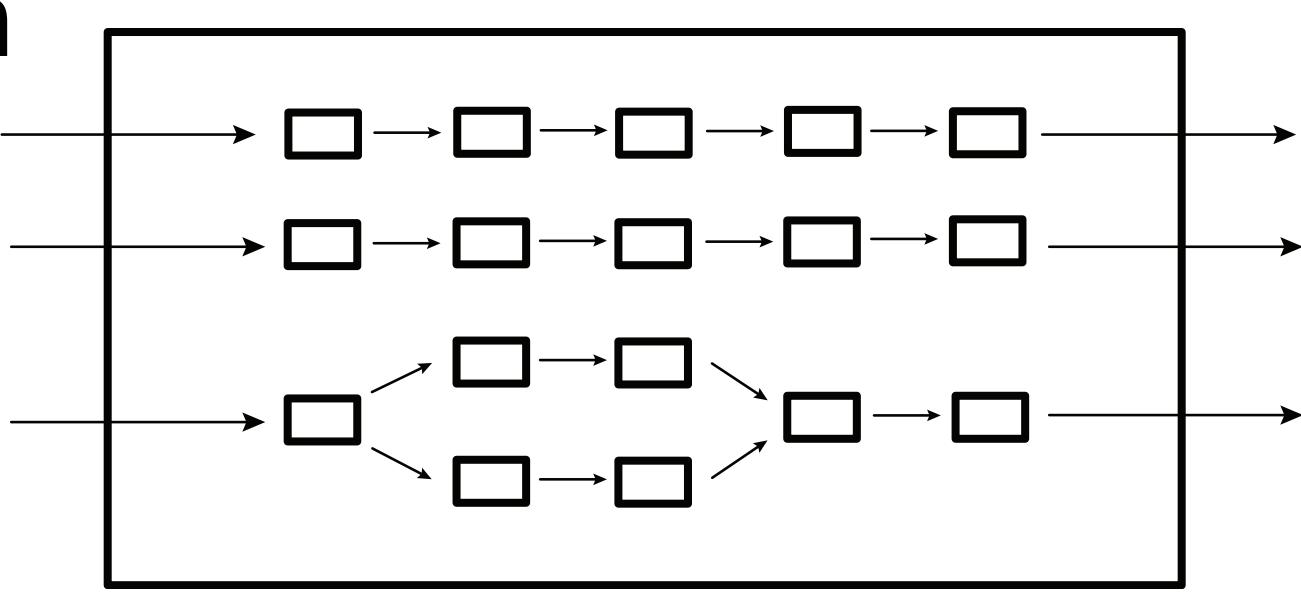
Modular Privacy Flows: A Design Pattern for Data Minimization

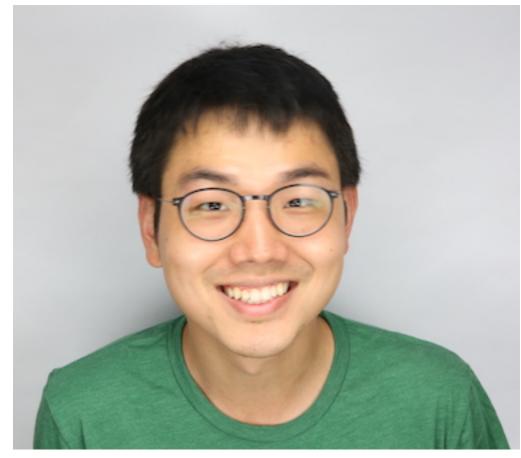




Haojian Jin Mar. 29, 2023



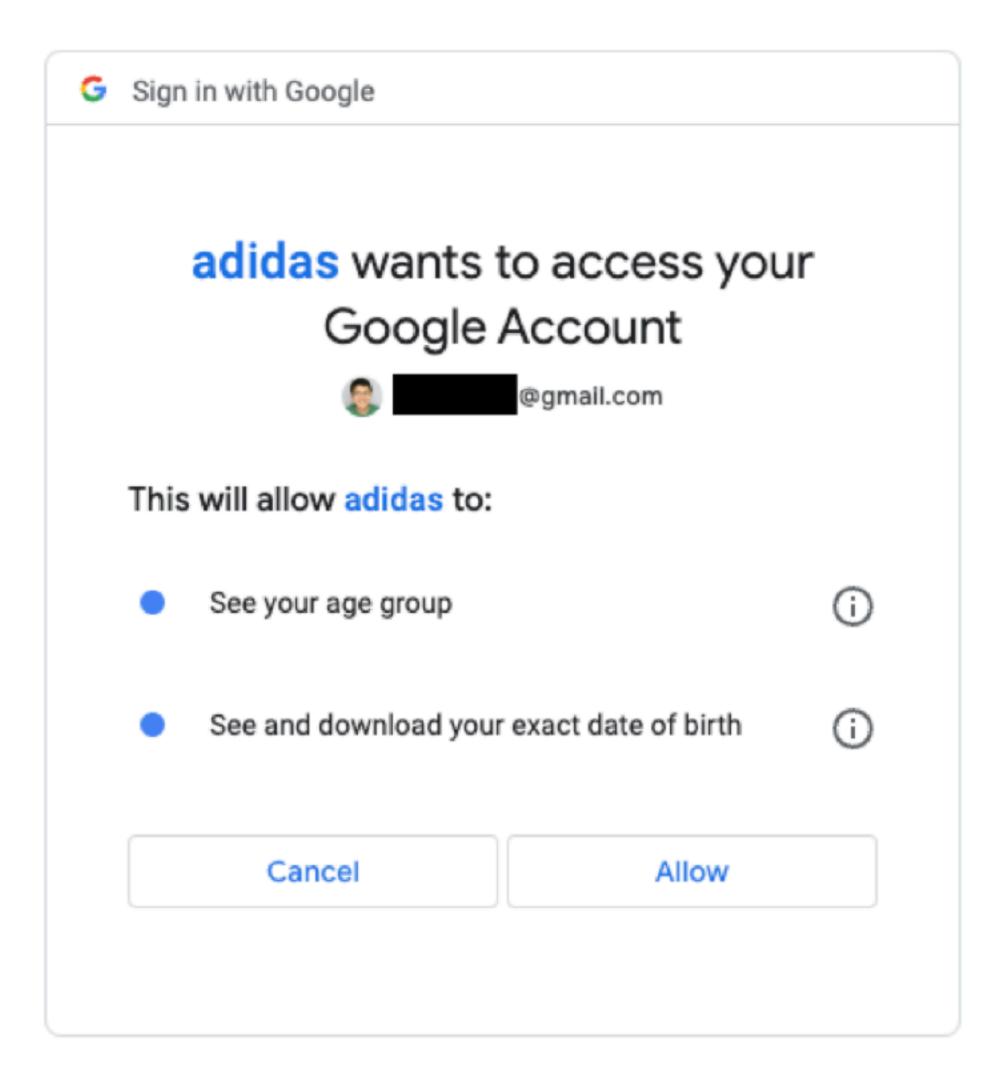
Haojian Jin (http://haojianj.in/) Asst. Prof @ UCSD-HDSI Data Smith Lab: people who design, implement, and use these systems. Ph.D. from CMU Human-Computer Interaction Institute Before Ph.D.: worked at Yahoo Research, ran a startup



- We study the security and privacy of data systems by researching the



Permissions



"Uber" Would Like to Use Your Location.

Uber picks you up exactly where you are. To start riding, choose "Allow" so the app can find your location.

Don't Allow

OK

Notice and choice

Informed decisions

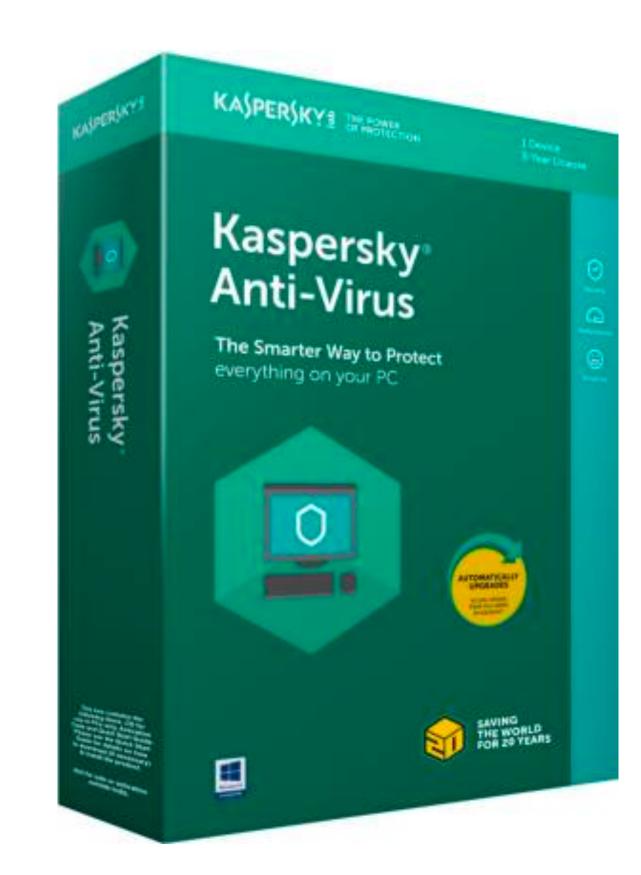
This website uses cookies to enhance user experience and to analyze performance and traffic on our website. We also share information about your use of our site with our social media, advertising and analytics partners. **Cookie Notice**

Customise your Preferences

Accept All



APIVISION - 🗟 - 🚳 - 🛍 - 📾 - 🕅 🍇 * 🕹 A - 🍛 C - 🍛 D - 🍛 C - 😹 F - 😒 I - 😒 I - 😒 N - 🤅	Z - S APIVISION.COM
🚰 Microsoft Corporation - Microsoft Internet Explorer	X
File Edi: Vew Favorites Tools Help	
😋 Back + 🕥 - 💽 😰 🏠 🔎 Search 🤸 Favorites 📢 Media 🥝 🍰 + 🥃 🚍 + 🗔 🍫 💓 🖃 🎼	🍃 Folders 🛛 🐥 - 🔏 📄 🗊 🐨 💙
Address 🙆 http://www.microsoft.com/	💌 🔁 Gn
🔄 Ġ 🖉 💽 🖉 🐨 💽 Google -	
🛛 🍸 🖉 - 👘 Search 🕞 📼 🔹 Sign in 🚳 My Yahoo! 👻 🧐 HotJobs 🔹 🖅 Yahoo 🔹 👘 Shopping 🔹 🎼	Games 🔹 🖂 Yahco! Mail 🔹 🏈 News 🔹 🛛 😕
🖉 🐠 🖬 Search News 🎽 Highlight	🔬 Email this page to a friend 🚺 🛞 JeevesLinks
AP Option: Soltware - P 🗄 D/L 0 liles 🕰 🖉 - 🛃	more about the new DAF 1
AimAtšite 🔹 🙀 History 🧟 Favorites 🔬 This Page 🦚 Online 👻 🏪 Tools 👻	
🖌 🕐 🕐 🕐 🖉 Search 🐁 Channels 🕶 My Praize 🔹 🖂 Praize Mail 🔹 🖉 M 🔹 💼 Favo	prites 🕶
🐨 MapStan.net 🔹 💿 🐨 🐨 🐨 🐨 🐨 My Plan 🖉 Discover 🥞 My Plan 🖉 Daily Plan 🆓 Best Sites	
Search	
Addresses.com Search Type: Email Search 💌 Name/Business: 💽 Locale: 💌 Search 🔞 Search	n Tips
TEOMA -	🖞 🌽 Highlight 🛛 🛃 Email this page to a friend
Microsoft 😵 Software) 🕼 Tech News) 📢 Windows 2000 🌑 Intemet) 🗐 Computer Hardware	Search 🔍 Skins Web 🚫
🖌 🖌 🖌 🛃 🗹 🖉 🖉 Do Search 📲 Highlight	
🔄 🛃 🔹 🔁 Search eBay 🔹 – 🙀 MyleBay 😪 Bid Alert 🔹 🧥 Watch Alert 🔹 🥓 Top Picks 🔹 🕐 Help 🔹	
📔 🗚 🖉 🖉 🖉 Al Lsers Services 🕞 🖉 Abilene Area Sites 🔹 🃁 Al Recommended Links 🔹	
Image: Construction of the second	rafx • Metatec International
🔹 🖉 bearch 👻 📑 News 🕶 Email 👻 🚰 Hill 👻 🗇 Popub 🛛 💽 M-3 📧 💽 💌	Programs
Links 👸 Customize Links 👩 Free Hotmal 🔞 Windows 🙆 Wincows Media	
All Products	Support Search Microsoft.com Guide 📥
Microsoft	
Microsoft Home MSN Home Subscribe Manage Your Profile	
Go Go Control	
Product Families Server 2003 RC2	
i www.metatec.com/	
🛃 Start 🧑 🚱 🖗 Microsoft Corporation	🥜 🔯 🛃 ờ 🏧 🍓 🛦 🌽 N 🕜 8:11 AM



App Stores

"changed how software development worked, and expanded the **number of people who could comfortably, safely use a computer** from a few hundred million to a few billion."

[1] Benedict Evans, App stores, trust and anti-trust



Technical idea #1 Putting apps in a sandbox

Apps can only do things that Apple allows and cannot ask (or persuade, or trick) the user for permission to do 'dangerous' things.



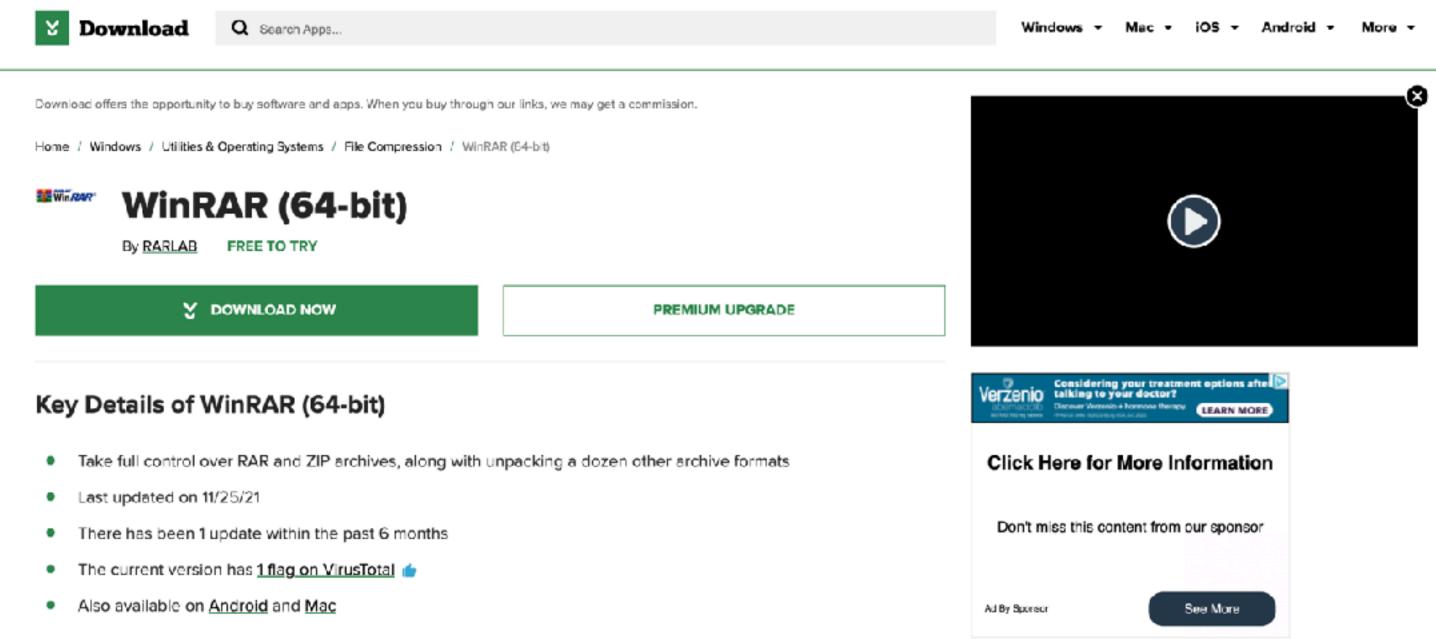




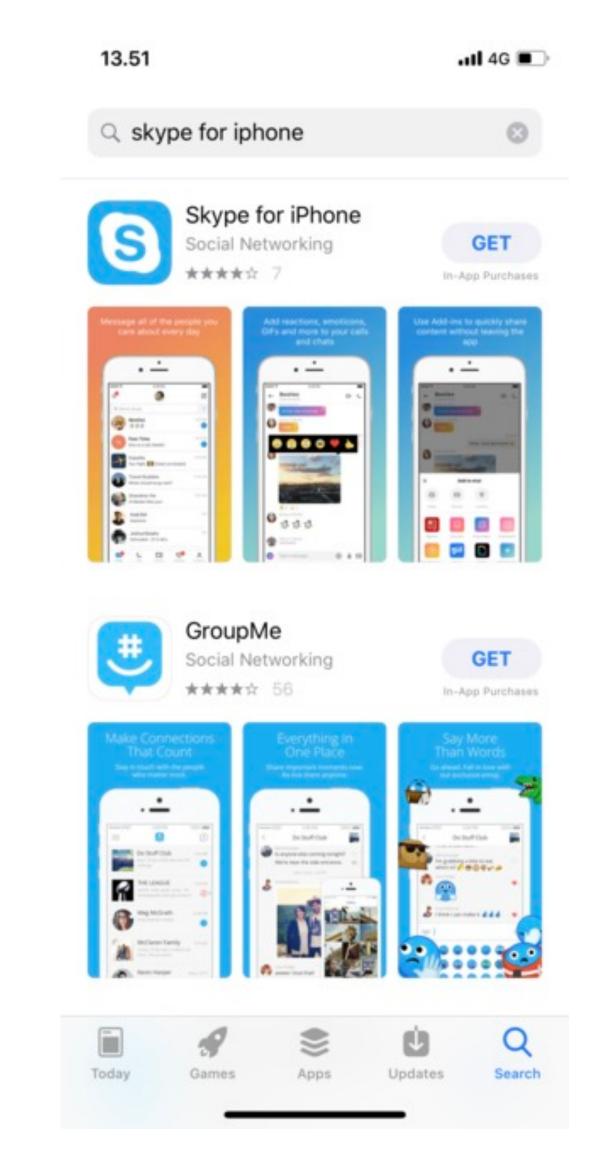
- Would this break my phone?
- Would this run my battery down?
- Steal my bank details?



Technical idea #2 Distributing software through a centralized app store



Archive name and parameters	? X	Downloads (evaluation -		- 0	5
General Advanced Options Files	Backup Time Comment	Add Extract To	I 🛄 🗑 🔍	Wizard Info	
Downloads.rar	~	+ C:\Users\winn	NDownloads		_
Default Profile	Update mode	Name	Size Type	Modified	
Profiles	Add and replace files	↓ .	System Folder		
		convert	File folder	5/22/2019 10:23 AM	
Archive format	Archiving options	older_downloads	File folder	5/22/2019 9:48 AM	
		Backup_data.rar	10,573,324 WinRAR archive	5/22/2019 10:14 AM	
BAR ORARA OZIP	Delete files after archiving	Documents.rar	4,587,396 WinRAR archive	5/22/2019 10:10 AM	
CETTINGS	Create SFX archive	Downloads.rar	2,964,465 WinRAR archive	5/22/2019 10:03 AM	
SETTINGS	Create golid archive	music.zip	15,163,290 WinRAR ZIP archive	e 5/22/2019 10:22 AM	



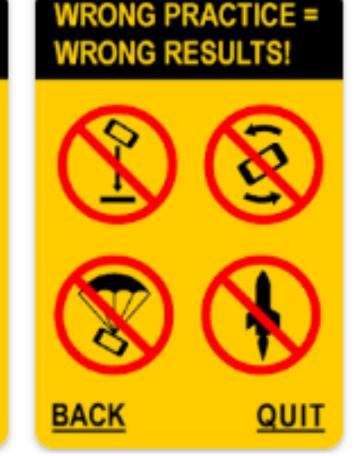
Technical idea #2 Distributing software through a centralized app store

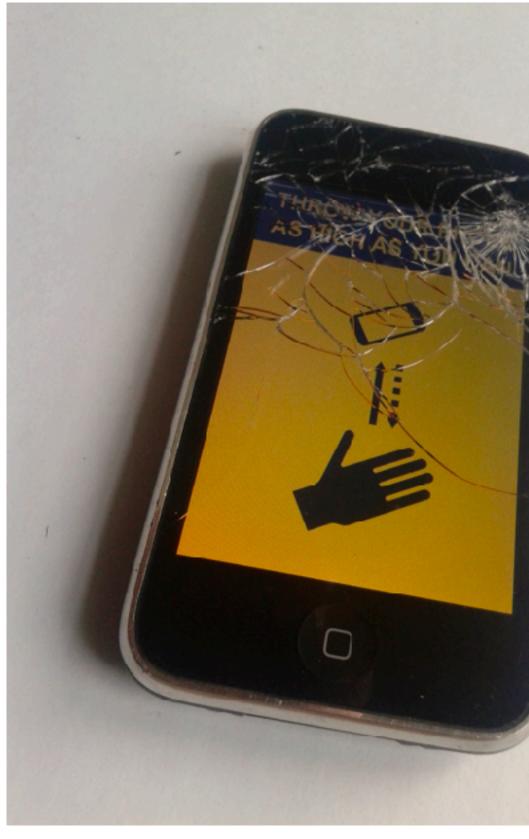


S.M.T.H.: Send Me to Heaven

WORLD TOP 10

1.	ME	0.63m
	COCOFOX	1.76m
	ABRAHAM	1.63m
	VICTOR	1.56m
	RICHARD	1.39m
	ERZIKA	1.39m
	MARIUS	1.13m
_	MARTIN	1.12m
	NIN	1.10m
	MARTIN	0.93m
0.	PETER	0.78m
PL	.AY	







Issues around idea #1 Putting apps in a sandbox

Apps can only do things that Apple allows and cannot ask (or persuade, or trick) the user for permission to do 'dangerous' things.

- What are 'dangerous' things?
- How can we trust Apple?

.

- How can we detect if apps trick the user?

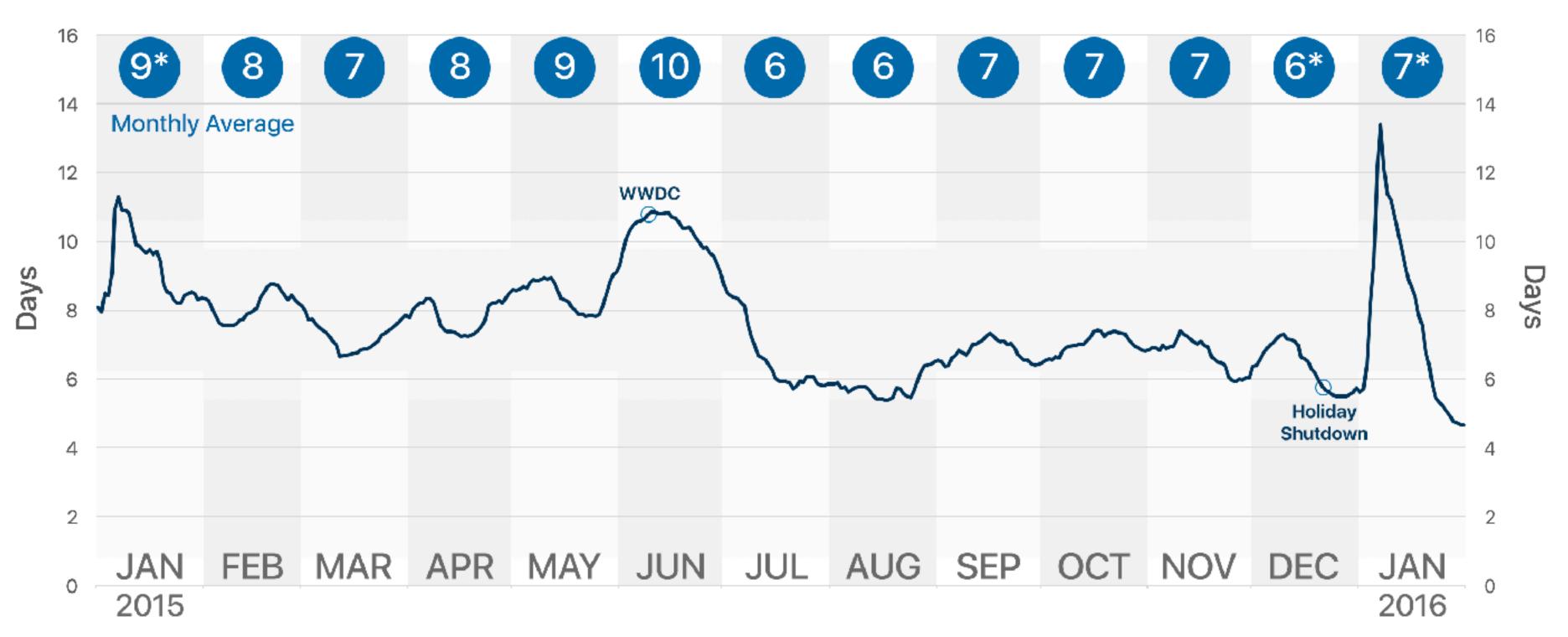
Issues around idea #1 After 15 years - still iterating







Issues around idea #2 Distributing software through a centralized app store



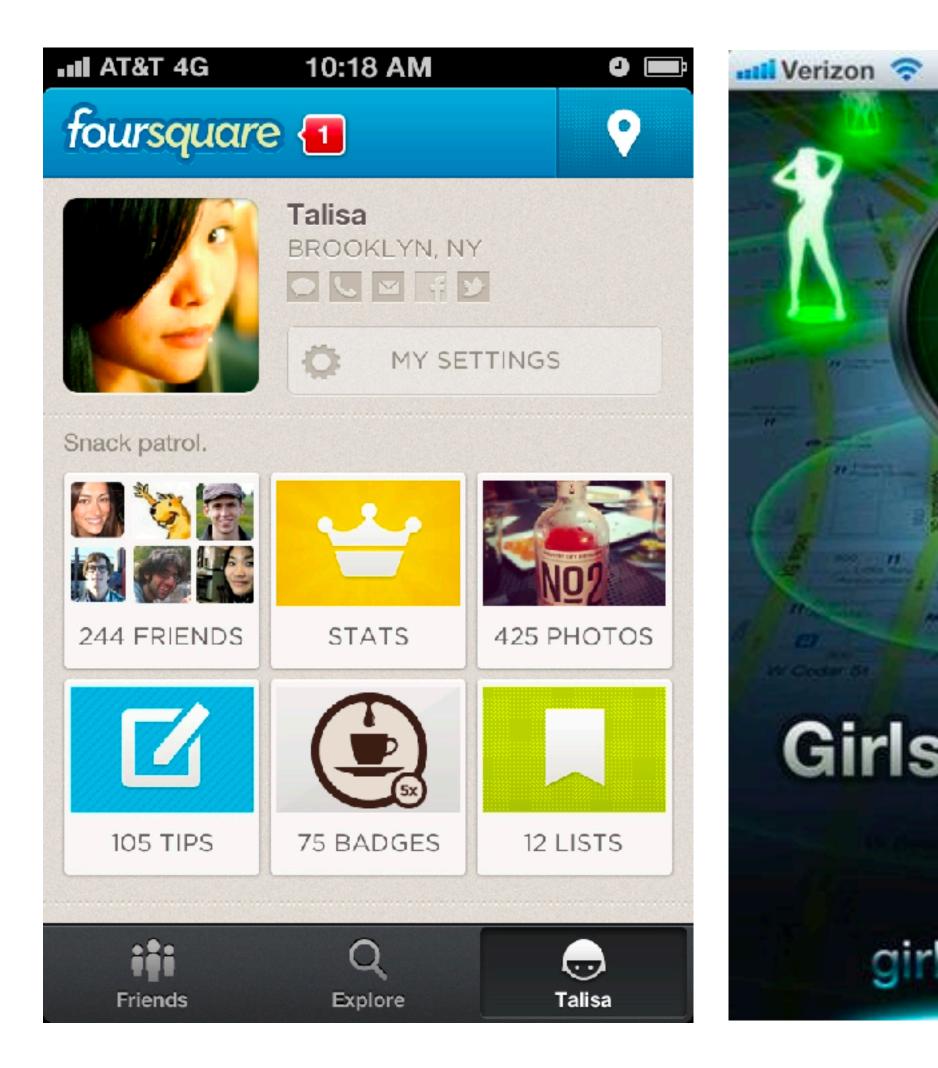
MacStories.net | Graham Spencer

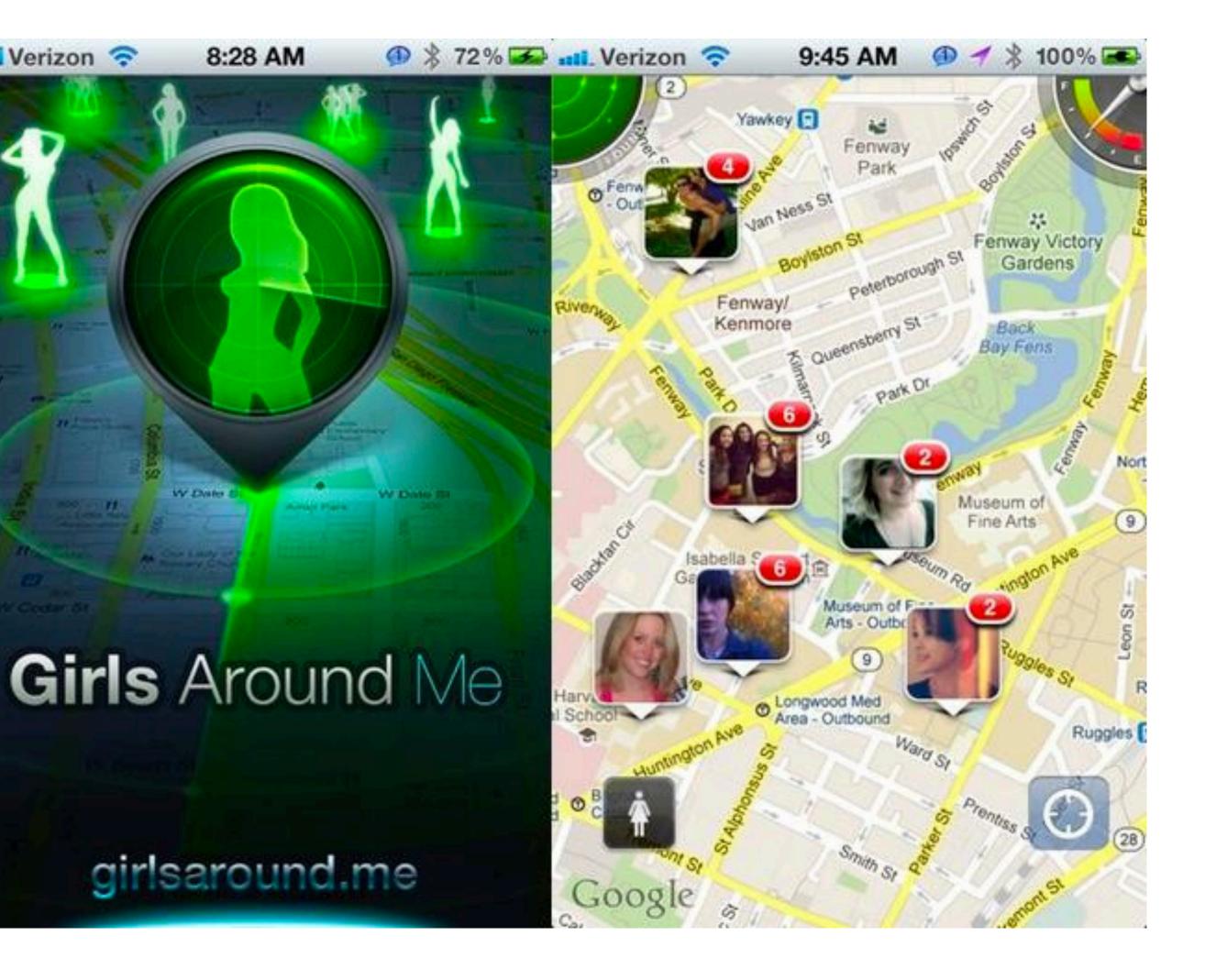
iOS App Store Review Times

- Unofficial data, courtesy of AppReviewTimes.com
 - (14 day trailing average)

* affected by Holiday Shutdown | Datapoints: 6,999

Issues around idea #2 Software often has a cloud component.





Linux systems

💋 🗊 aditya314@ubuntu: ~									
aditya314@ Desktop Documents Downloads aditya314@ total 52	e) ge l	xamples.des gf.txt istfile		new	ic one tures	Тег	npla	ates	cument
drwxr-xr-x drwxr-xr-x drwxr-xr-x -rw-rr -rw-rw-r	2 2 1 1	aditya314 aditya314 aditya314 aditya314	adity adity adity adity	a314 a314 a314 a314 a314	4096 4096 8980 0	Mar Mar Mar Mar	5 5 5 5	01:21 01:21 01:05 03:53	Desktop Documen Downloa example ggf.txt listfil
drwxr-xr-x drwxrwxr-x drwxr-xr-x drwxr-xr-x drwxr-xr-x -rw-rw-r drwxr-xr-x -rw-rw-r	2 2 2 2 1 2 1	aditya314 aditya314 aditya314 aditya314 aditya314 aditya314 aditya314 aditya314 aditya314	adity adity adity adity adity adity adity	a314 a314 a314 a314 a314 a314 a314 a314	4096 4096 4096 4096 4096 0 4096	Mar Mar Mar Mar Mar Apr Mar	5 5 5 27 5	01:21 03:53 01:21 01:21 01:21 02:55 01:21	Music new one Picture Public Templat Untitle Videos
0000905140									

Videos xyz.txt

op ents oads les.desktop kt ile

e es

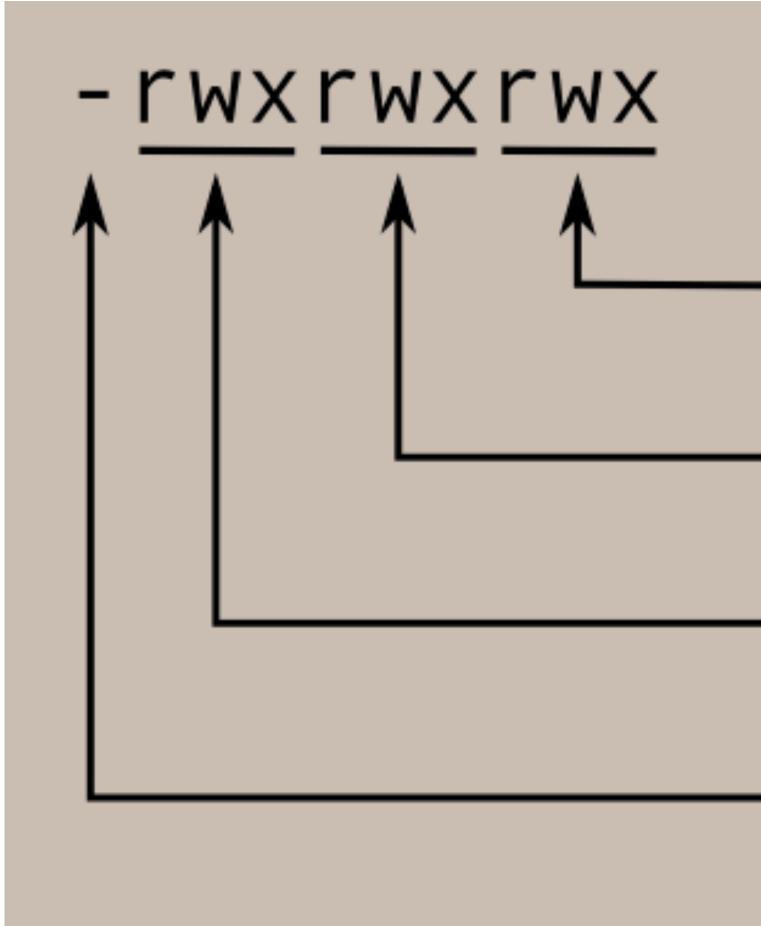
i<mark>tes</mark> .ed Document

t

• Users

• File permissions

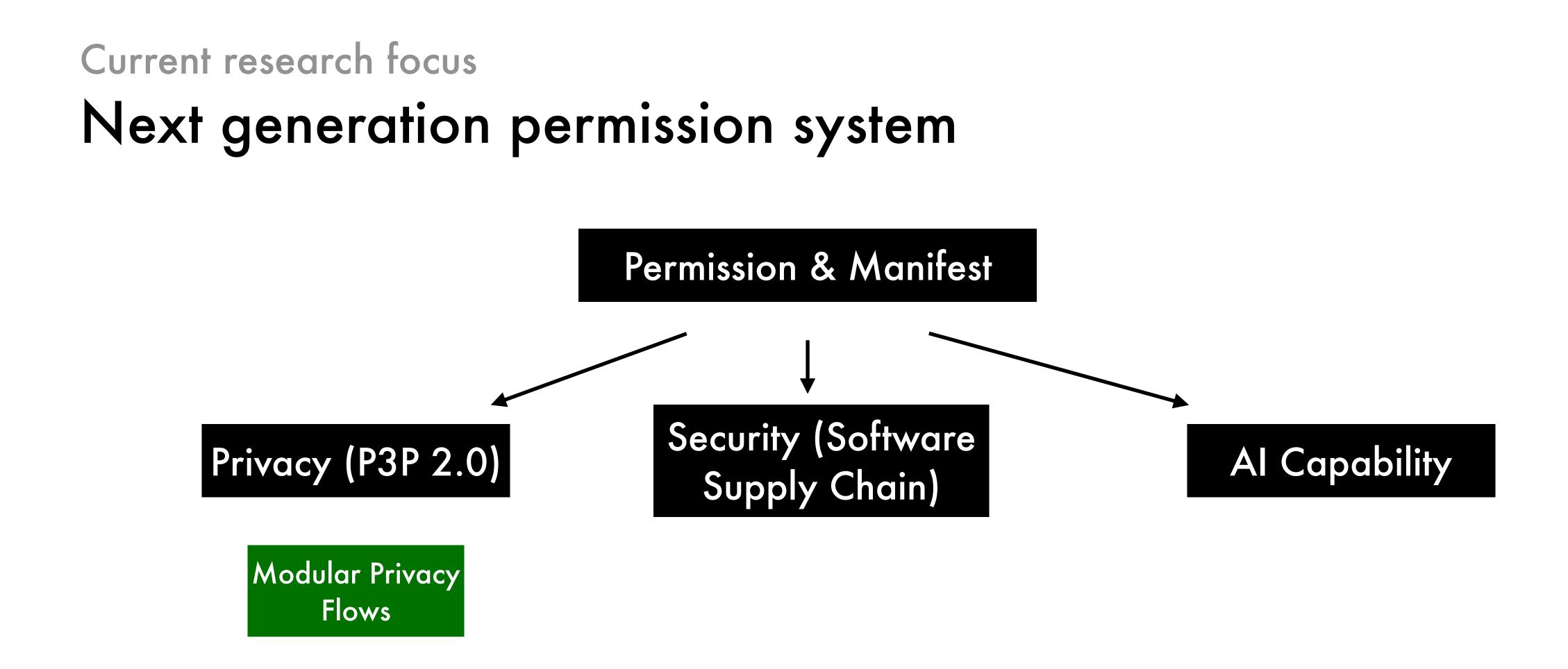
File permissions



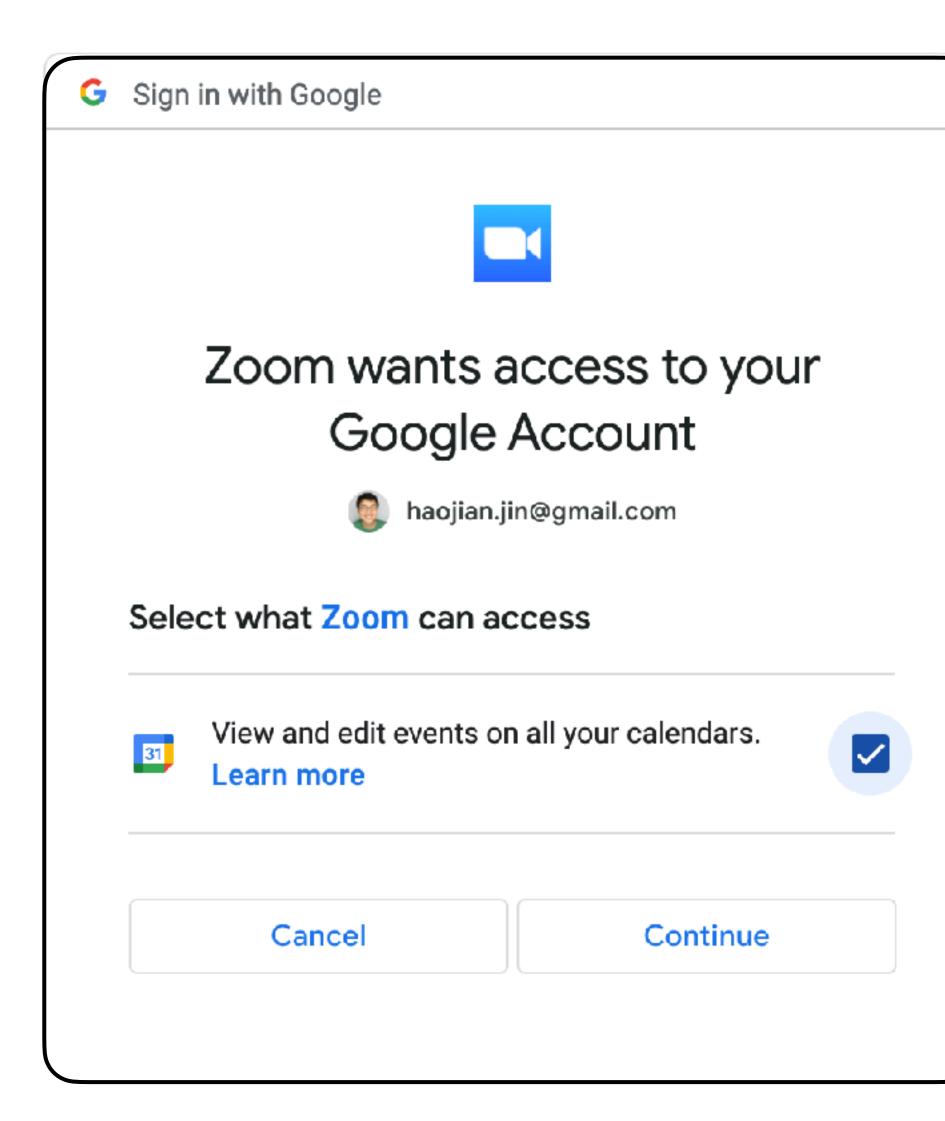
- Read, write, and execute permissions for all other users.
- Read, write, and execute permissions for the group owner of the file.
- Read, write, and execute permissions for the file owner.

File type:

 indicates regular file d indicates directory



Zoom accesses all your calendar events continuously!



Calendar events that contain https://zoom.us/xxxx

Uber wants to see all your emails.

Stay organized for your next trip

Uber Travel

9:41

←

8

Ð

.ul 🗟 🛙

+

Uber has partnered with AwardWallet to automatically import and organize your hotel, flight, and restaurant reservations into one view

We respect your privacy

We'll only import travel-related emails to create trip itineraries

Let us do the work for you

If anything changes, we'll update your itineraries so you can stay organized

By linking your Gmail account, you will agree to let Uber and AwardWallet import your travel-related emails in accordance with Uber's Privacy Notice and AwardWallet's Privacy Policy.

🬀 Sign in with Google

Principle of data minimization

"Personal data shall be limited to what is they are processed."

necessary in relation to the purposes for which

- GDPR, Article 5 (1) (c)

Principle of least privilege

to perform its function.."

"A security architecture should be designed so that each entity is granted the minimum system resources and authorizations that the entity needs

Google APIs - All-or-nothing binary permissions

Scope

https://www.googleapis.com/auth/calendar

https://www.googleapis.com/auth/calendar.read

https://www.googleapis.com/auth/calendar.even

https://www.googleapis.com/auth/calendar.even

https://www.googleapis.com/auth/calendar.sett

https://www.googleapis.com/auth/calendar.addo

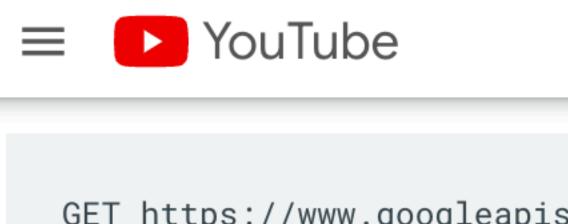
	Meaning
	read/write access to Calendars
donly	read-only access to Calendars
nts	read/write access to Events
nts.readonly	read-only access to Events
tings.readonly	read-only access to Settings
ons.execute	run as a Calendar add-on

https://developers.google.com/calendar/api/guides/auth



Program data transformation functions using chainable operators



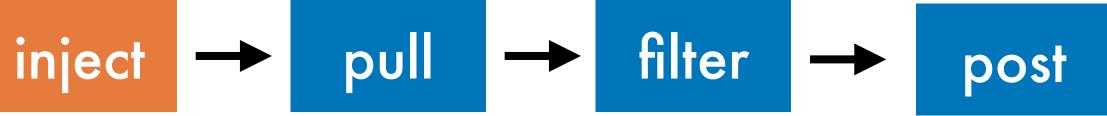


Operator-based APIs



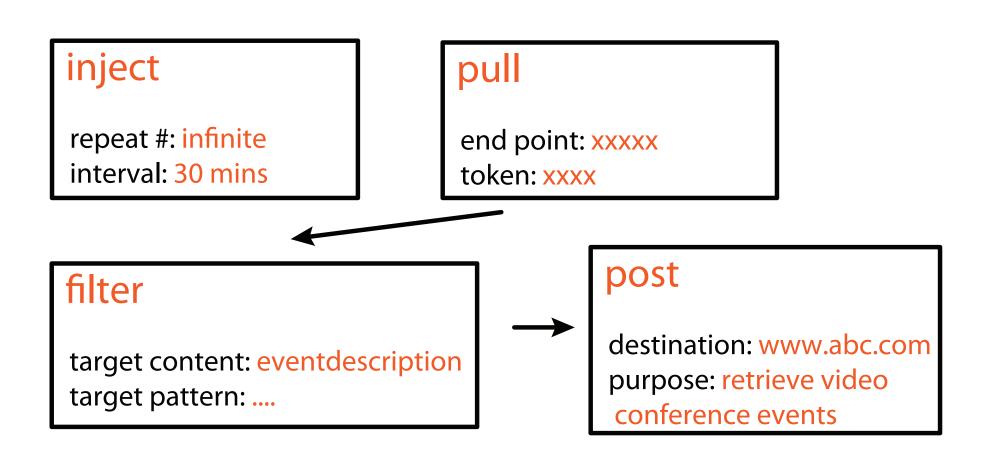
Ð

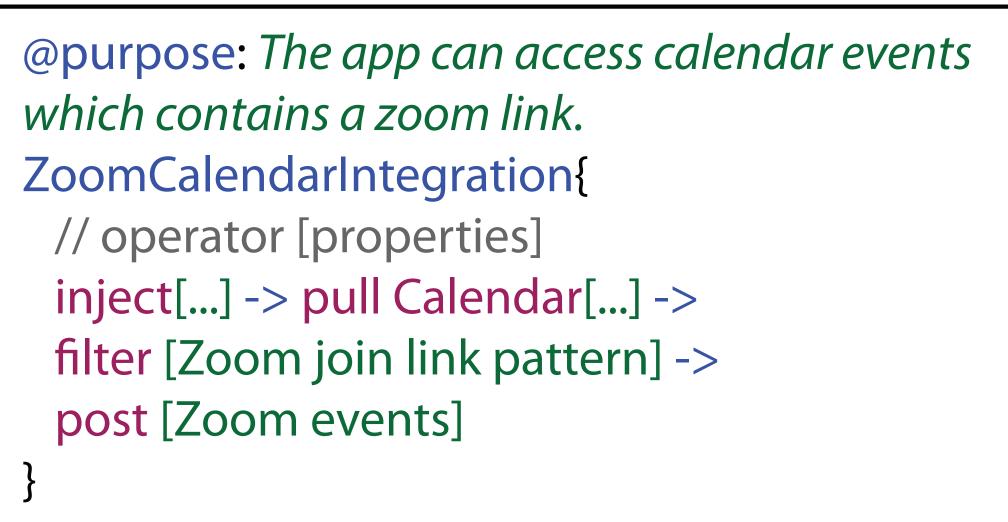
GET https://www.googleapis.com/youtube/v3/playlists





A text-based whitelist manifest (i.e., program representation)







System builders



Offer a set of operators as the API



Execute the manifest using preloaded implementations

Developers

@purpose: The app can access calendar events
which contains a zoom link.
ZoomCalendarIntegration{
 // operator [properties]
 inject[...] -> pull Calendar[...] ->
 filter [Zoom join link pattern] ->
 post [Zoom events]

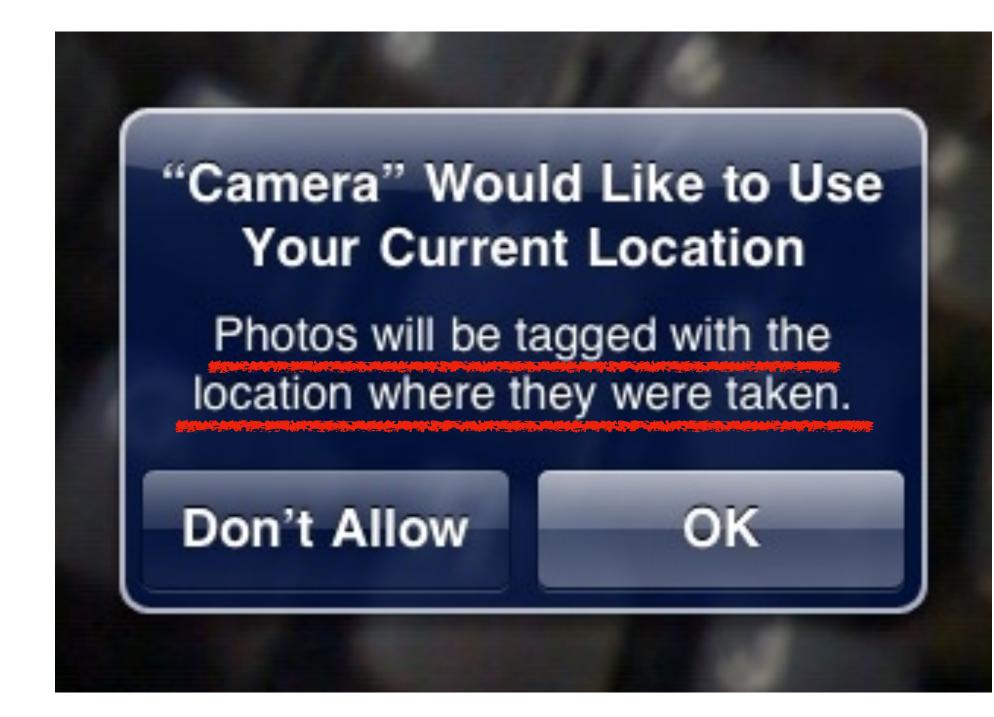
Author a manifest by connecting operators



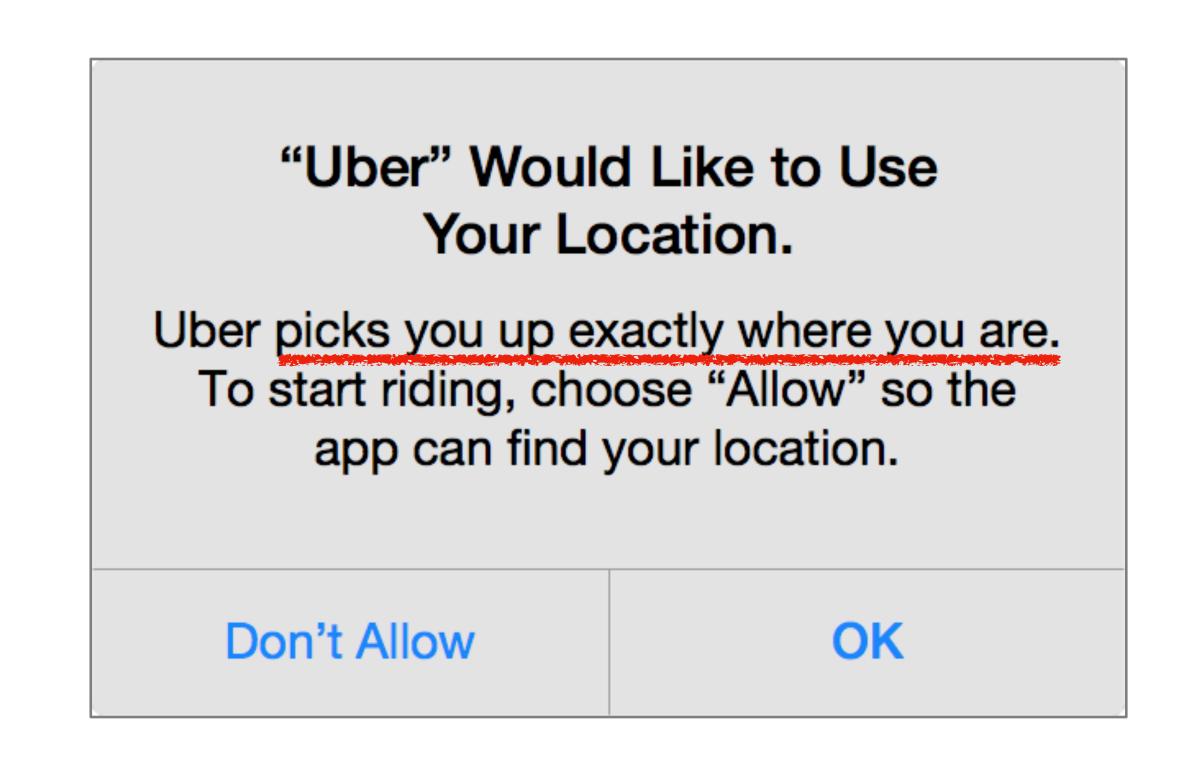
Talk outline

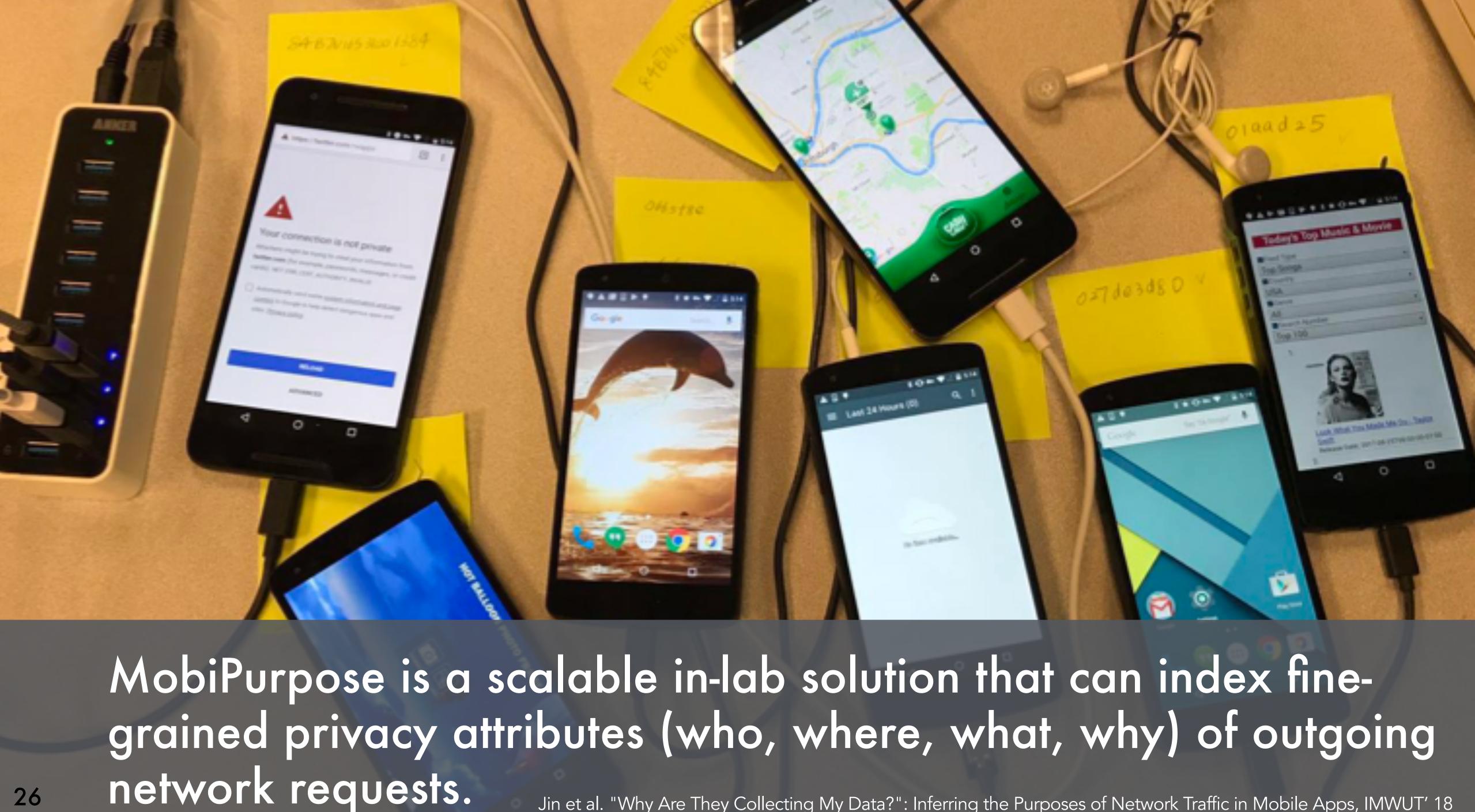
- 1. Modular Privacy Flows (MPF) in a Nutshell
- 2. Why MPF
- 3. How MPF
- 4. When and when not MPF
- 5. Future Work

Purpose strings.



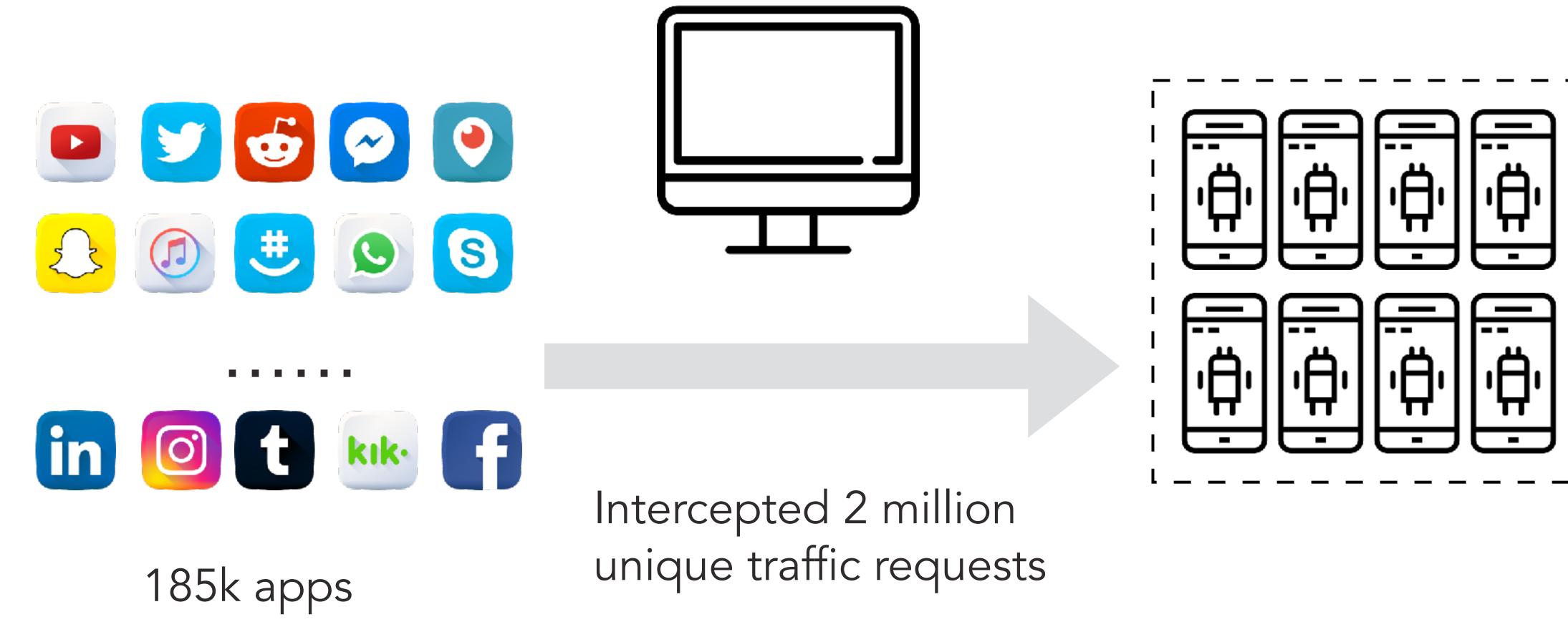
arbitrary text, manually annotated, hard to validate/assess.





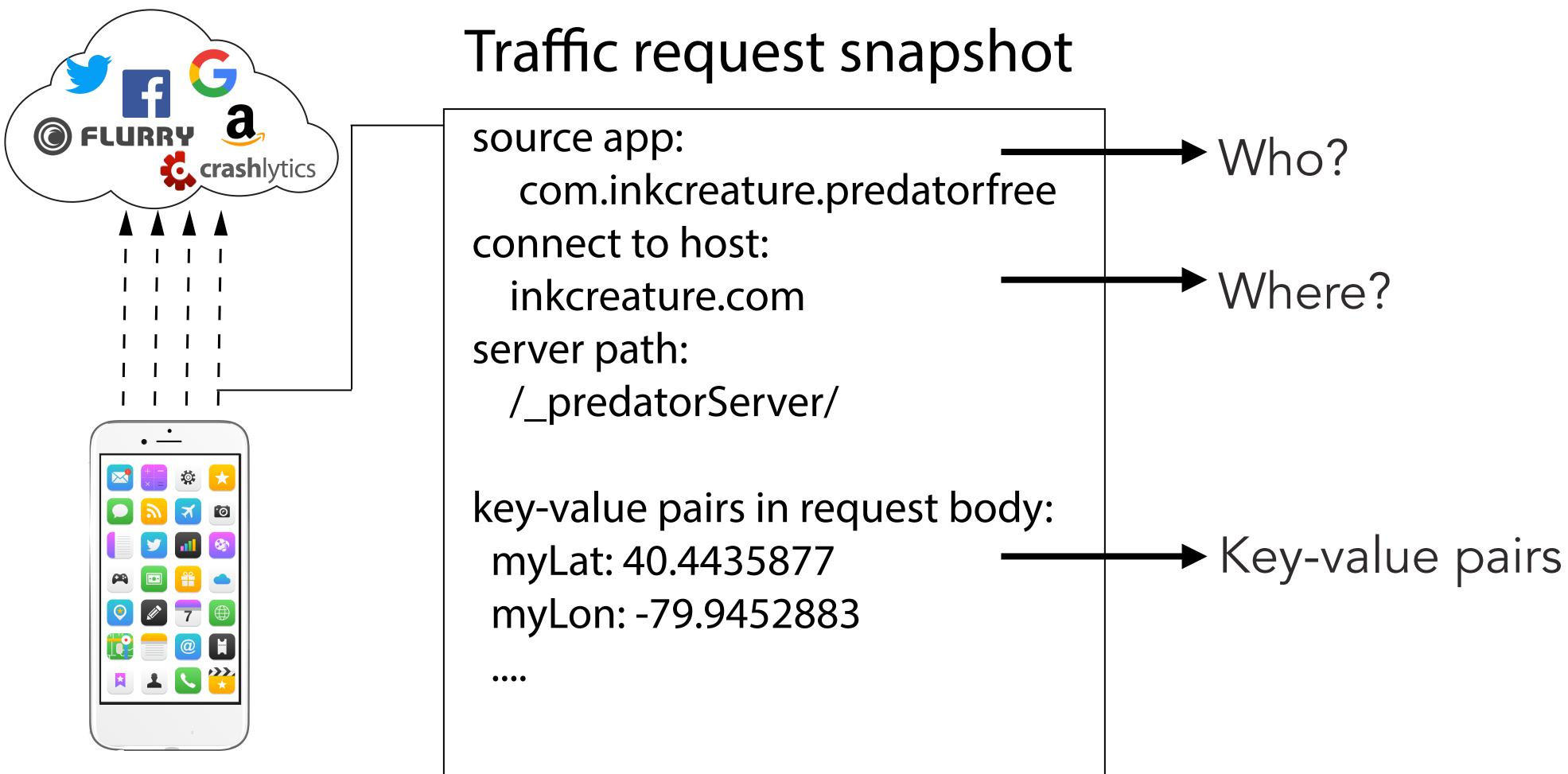
Jin et al. "Why Are They Collecting My Data?": Inferring the Purposes of Network Traffic in Mobile Apps, IMWUT' 18

MobiPurpose - network tracing





MobiPurpose - network tracing





Traffic request snapshot

source app: com.inkcreature.predatorfree connect to host: inkcreature.com server path: /_predatorServer/

key-value pairs in request body: myLat: 40.4435877 myLon: -79.9452883

 $\bullet \bullet \bullet \bullet$

2,008,912 unique traffic requests from 14,910 apps

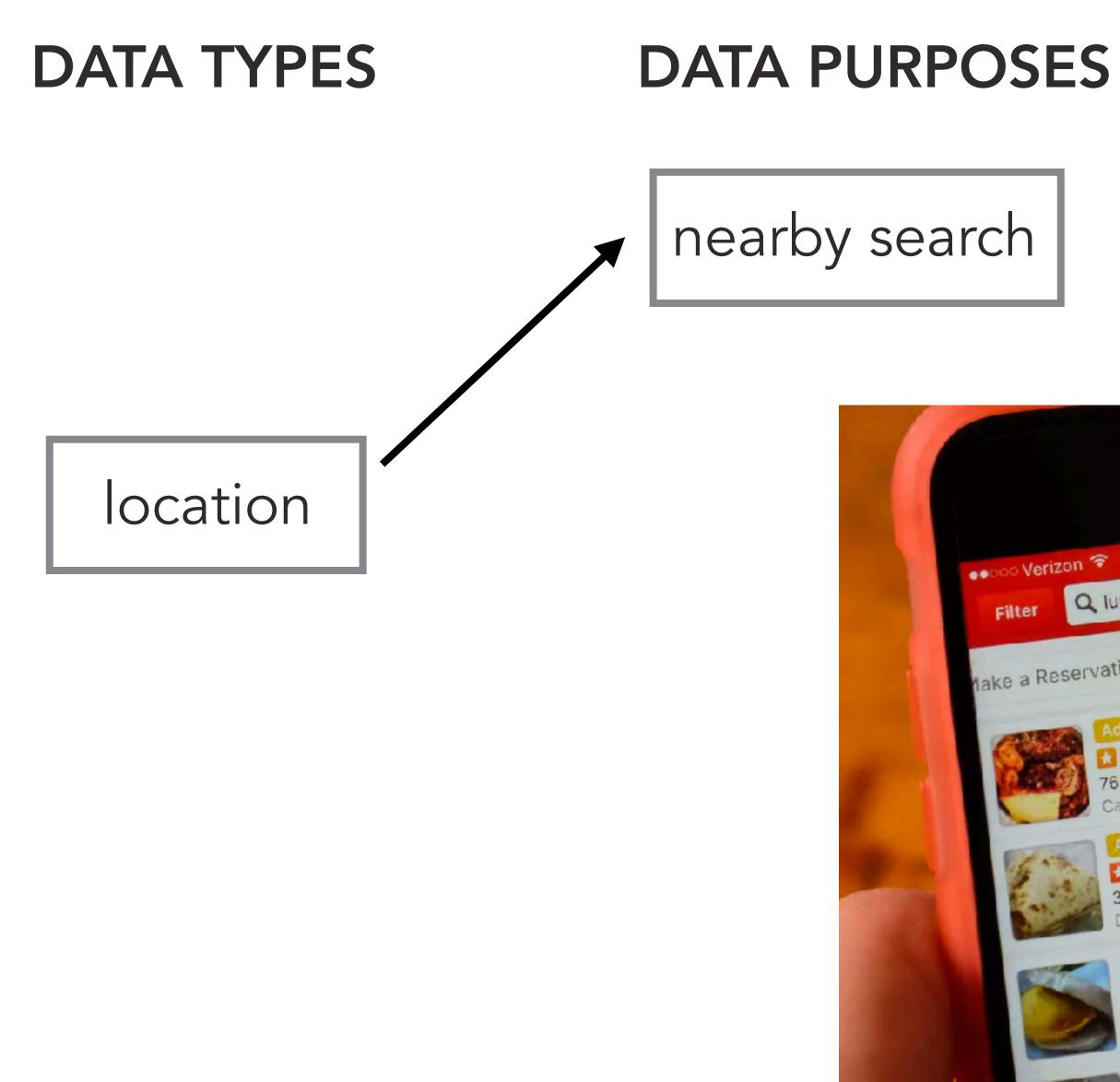
contacting

12,046 unique domains 302,893 unique URLs

We publish the dataset at: <u>http://bit.ly/purposedata</u>

Traffic Data stats



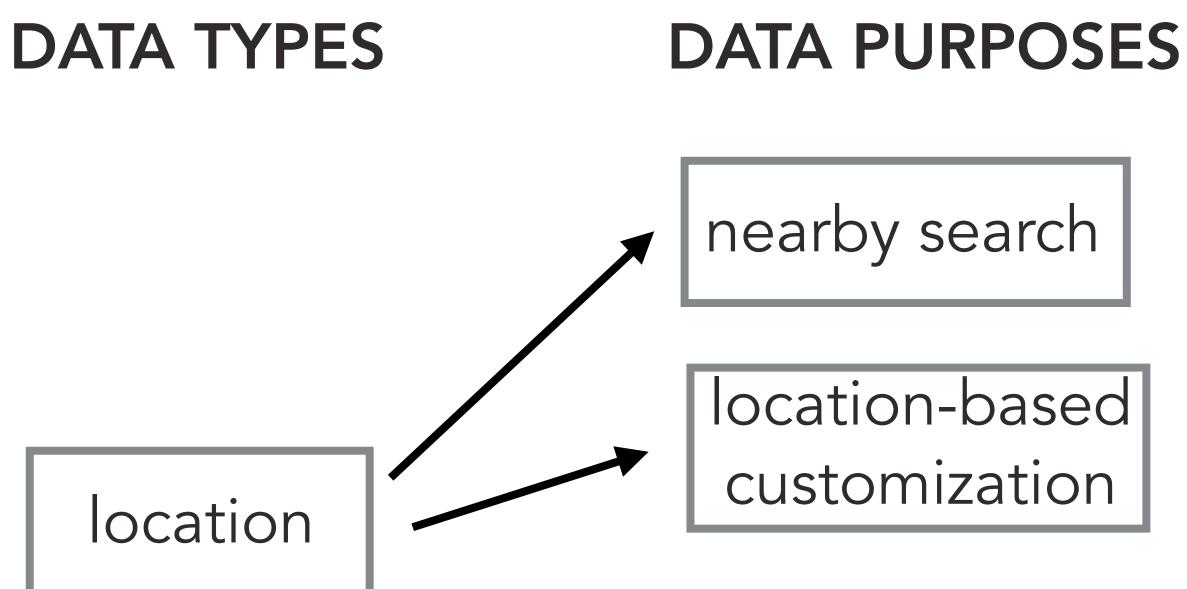


EXAMPLES



	9%
12:49 PM	Map
Filter Q lunch Current Location	
Filter Cluter Take a Reservation PokéStop Nearby Outdoor Set	
lake a Reserver	0.9 mi \$
Ad Gloria's Caribbean Cuisine Ad Gloria's Caribbean Cuisine 196 Reviews 764 Nostrand Ave, Brooklyn	100000
(hka2)	1.7 mi
Contana Deli & Grocery	
A Contral Aver	
Dalis Sandwichter	0.2 mi
Luquise Baking	\$
1. Tropical House L 30 Reviews Constant Ave, Crown Heights	and the second se
267 Schenectady Ave, Crown Heights	and the second sec
267 Schenectory Bakeries, Caribbean	0.6 mi
CreanS.	5
2. Daddy Greens 2. Daddy Greens 189 Reviews 189 Reviews Status and Stuyvesant	
2. Daddy 189 Reviews 189 Reviews 352 Malcolm X Blvd, Bedford Stuyvesant 352 Malcolm X Blvd, Coffee & Tea	
352 Malcolm X Blvd, Bedicie Sandwiches, Pizza, Coffee & Tea	0.8 mi
Sandwichter	\$\$
3. Manny's 158 Reviews	
3. Manny's 3. Manny's 158 Reviews 158 Reviews Stuyvesant Conffee & Tea	
a Detriben Ave, Beuter & Tea	

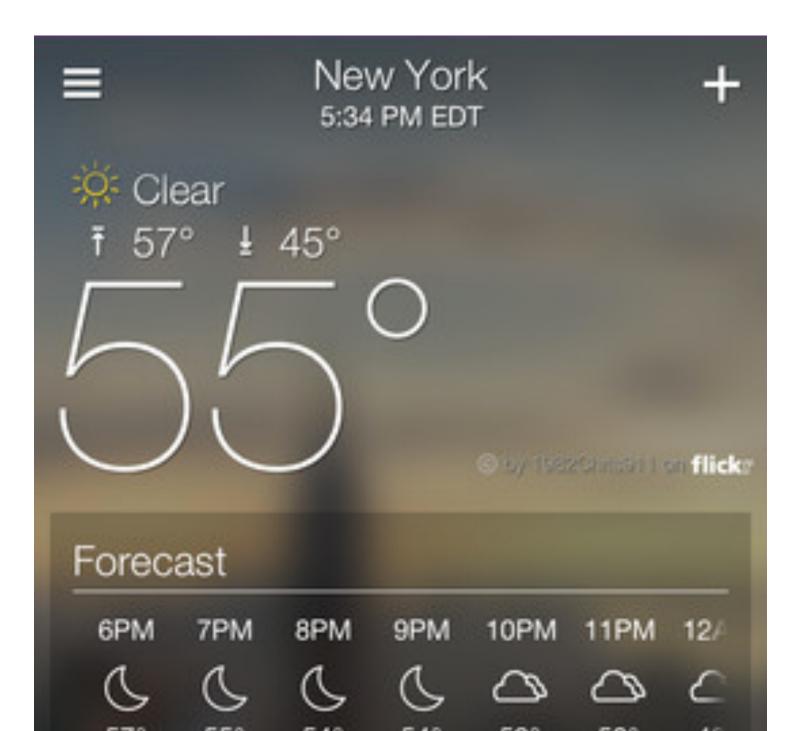


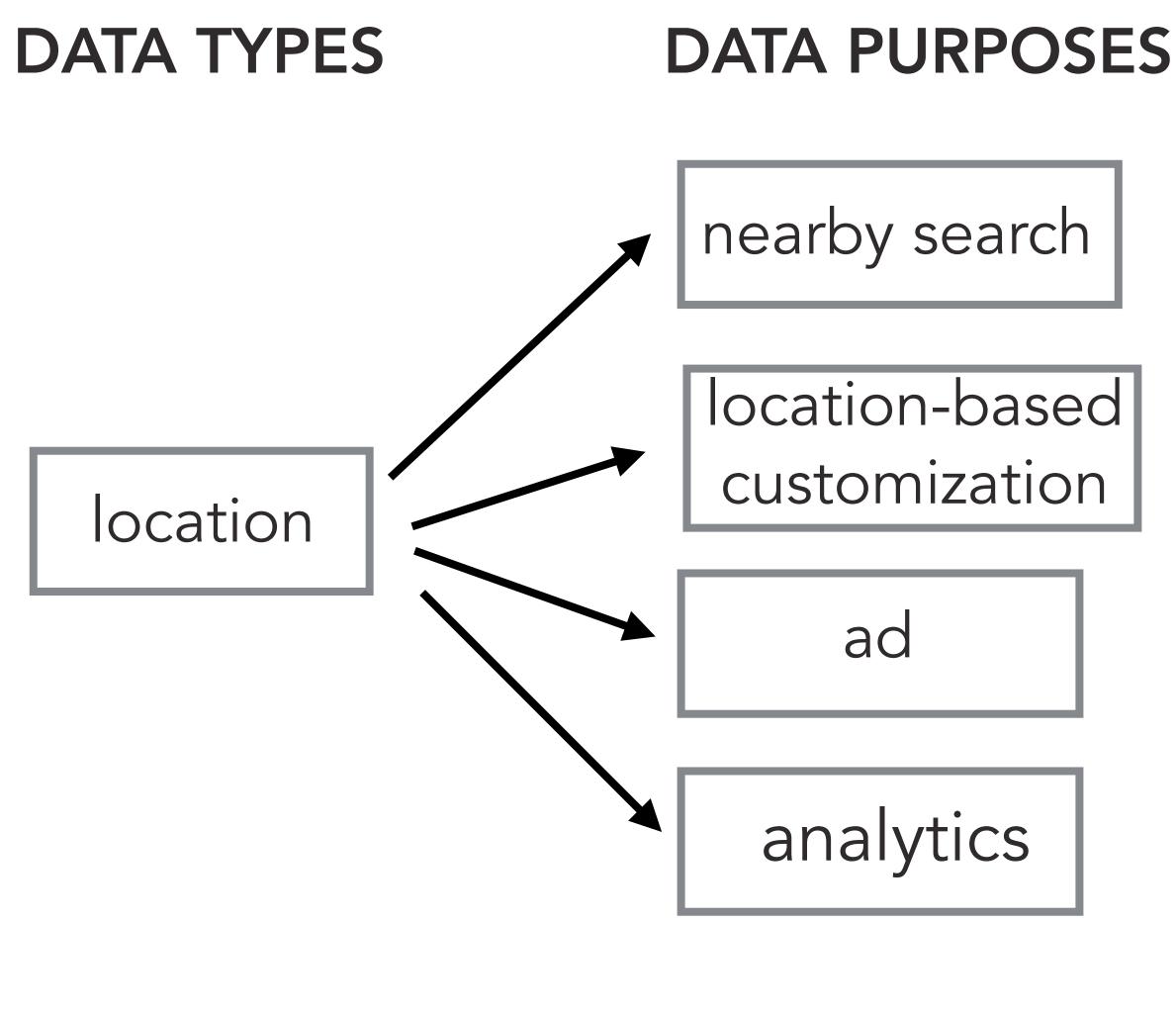


EXAMPLES









.

EXAMPLES







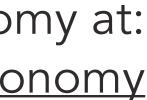


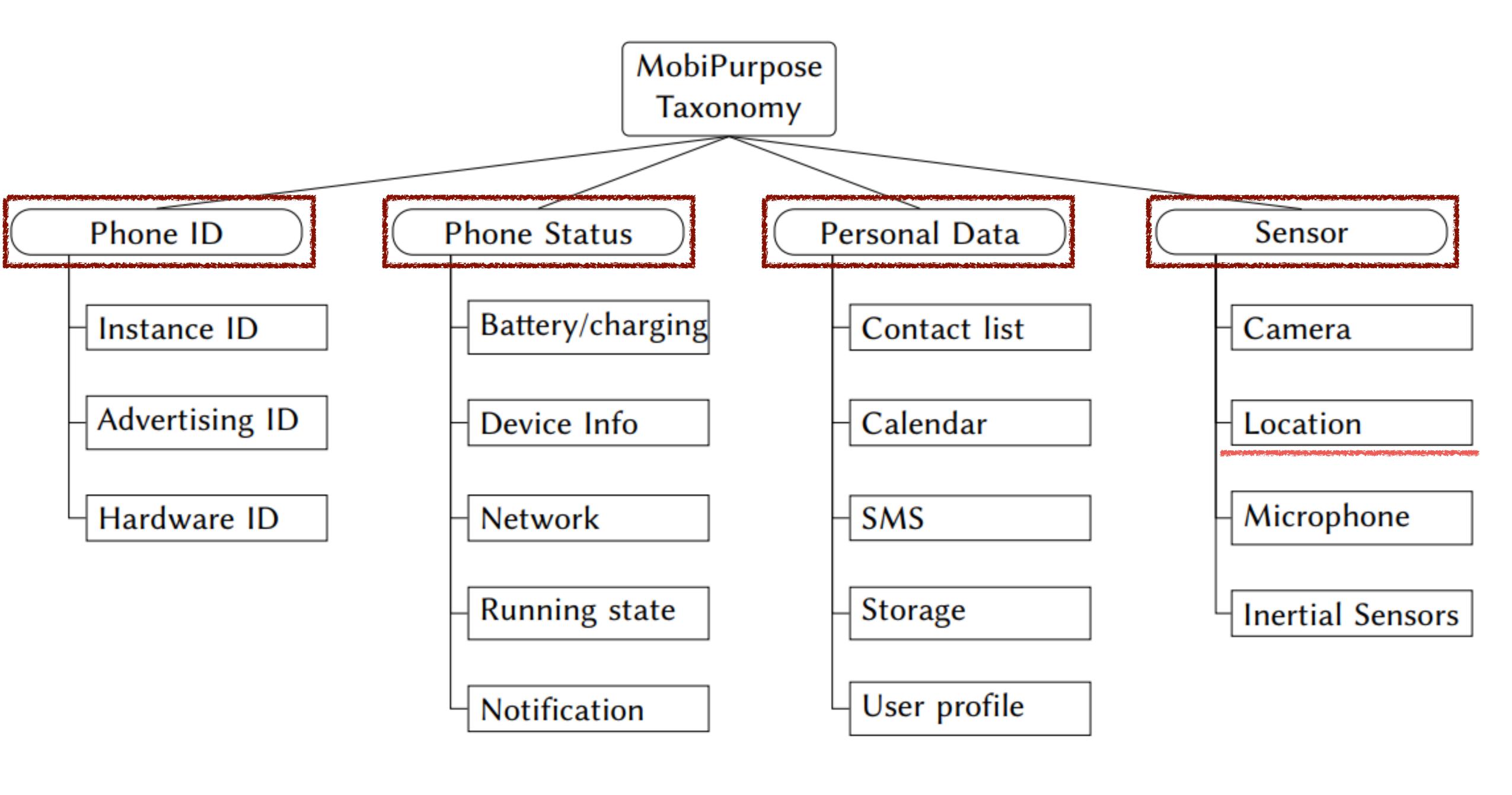
.

13 common data collection purposes for location data

Location ⁷	Nearby Search	Search nearby POIs/real estates			
Location	Location-based	Fetch local weather/radio information			
	Customization				
	Query Transportation	Estimate the trip time through Uber API			
	Information				
	Recording	Track the running velocity			
	Map and Navigation	Find the user location in Map apps			
	Geosocial Networking	Find nearby users in the social network			
	Geotagging	Tag photos with locations			
	Location Spoofing	Set up fake GPS locations			
	Alert and Remind	Remind location-based tasks			
	Location-based game	Play games require users' physical location			
	Reverse geocoding	Use the GPS coords to find the real world address.			
	Data collection for analytics	Collect data for marketing analysis			
	Data collection for ad	Collect data for ad personalization			

See the complete taxonomy at: http://bit.ly/mobitaxonomy

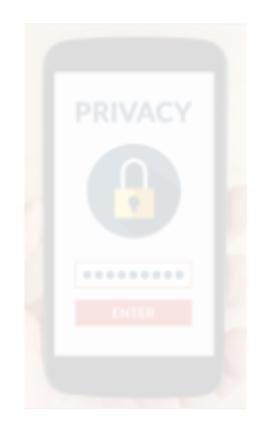




data types



Three empirical studies





Large-scale mobile network tracing^[1]

Smart home ^[2] applications



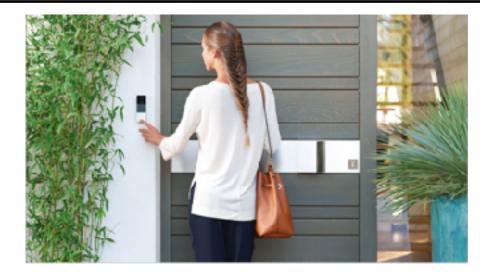
[2] Peekaboo, Jin et al., S&P'22



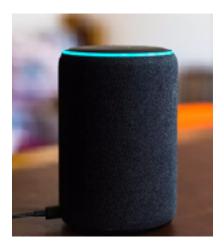
77% Smart home apps do not need raw data.

Sensor

Hello visitor

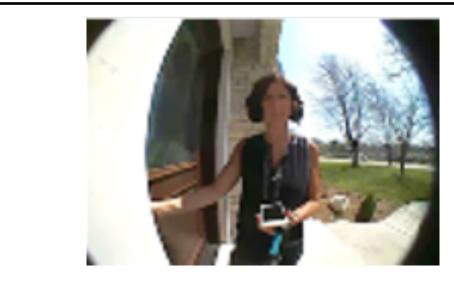


Noise level



Raw Needed data

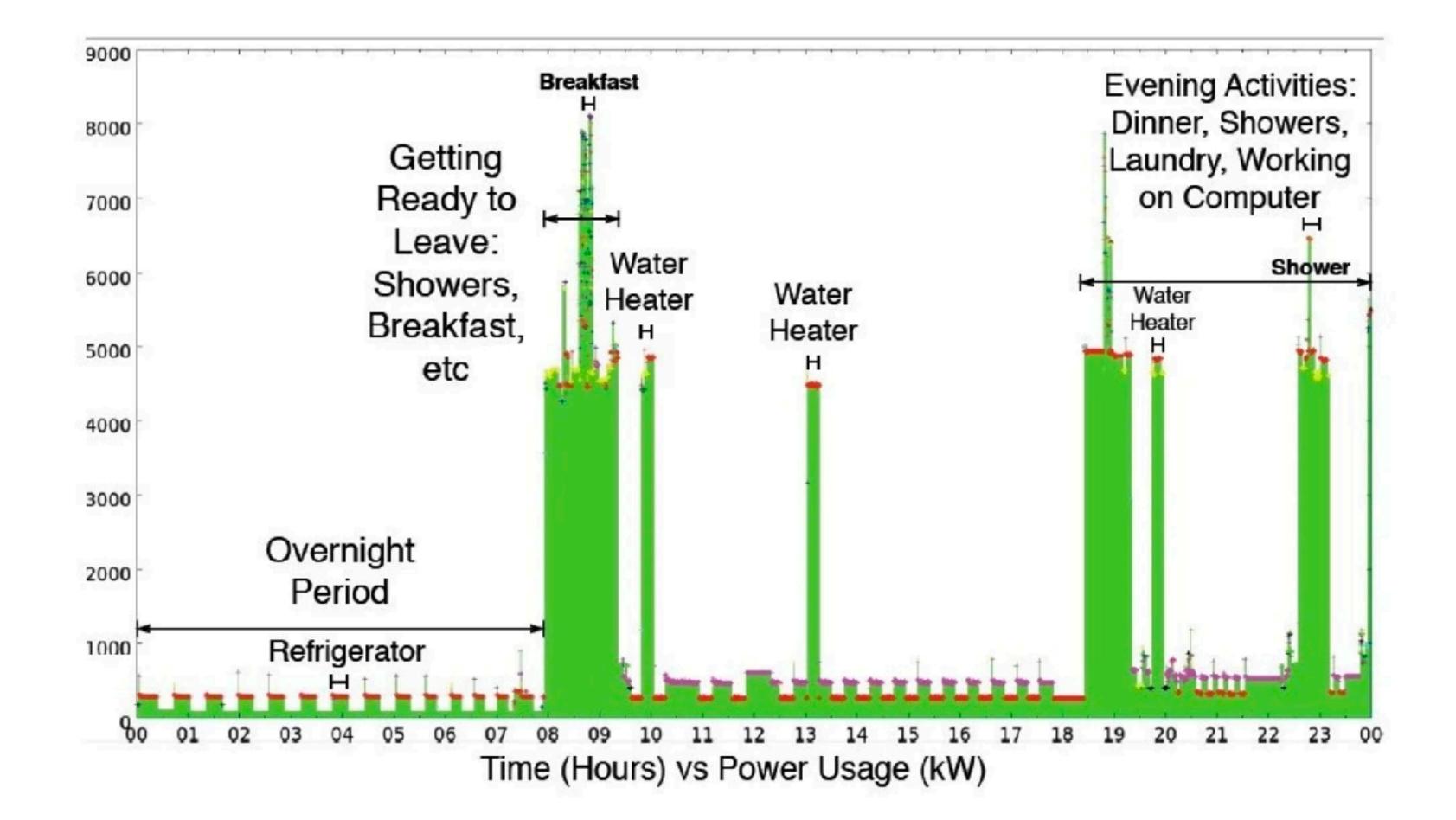






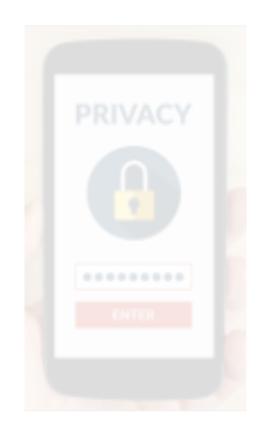
55 db

Oversensing





Three empirical studies





Large-scale mobile network tracing^[1]

Smart home ^[2] applications



[2] Peekaboo, Jin et al., S&P'22



76/80 Smart city apps only need aggregated data.

Aggregation types	Applications
Count/Sum/Average	Electricity usage (su sumption (sum/average emission tracking (su outdoor water usage sum), electric vehicle (sum), rainwater collection (sum), rainwater collection (sum), rainwater collection (count), pedestrian for (average), number of (GPS), fire department tracking (sum/average food consumption (se occupancy (count/average)
Rank/Median/Top X	Snow plowing fairnes
Histogram/Heatmap	Network speed monit erage), food consump noisiest neighborhood
Route	Delivery route with b polluted biking path
TF-IDF	Determine restaurant
Trilateration	Gunshot localization,
Time-series forecasting	Parking occupancy pr
Non-aggregatable apps	Searching for full tra visitors Counting thre

um/average), water consumption (sum/average), heat conage), water leaks (sum/average), fire alerts (average), CO2 um), power-efficient appliances (sum/percentage), indoor vs. e (group sum), building vs. home energy efficiency (group e usage (sum), waste management (sum), renewable energy lection (sum), public transport tracking (count), intersection ents physical activity (average), pedestrian entrance tracking foot traffic (count), user mobility (count), time spent asleep f users that pass by a billboard (count), DMV visitor counting ent delay (average), traffic light wait time (average), irrigation ge), electromagnetic field level monitoring (sum/average), sum/average), indoor home temperature (average), parking erage/sum)

ess, public space noise level monitoring

itoring (group average), water quality monitoring (group avption (group average), street light brightness (group average), ds (group average)

best parking availability, Smart garbage pickup route, least-

type for region, trending TV search queries

air pollution source localization

prediction

ash cart, On-demand rides with autonomous vehicles, DMV rough a camera, Garbage collection prediction

New understanding about privacy

State of the art

MPF

25% users were surprised this app sent their approximate location to dictionary.com for searching nearby words.

Sensor

Hello visitor



Enumerable data collection purposes.

arbitrary text manually annotated hard to validate/assess

Needed data



Raw

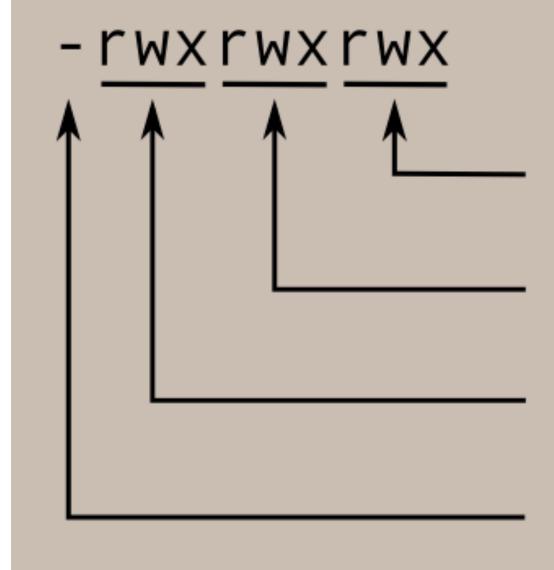


Given a purpose, developers do not need raw data.



File permissions are insufficient.

😣 🖻 🗉 aditya314@ubuntu: ~							
aditya314@u	ıbuntu:~\$ ls						
Desktop	examples.deskt	op Music	Public	Videos			
Documents		new one	Templates	2			
Downloads		Pictures	Untitled Do	cument			
	ıbuntu:~\$ ls -l						
total 52							
	2 aditya314 ad			-			
	2 aditya314 adi						
	2 aditya314 ad			Downloads			
	1 aditya314 adi	-	Mar 5 01:05	examples.desktop			
- FW- FW- F	1 aditya314 adi	itya314 0	Mar 5 03:53	ggf.txt			
	1 aditya314 adi		•				
drwxr-xr-x	2 aditya314 adi	itya314 4096	Mar 5 01:21	Music			
drwxrwxr-x	2 aditya314 adi	itya314 4096	Mar 5 03:53	new one			
drwxr-xr-x	2 aditya314 adi	itya314 4096	Mar 5 01:21	Pictures			
drwxr-xr-x	2 aditya314 adi	itya314 4096	Mar 5 01:21	Public			
drwxr-xr-x	2 aditya314 adi	itya314 4096	Mar 5 01:21	Templates			
- FW- FW- F	1 aditya314 adi	itya314 0	Apr 27 02:55	Untitled Document			
drwxr-xr-x	2 aditya314 ad	itya314 4096	Mar 5 01:21	Videos			
- FW- FW- F	1 aditya314 ad	itya314 268	Mar 5 04:17	xyz.txt			
aditya314@u	ibuntu:~\$						

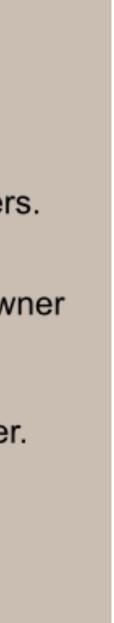


Read, write, and execute permissions for all other users.

Read, write, and execute permissions for the group owner of the file.

Read, write, and execute permissions for the file owner.

File type: - indicates regular file d indicates directory



State of the Art: fine-grained permission manifest

```
<manifest ... >
```

```
<!-- Always include this permission -->
```

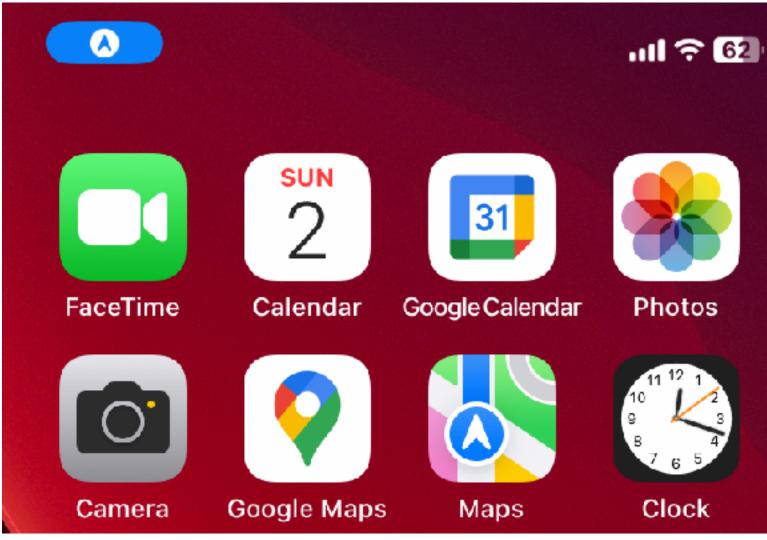
<uses-permission android:name="android.permission.ACCESS_COARSE_LOCATION" />

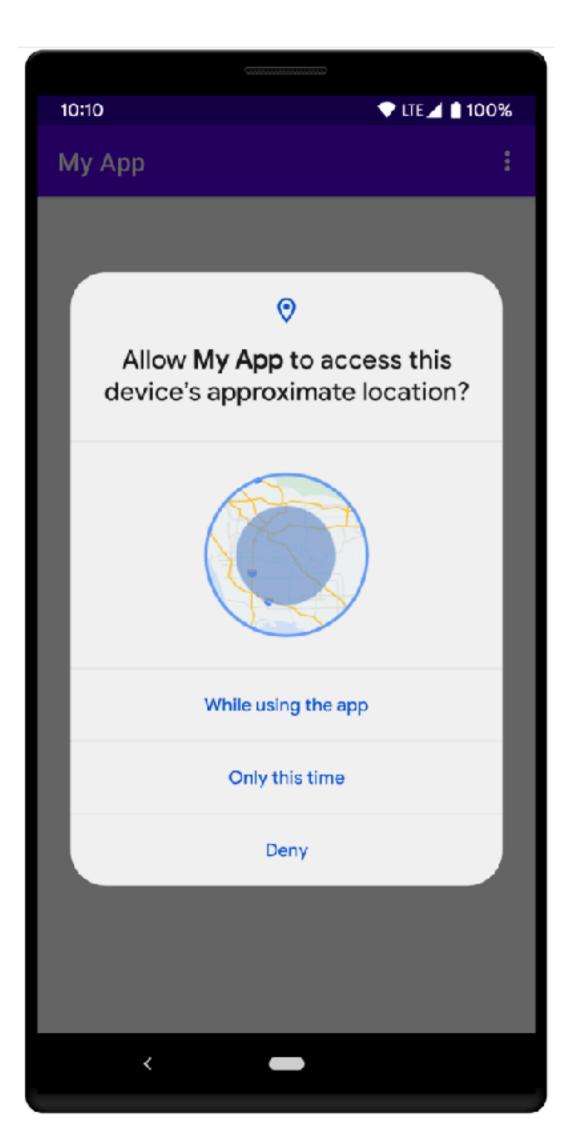
<!-- Include only if your app benefits from precise location access. --> <uses-permission android:name="android.permission.ACCESS_FINE_LOCATION" /> </manifest>

```
<manifest ... >
```

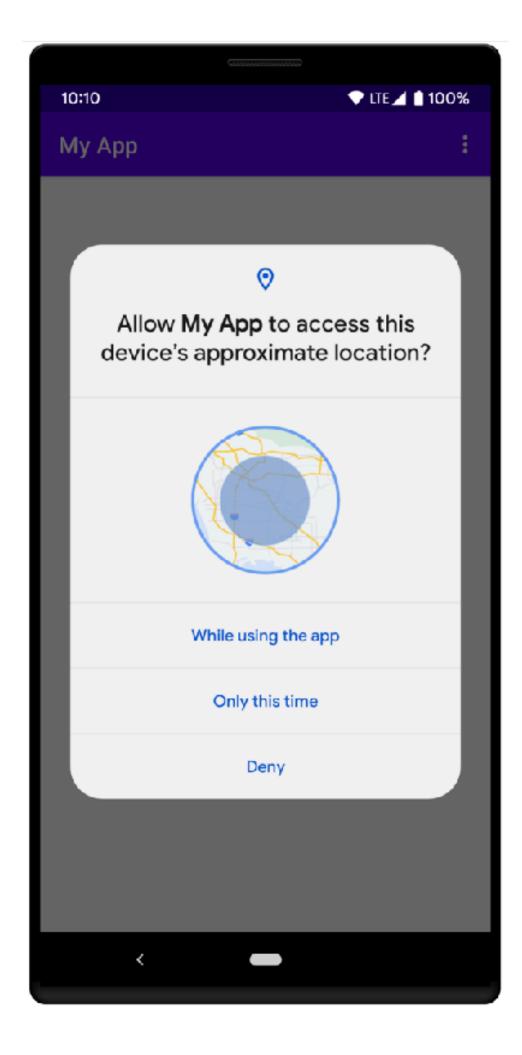
<!-- Required only when requesting background location access on Android 10 (API level 29) and higher. -->

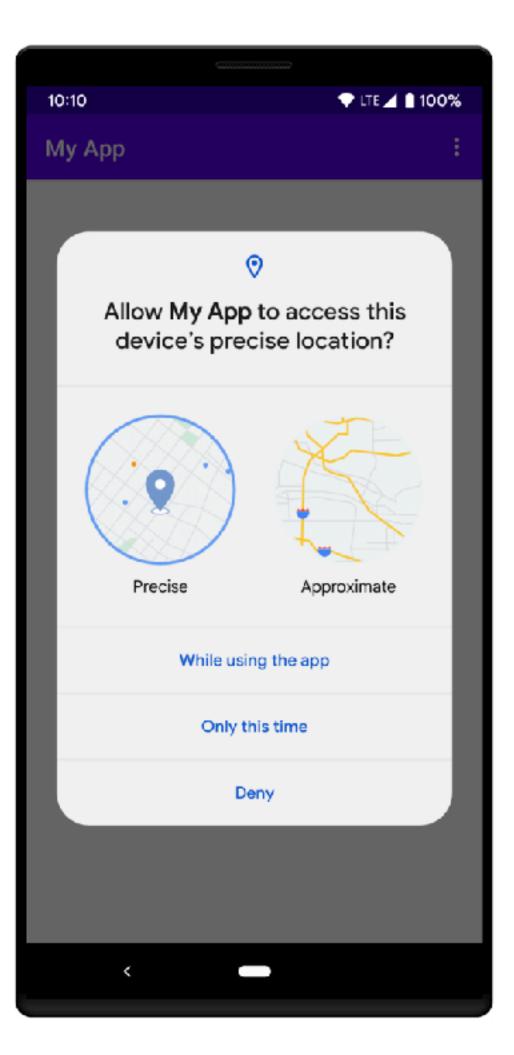
<uses-permission android:name="android.permission.ACCESS_BACKGROUND_LOCATION" /> </manifest>





State of the Art: User choices







	Precise	Approximate
While using the app	ACCESS_FINE_LOCATION and ACCESS_COARSE_LOCATION	ACCESS_COARSE_LOCATION
Only this time	ACCESS_FINE_LOCATION and ACCESS_COARSE_LOCATION	ACCESS_COARSE_LOCATION
Deny	No location permissions	No location permissions



The permission granularity dilemma

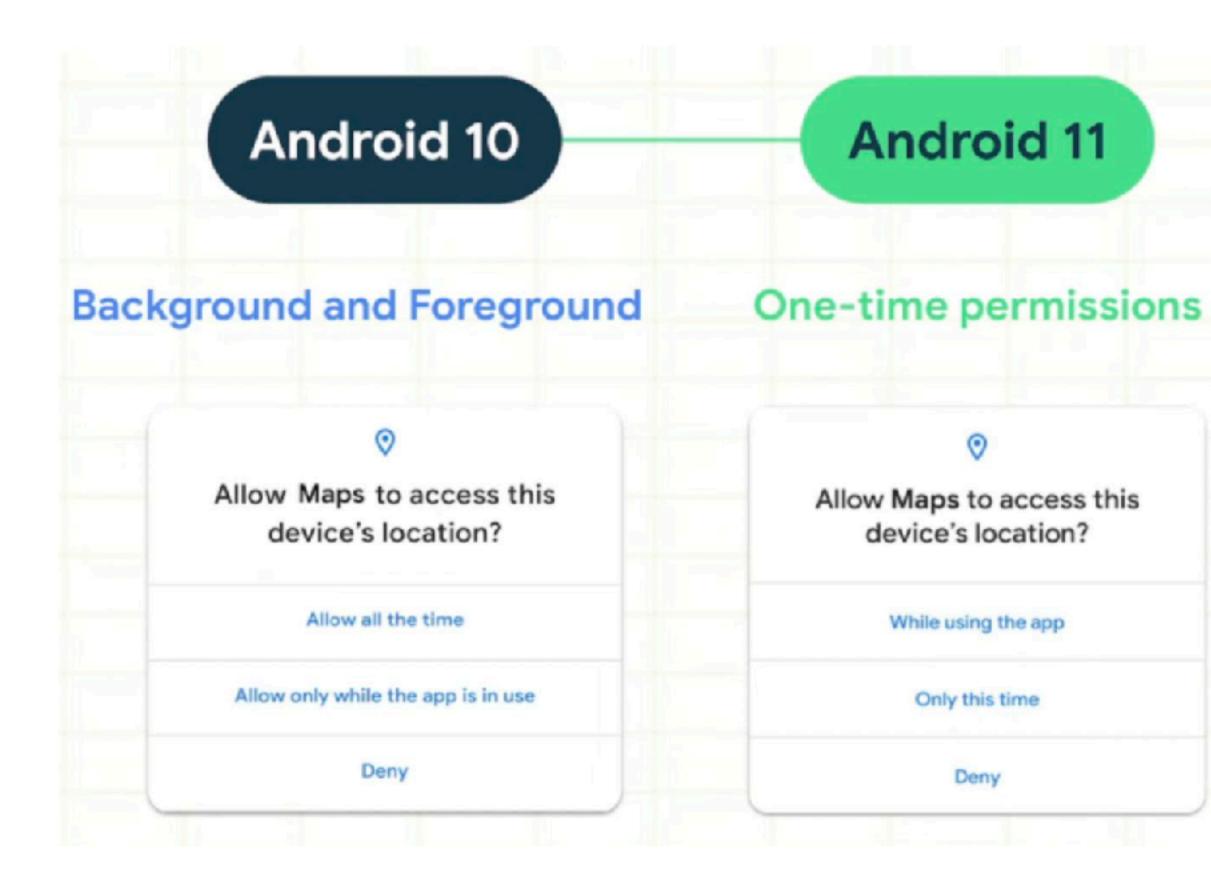
More fine-grained permissions

- → Better privacy
- → More management burden for users Harder learning curve for app developers More implementation efforts for system builders

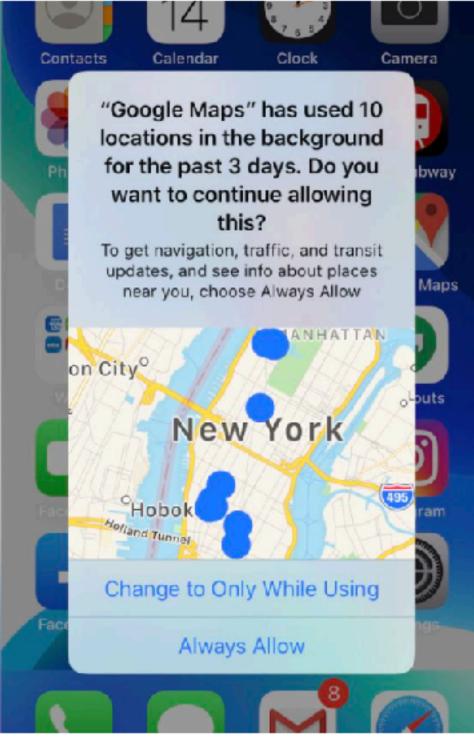
More coarse-grained permissions

- \rightarrow Worse privacy
- → Oversensing risks More users deny data requests More complaints for system builders Hard to gain trust from users for app developers

On-going permission dilemma



iOS 12



iOS 13
.ıll 1:52 PM ֎ ◀ 60%
C
Discover
New
Allow "Apple Store" to access your location? This app will use your location information to identify what's happening at nearby Apple Retail stores and to notify you about and access services available in store. Your location is used to identify, notify you about, and access services available in nearby Apple Retail Stores while you are using the app.
Allow Once
Don't Allow
New

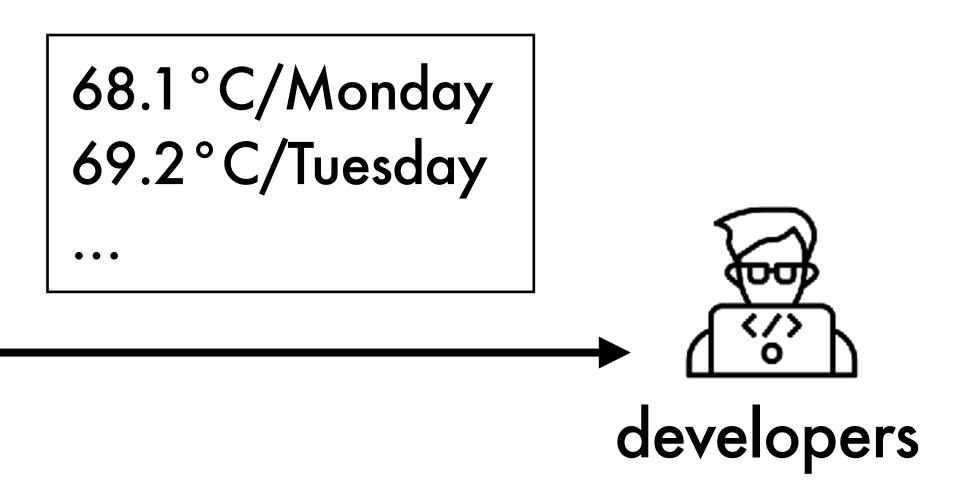


Talk outline

- 1. Modular Privacy Flows (MPF) in a Nutshell 2. Why MPF
- 3. How MPF
- 4. When and when not MPF
- 5. Future Work

How can Nest prove that they only collect aggregated data?



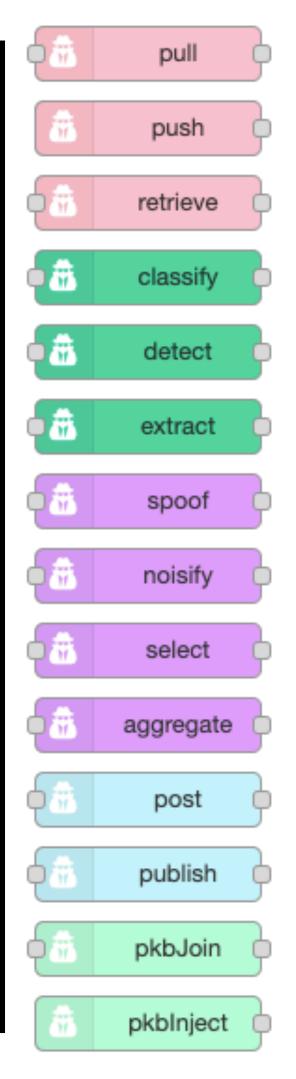


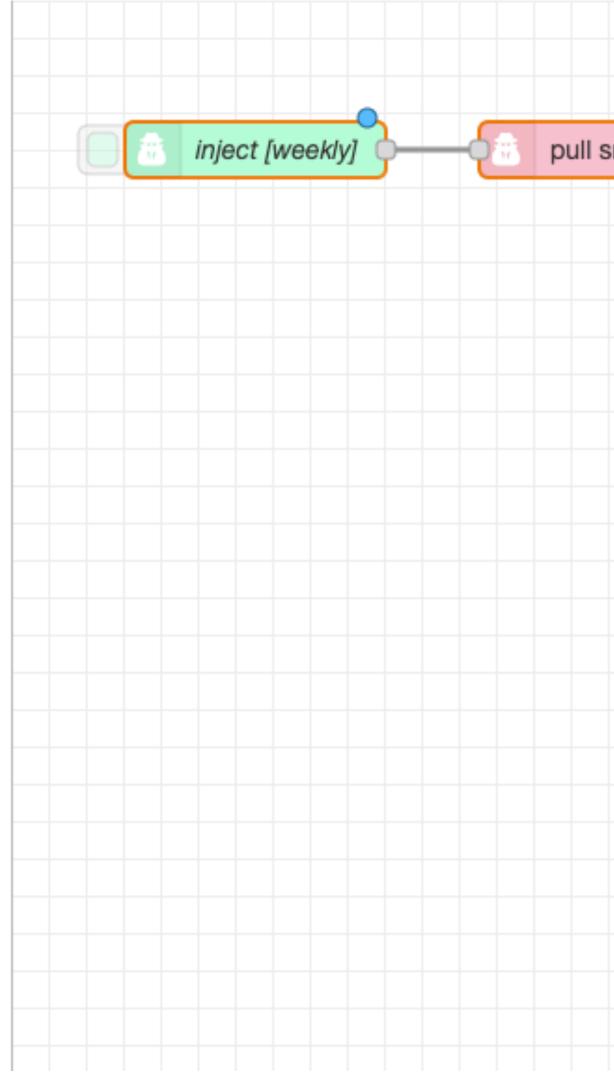
Open source?

[1] Jin et al., Peekaboo: A Hub-Based Approach to Enable Transparency in Data Processing within Smart Homes. S&P'22



Program pre-processing functions using chainable operators





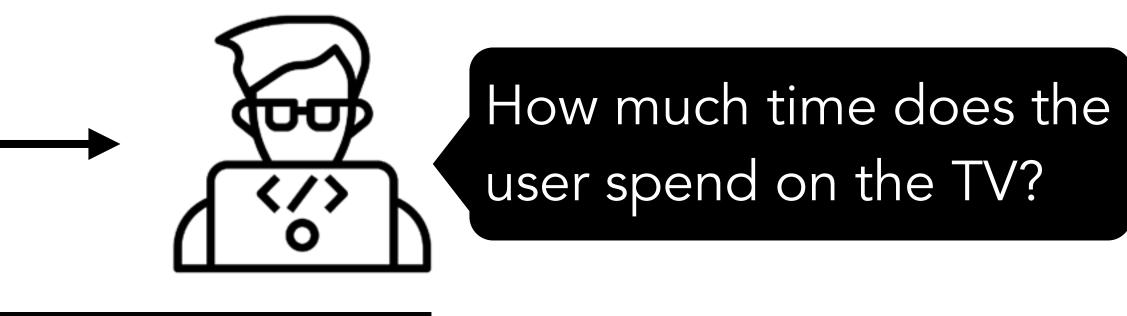
	sum duration]
Edit aggregate	node
Delete	Cancel Done
Properties	
4 1 10001100	
Name	aggregate [sum duration]
ද්ථ Data Type	tabular
⊙ Target	custom
Olarget	
🗋 Tabular field	duration
C) Operation	sum 🗸
C) Operation	sum

2. Implement - Peekaboo A text-based whitelist manifest (i.e., program representation)



@purpose: To measure dev WeeklyUsageHours{ // operator [properties] inject [weekly] -> pull [smart TV driver] ->

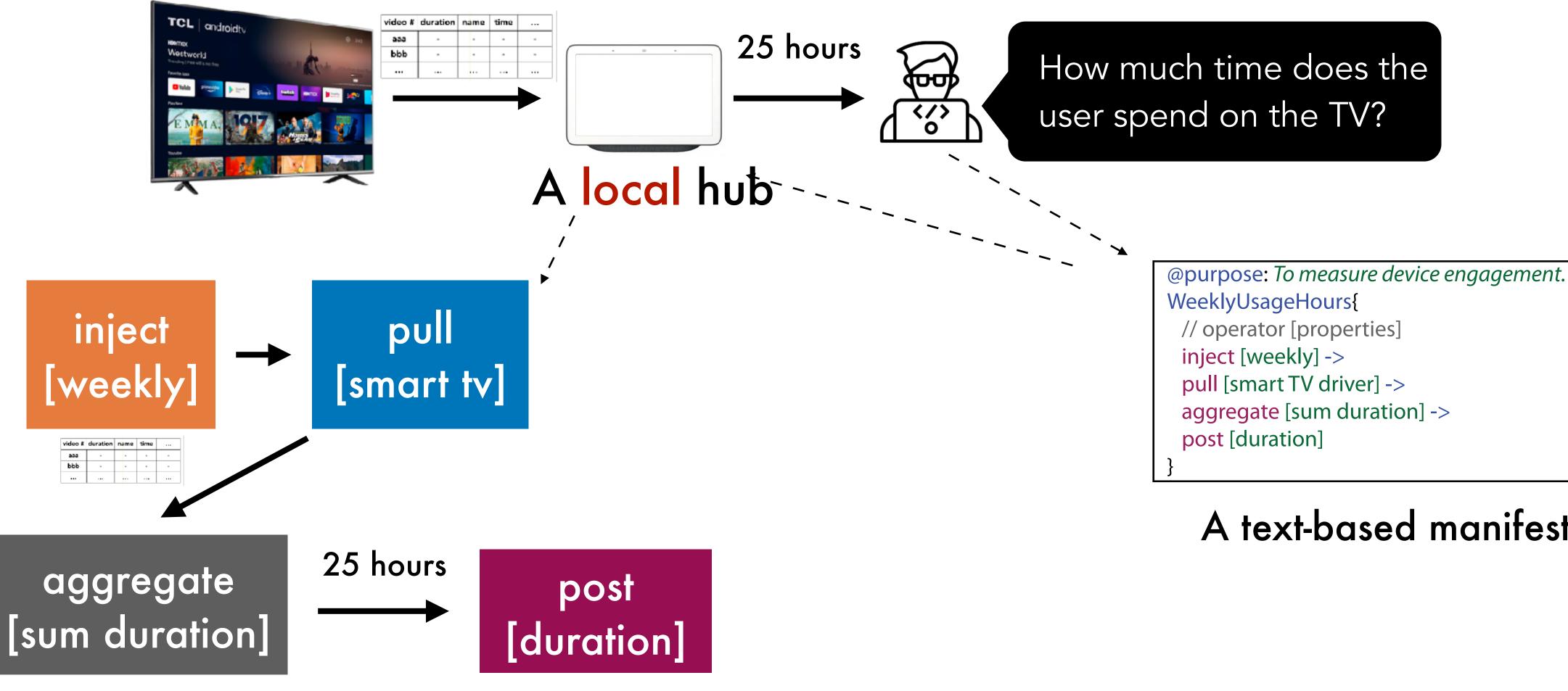
aggregate [sum duratio **post** [duration]



vice engagement.	
n] ->	



2. Implement - Peekaboo A trusted runtime with pre-loaded implementations



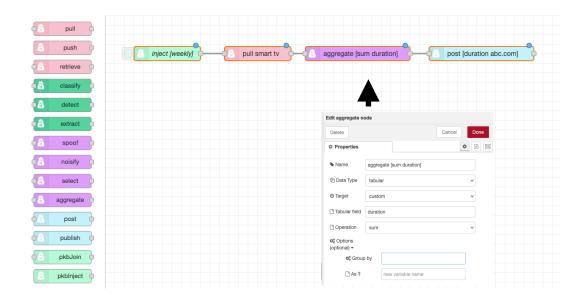
A text-based manifest

A trusted runtime with pre-loaded, open-source implementations





Smart home app store



Programming environment with operators



Runtime with preloaded implementations

App developers

@purpose: To measure device engagement.
WeeklyUsageHours{
 // operator [properties]
 inject [weekly] ->
 pull [smart TV driver] ->
 aggregate [sum duration] ->
 post [duration]

Manifest

Smart home hub → privacy firewall

Smart home app

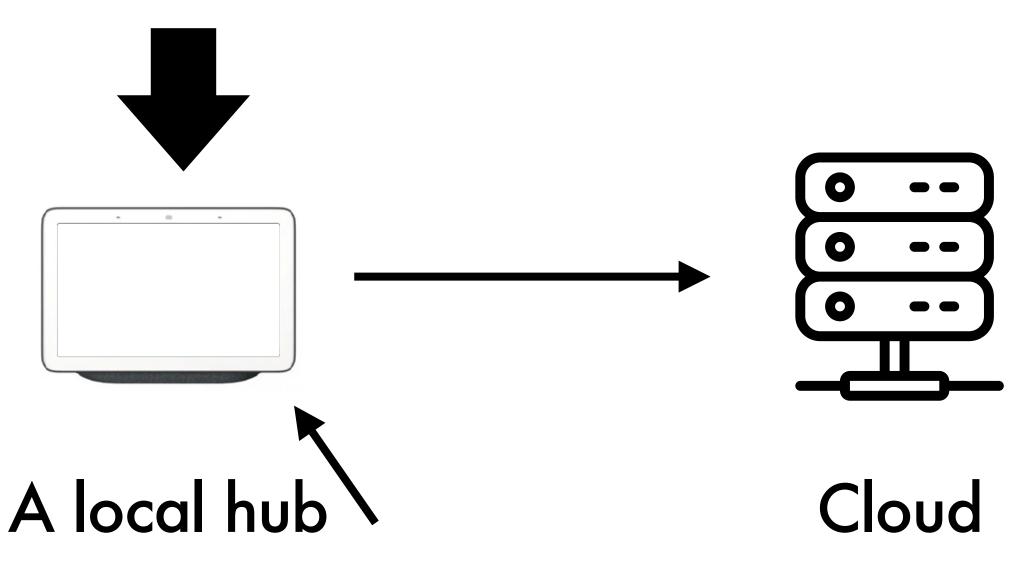




Edge devices

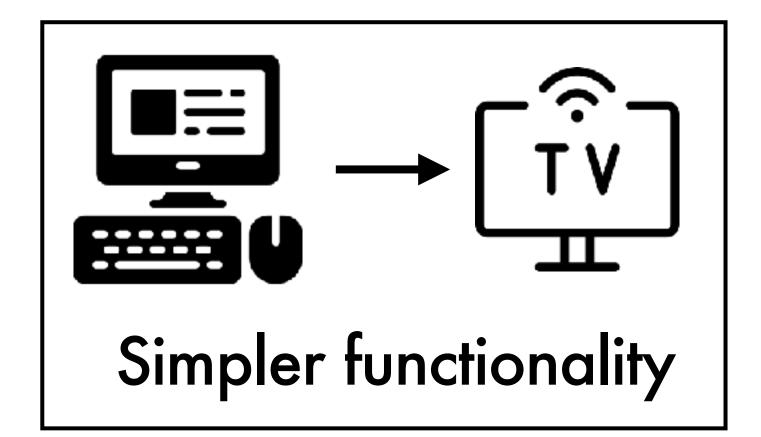
53

@purpose: To measure device engagement. WeeklyUsageHours{ // operator [properties] inject [weekly] -> pull [smart TV driver] -> aggregate [sum duration] -> post [duration]



"Privacy firewall"

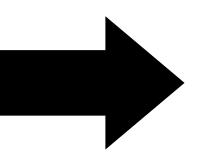
Peekaboo v.s. Firewall



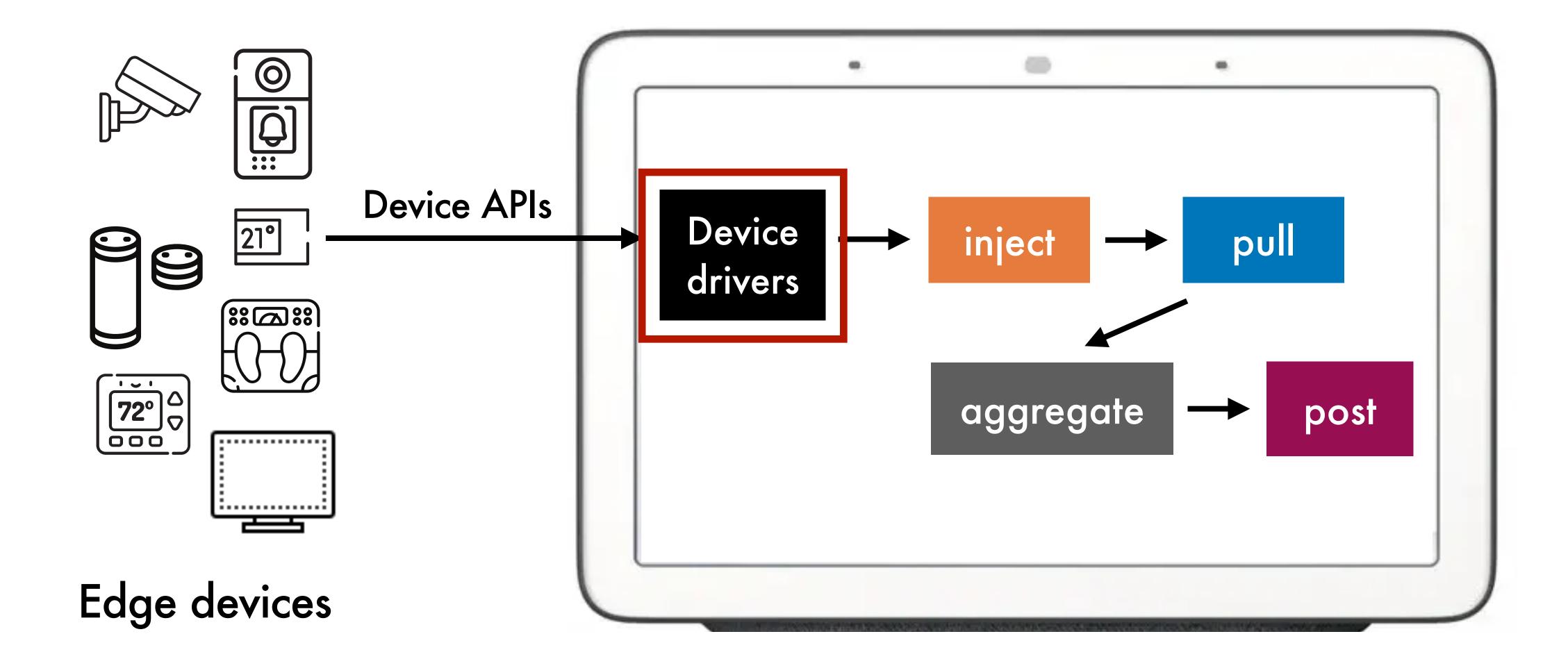
77% Apps do not need raw data.

Whitelist-only Developer-in-the-loop

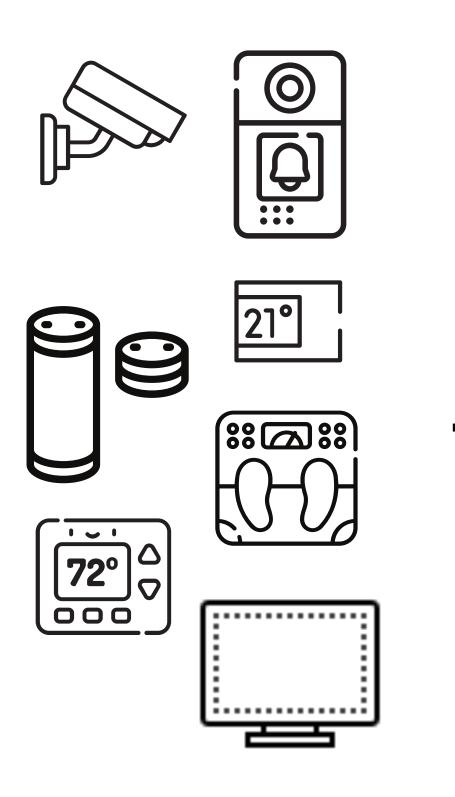
Pre-process users' data



How Peekaboo works Handle heterogeneous hardware with device drivers



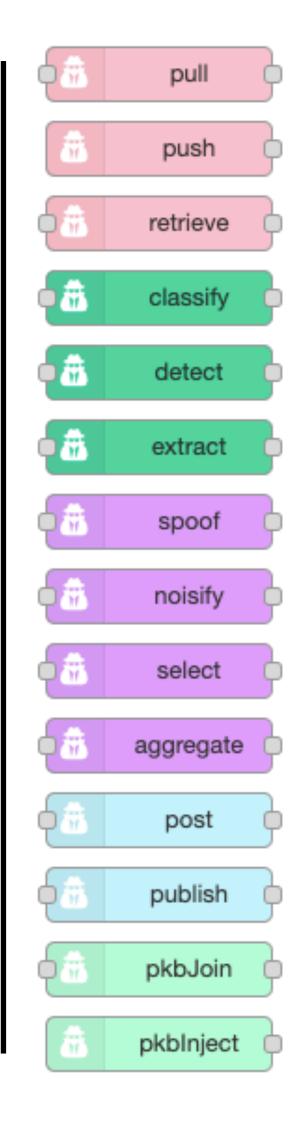
How Peekaboo works A fixed set of operators



Edge devices

video, image, audio, tabular, scalar

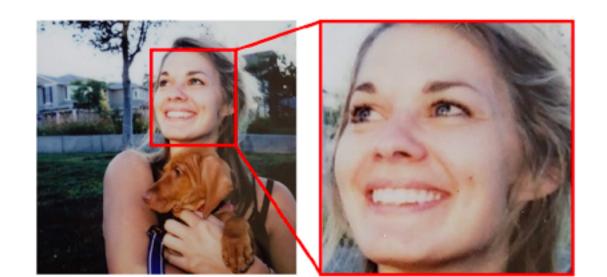
A fixed set of operators

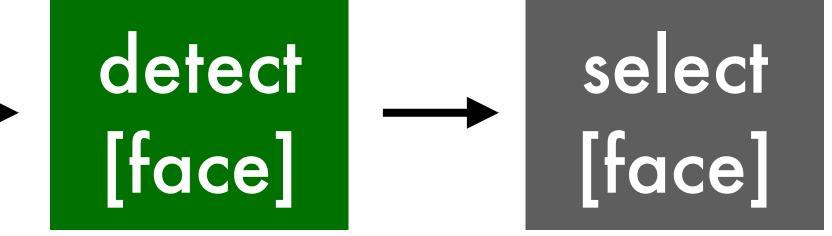


How Peekaboo works An operator = A verb keyword



	product_id	product_name	inventory_received	starting_inventory	inventory_on_hand	minimum_required
1	2	Booth	29pcs	27pcs	56pcs	20pcs
2	3	Maclean	23pkts	25pkts	48pkts	25pkts
3	4	Closeup	24pkts	25pkts	49pkts	25pkts

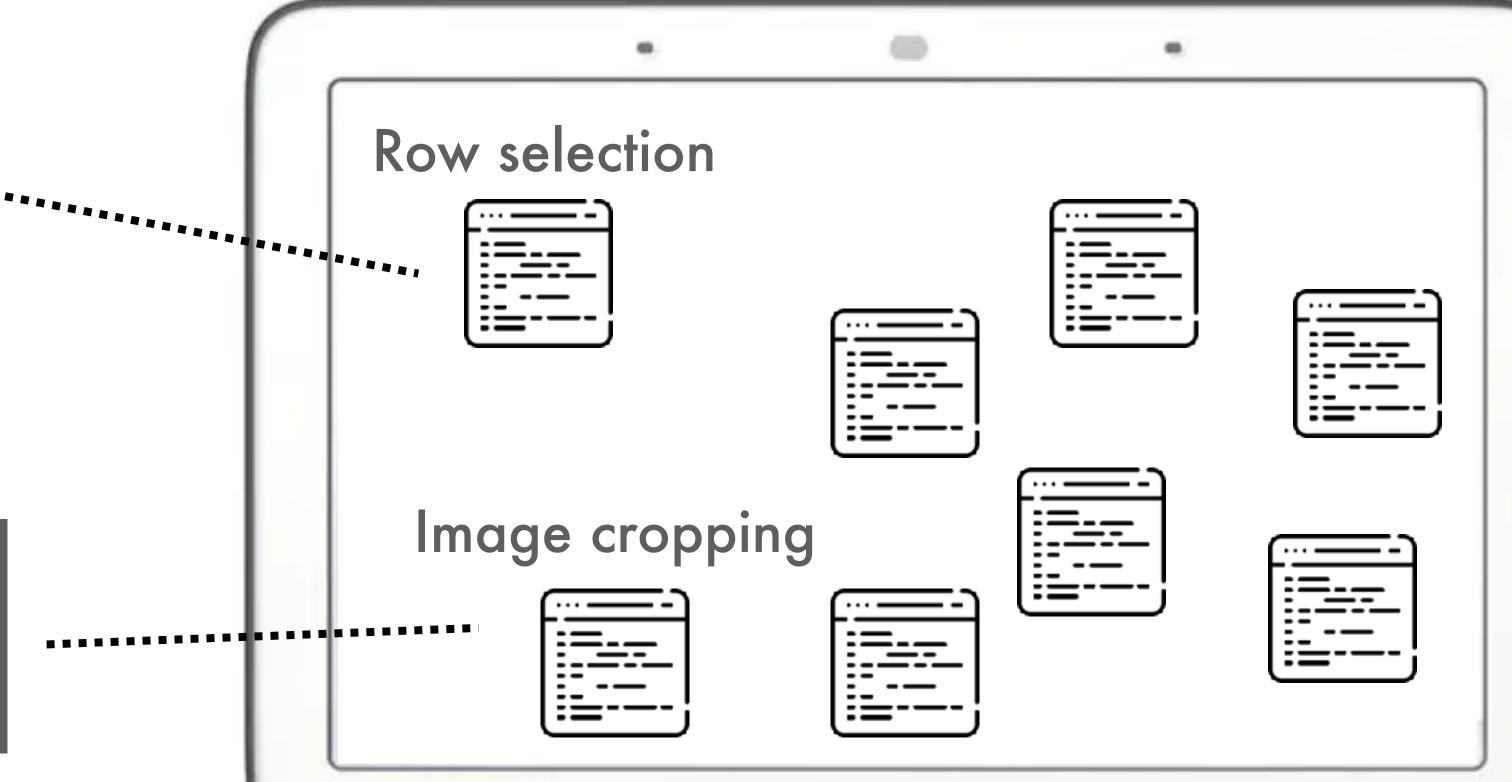




How Peekaboo works Operators are mapped to pre-loaded implementations



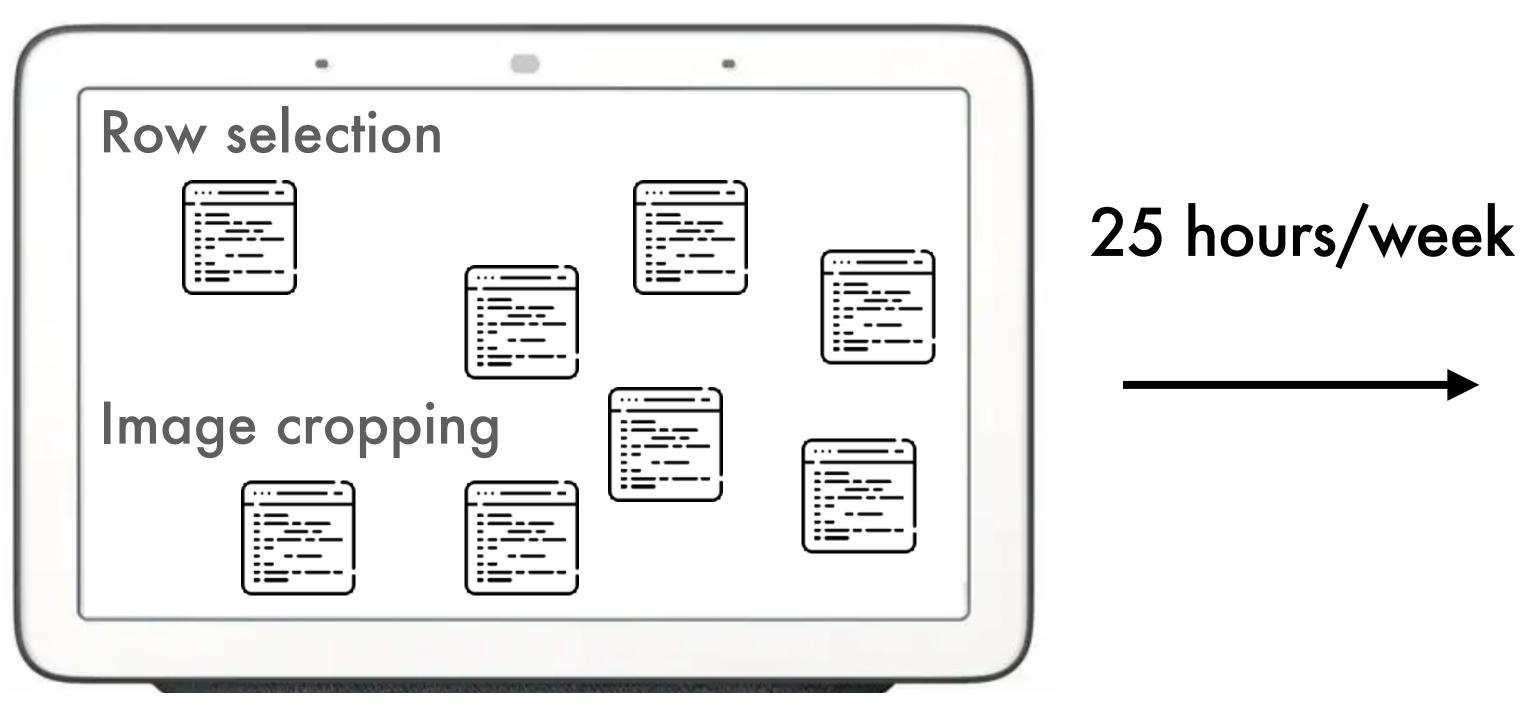
select [face]





How Peekaboo works A small set of pre-processing algorithms improve privacy

video #	duration	name	time	•••
aaa	-	-	-	-
bbb	-	-	-	-
•••		• • •	•••	•••



Implementation (hardware)



Edge devices



60

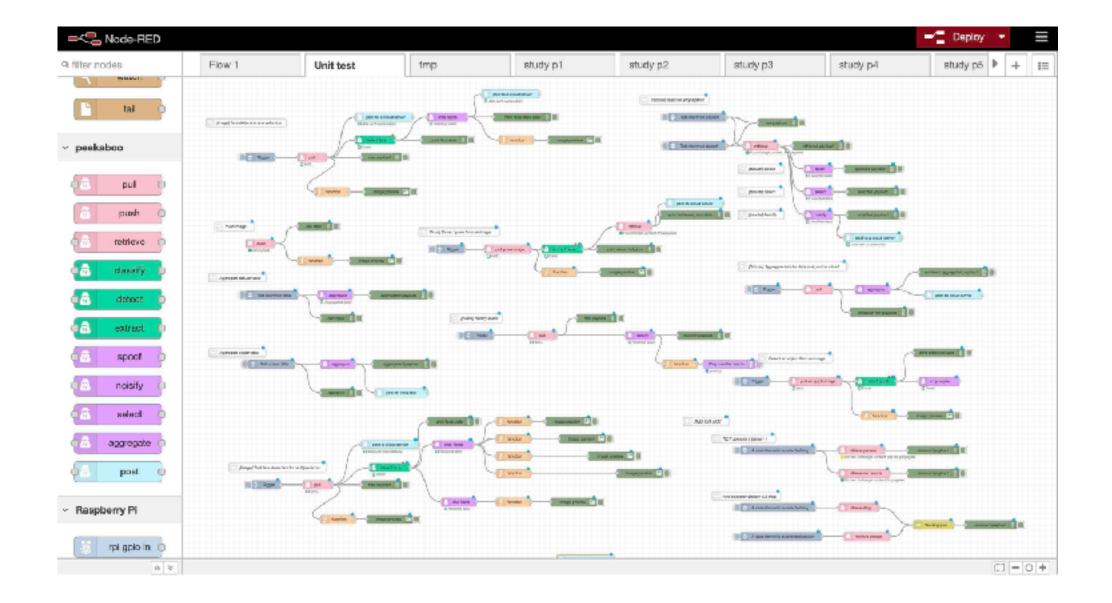
Raspberry PI + TPU





Implementation (software)

- 1. Operators: Node.JS package
- 2. Programming IDE: NodeRed
- 3. Drivers: 5 data types
- 4. 23 Preloaded implementations



Expressiveness (200+ smart home cases)

→ c c ▲	lot Secure 128.237.99	9.157:1880/#flow/82c	51466.bda598				x G	• • 🖷	' 흚 ۱	• 😳	* 🛞
Node-RED										Deploy	-
ilter nodes	Flow 1	Flow 2	Test playground	Unit test	evaluation	HJ's testbed 1	HJ's testbed 2	Graph simp	lification	Perfor	+
file o						~					
👌 filo in 🖕											
२ watch 🖕											
🗅 tail 🔶											
eekaboo		vater leak detection on	the floor using carneras								
CERADOO											
🗈 puil 🔶											
🚯 push 🍦		under leek detection o	n the floor using micropi	hanne							
retriove		water leak selection o	n the noor dising micropi	liones							
🔹 classify	a wa	it for pushed audio (recognize drip;	ping sound	retrieve dripping even	nt o	king join (if appear 5 time	ес) —	pulla	udio 🕞	
	a discon		-02 recognize drip;	ping sound	- retrieve dripping ever	nt o	king join (if appear 5 time	es) — - (>
detect	C discon	nneoted	n the floor using humidit		- retrieve dripping ever	nt O Dioc	king join (if appear 5 time	es)(post to d	Dud
detect o	C discon	nneoted			• retrieve dripping ever	nt O To bloc	king join (if appear 5 time	ee)(Dud
detect detect extract spoof	C discon	nneoted		ly sensor	aggregate humidity scores		king join (if appear 5 time lassify dripping sound	es)((post to d	
detect iii extract iii spoof iii noisiíy	C discon	nneeted I water leak detection o	n the floor using humidit	ly sensor				es)(post to d	vent p
detect extract spoof noisify select	C discon	nnooted I water leak detection a trigger every 30 mins	n the floor using humidit	ly sensor						post to d	vent p
detect iii extract iii spoof iii noisify	C discon	nneeted I water leak detection o	n the floor using humidit	ly sensor						post to d	vent p
detect iii extract iii spoof iii noisily select	C discon	nnooted I water leak detection a trigger every 30 mins	n the floor using humidit	ly sensor						post to d	
iii detect iii extract iii spoof iii noisify iii select iii aggregate	C discon	nnooted I water leak detection a trigger every 30 mins	n the floor using humidit	hy sensor						post to d	vent p

Data overaccess mitigation breakdown

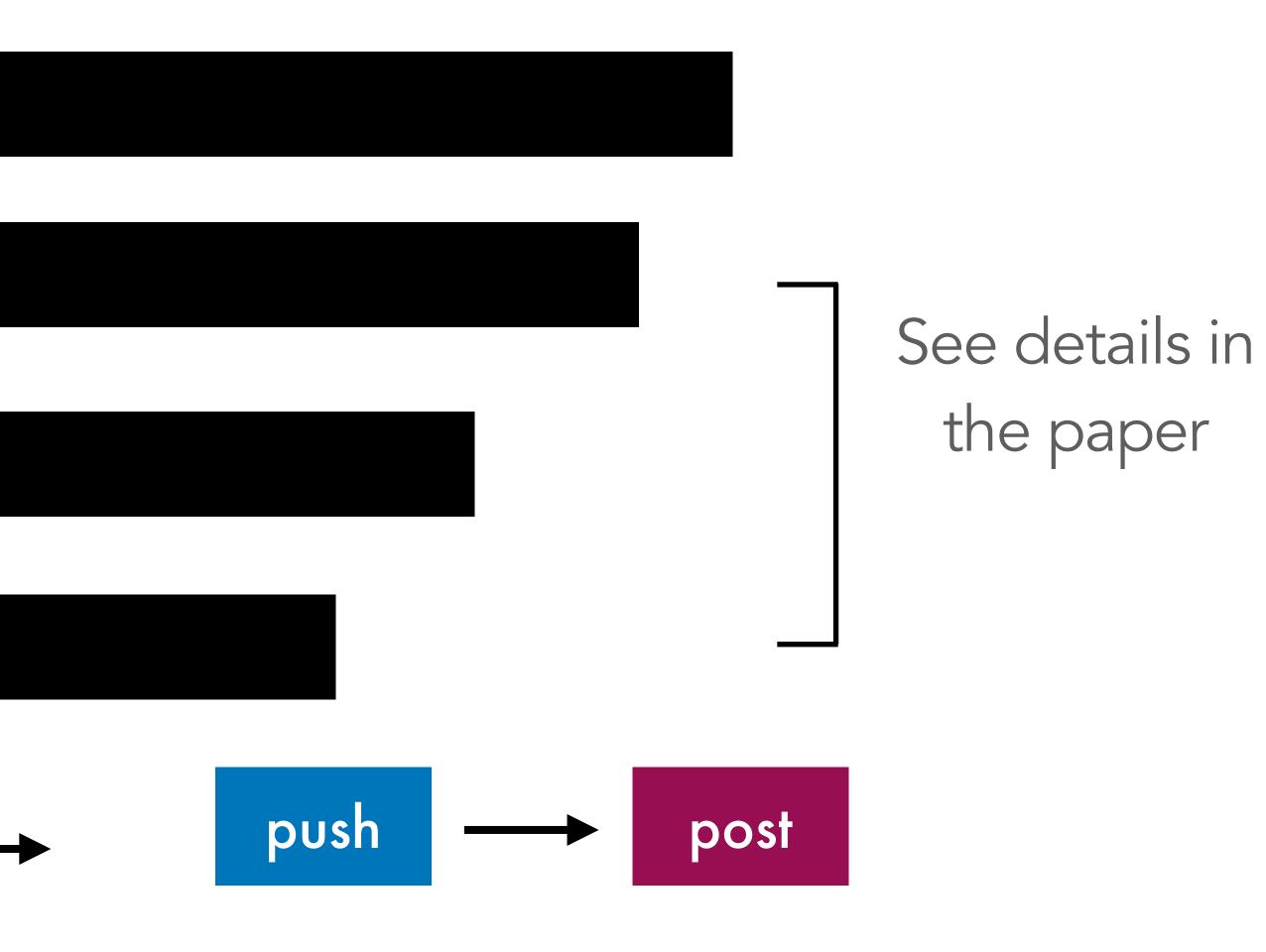
unique manifests: 68

content selection: 64

explicit noisification: 57

conditional filtering: 51





n

System performance



≈\$100

25 inference/s

100 filtering/s

1-80 ms per request

Utility privacy tradeoff example



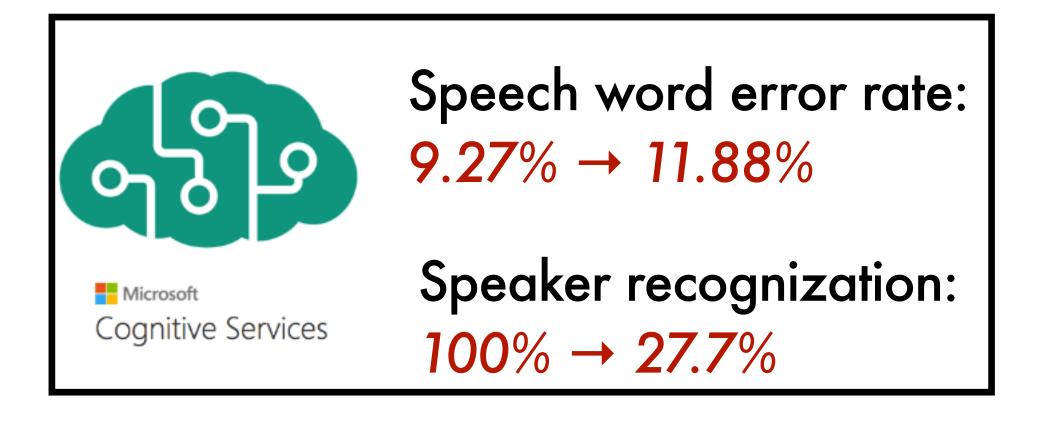
incognito voice assistant

6 speakers 112 audio files [1]





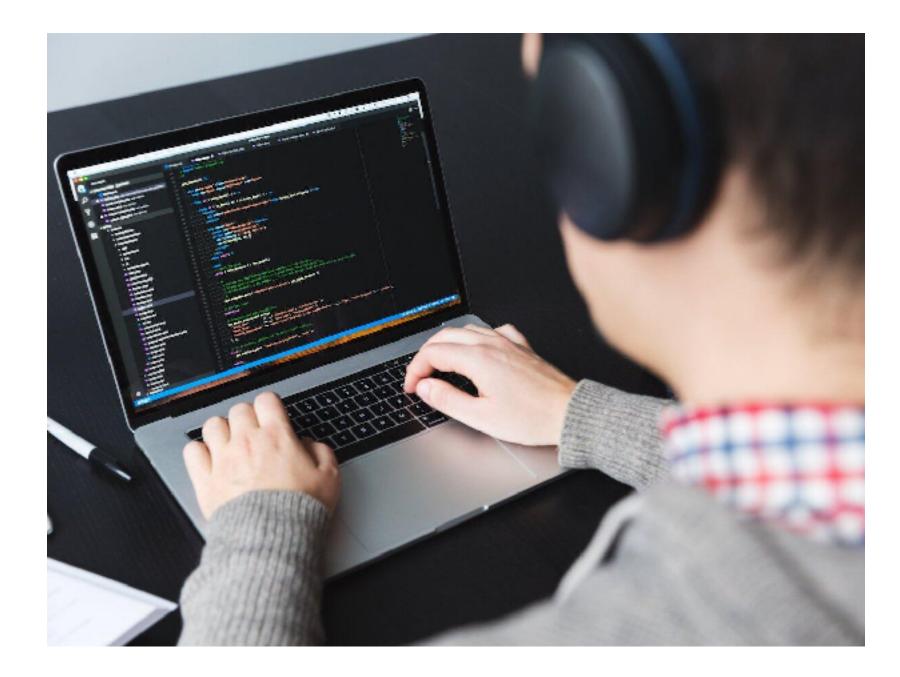
<5% random pitch shift



[1] CMU PDA Speech Database



Developer studies



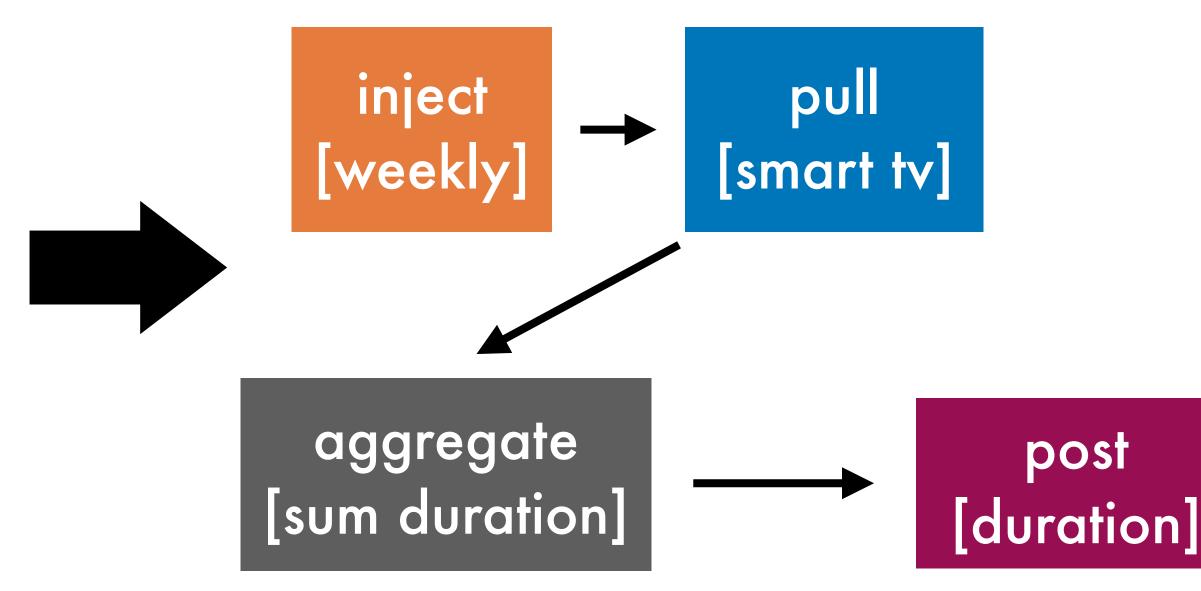
Task descriptions IDE & Unit tests

6 - 15 mins to author a manifest

Advantages Manifests enforce fine-grained data collection

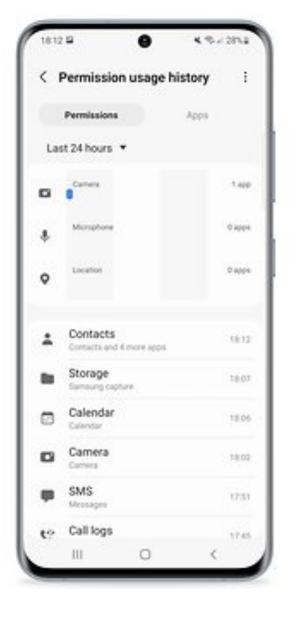
@purpose: To measure device engagement. WeeklyUsageHours{ // operator [properties] inject [weekly] -> pull [smart TV driver] -> aggregate [sum duration] -> post [duration]

public, non-proprietary





Advantages **Repetitive** implementation and **distributed** interfaces



Samsung





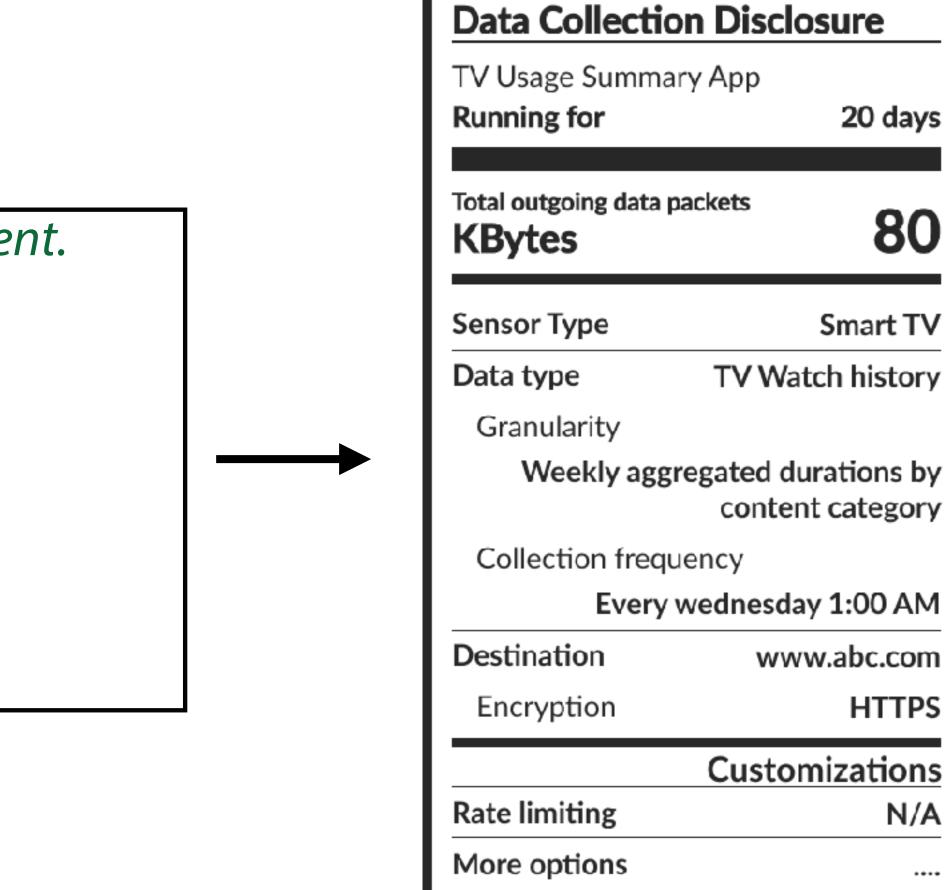
Nest



Users?

Advantages Manifests → enforceable/dynamic privacy nutrition labels

@purpose: To measure device engagement. WeeklyUsageHours{ // operator [properties] inject [weekly] -> pull [smart TV driver] -> aggregate [sum duration] -> post [duration]



[1] Security and Privacy "Nutrition" Label, P. Emami-Naeini et al, IEEE S&P'20



[1]

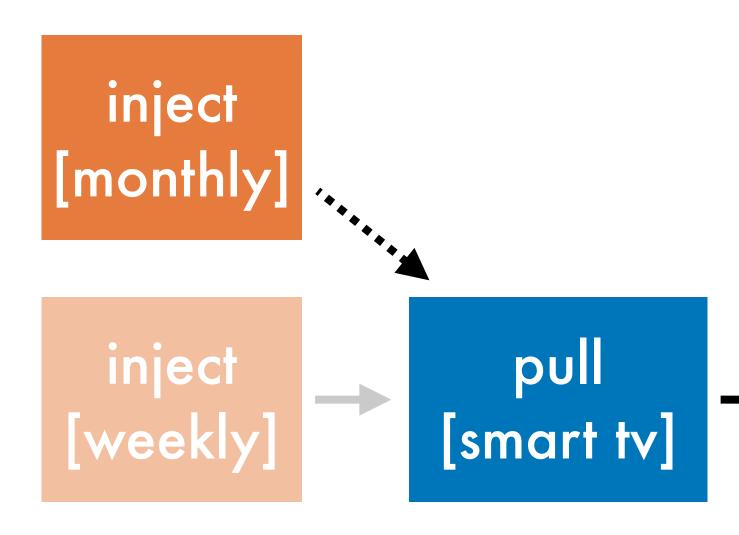
Advantages Built-in fine-grained control through manifest rewriting

Data Collection Disclosure

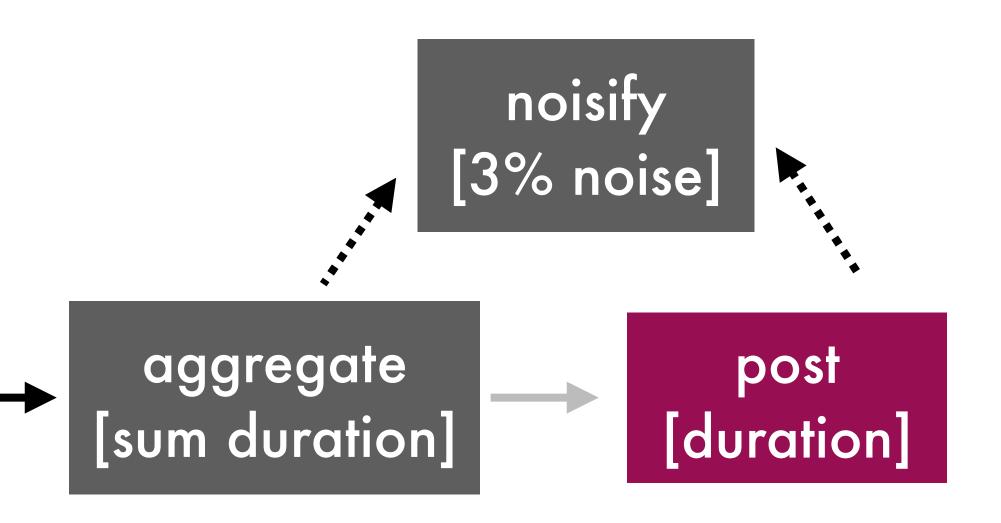
TV Usage Summary App

(Cu	st	on	niz	at	io	ns

Rate limiting	N/A
More options	



Change the rate to monthly



Revisit: The permission granularity dilemma

More fine-grained permissions

- → Better privacy
- → More management burden for users Harder learning curve for app developers More implementation efforts for system builders

More coarse-grained permissions

- → Worse privacy
- → Overaccess risks

More users deny data requests More complaints for system builders Hard to gain trust from users for app developers

Revisit: The permission granularity dilemma

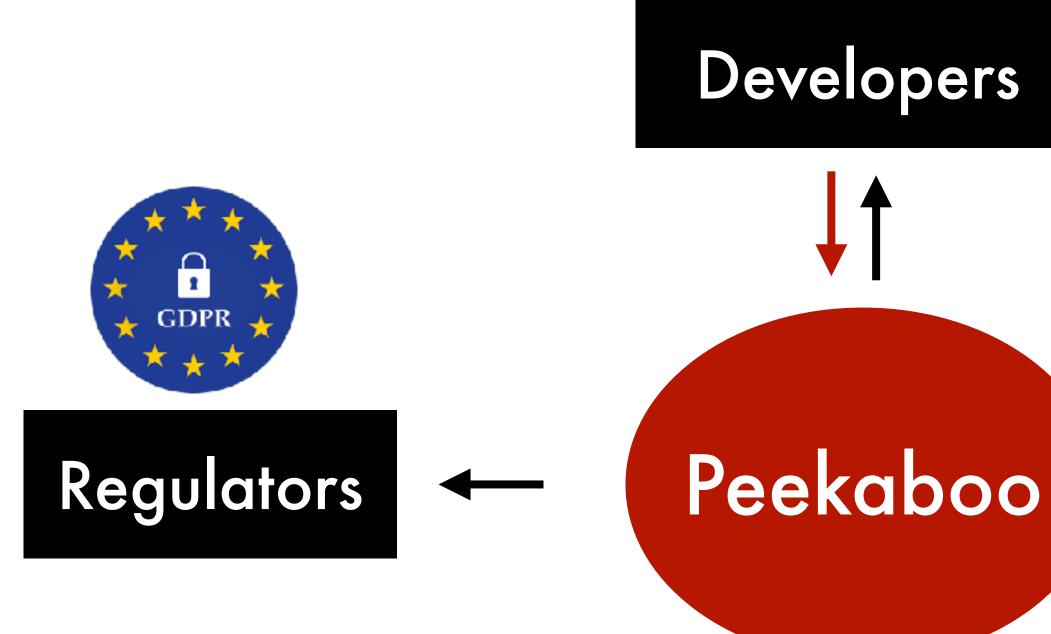
More fine-grained permissions.→ Better default options.

Machine-readable permissions
→ Easier to audit.

→ Better ecosystem. Good privacy drive-out bad privacy.
 → Aggregated management.

Decomposable (operator-based) permissions.→ Fast development.

Let the good privacy drive out the bad privacy



1. Identify overaccess



- 4. Free privacy features 5. Gain users' trust

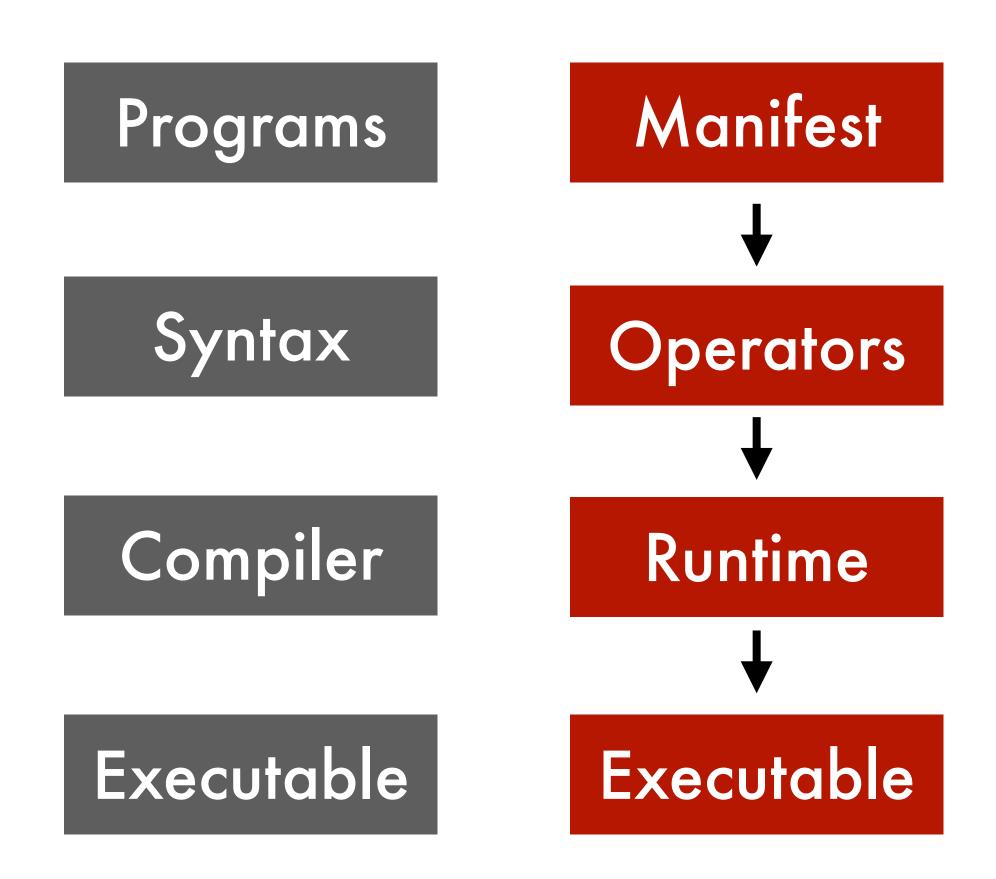


Privacy advocates

- 2. Rank app developers
- 3. Independent privacy features

- Users
- 6. Centralized and unified management 7. Enforceable controls

MPF is a simpler compiler architecture.



A fixed set of operators

A trusted runtime with a small set of pre-loaded implementations

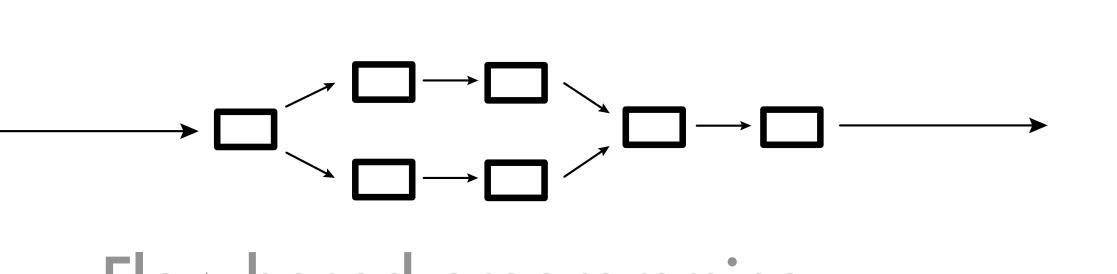
Talk outline

- 1. Modular Privacy Flows (MPF) in a Nutshell 2. Why MPF
- 3. How MPF
- 4. When and when not MPF
- 5. Future Work

Recap: Privacy as modular information flow

Who (which app) sends the data? Where the data is being sent to? What data is being collected? Why the data is being collected? How the data is being stored?







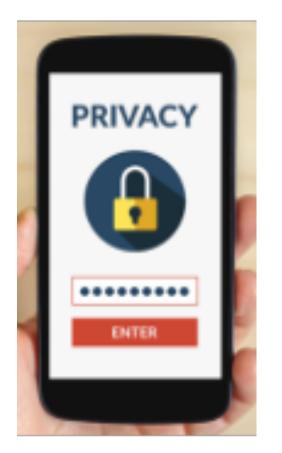
Future work: Broader application domains



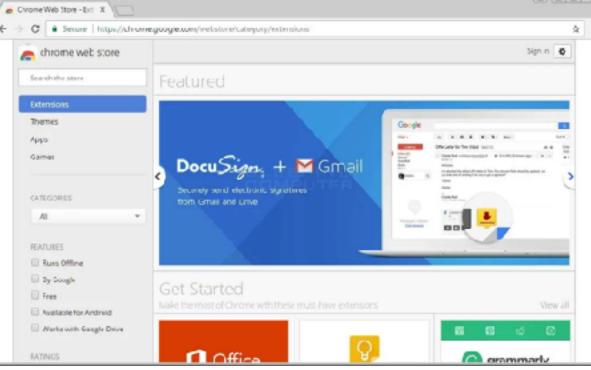
Peekaboo



Smart City



Mobile apps?



Chrome extensions?

Vox

recode

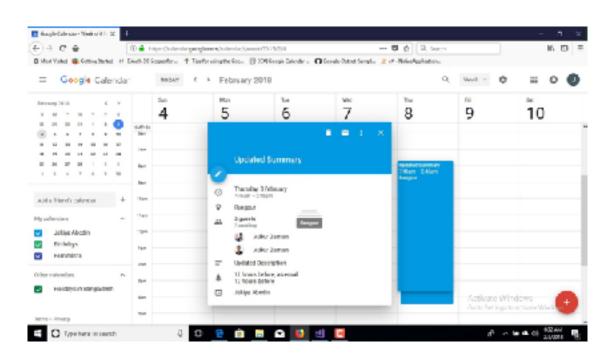
Here's how Facebook allowed Cambridge Analytica to get data for 50 million users

Facebook says it isn't at fault.

By Kurt Wagner | Mar 17, 2018, 3:47pm EDT

f 🔰 🕝 share

Social network?



Personal data API?



ChatGPT Plugin?





MPF v.s. Binary permissions

<manifest ...>

<uses-permission android:name="android.permission. ACCESS_COARSE_LOCATION" />

</manifest>

Android Permission Manifest

Allow "Weather" to access your location while you are using the app?

We need to check your location in order to let you know the weather forecast.

Don't Allow

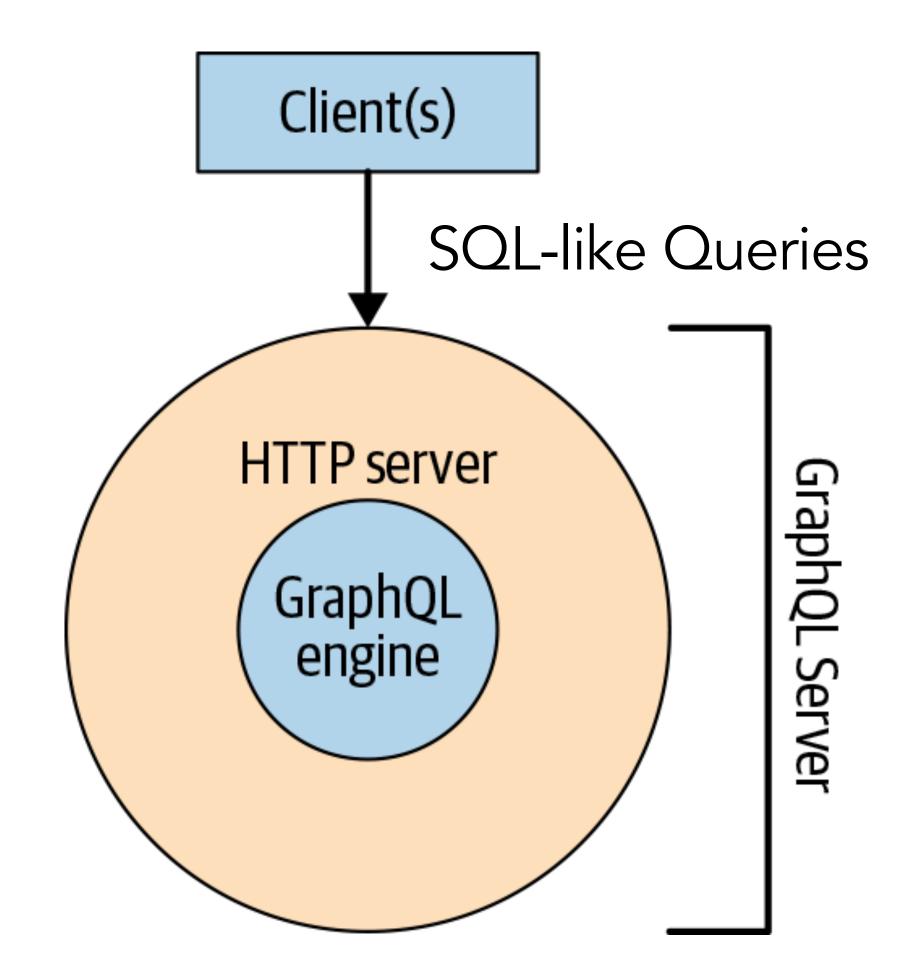
Allow

Popup window

1. System implementation

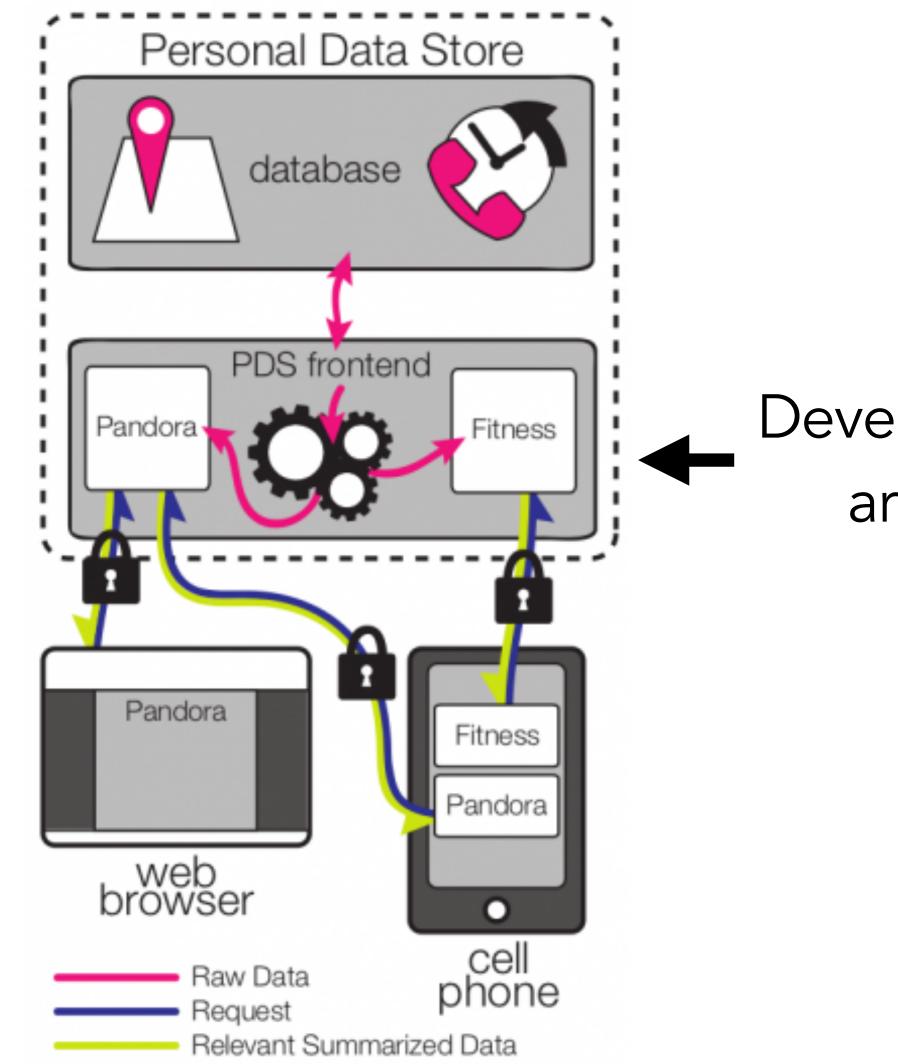
- 2. API complexity
- 3. End-user management

MPF v.s. Database approaches (e.g., GraphQL)



- 1. Flexibility/Extendability
- 2. Auditability

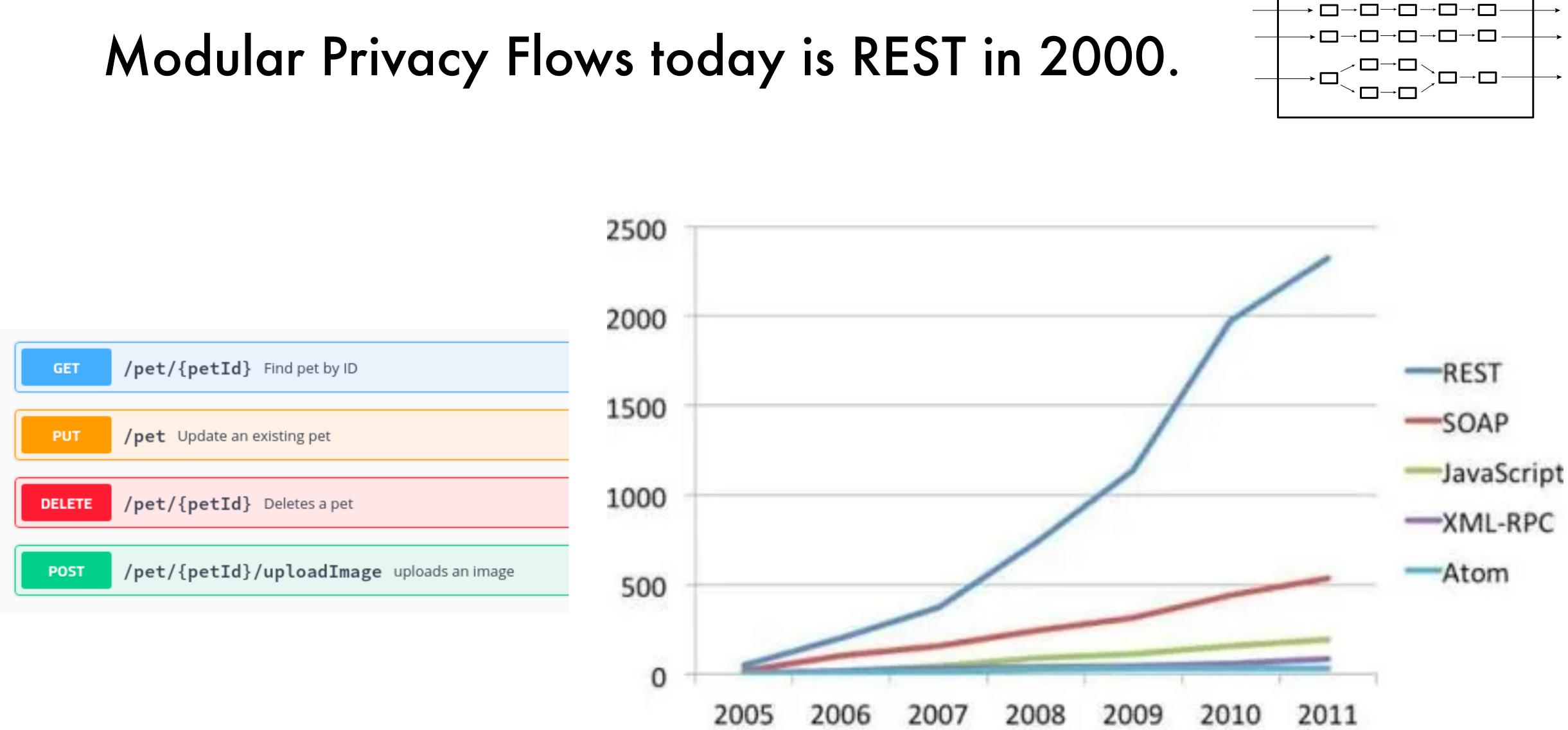
MPF v.s. Remote Code Execution



Developer-uploaded arbitrary code

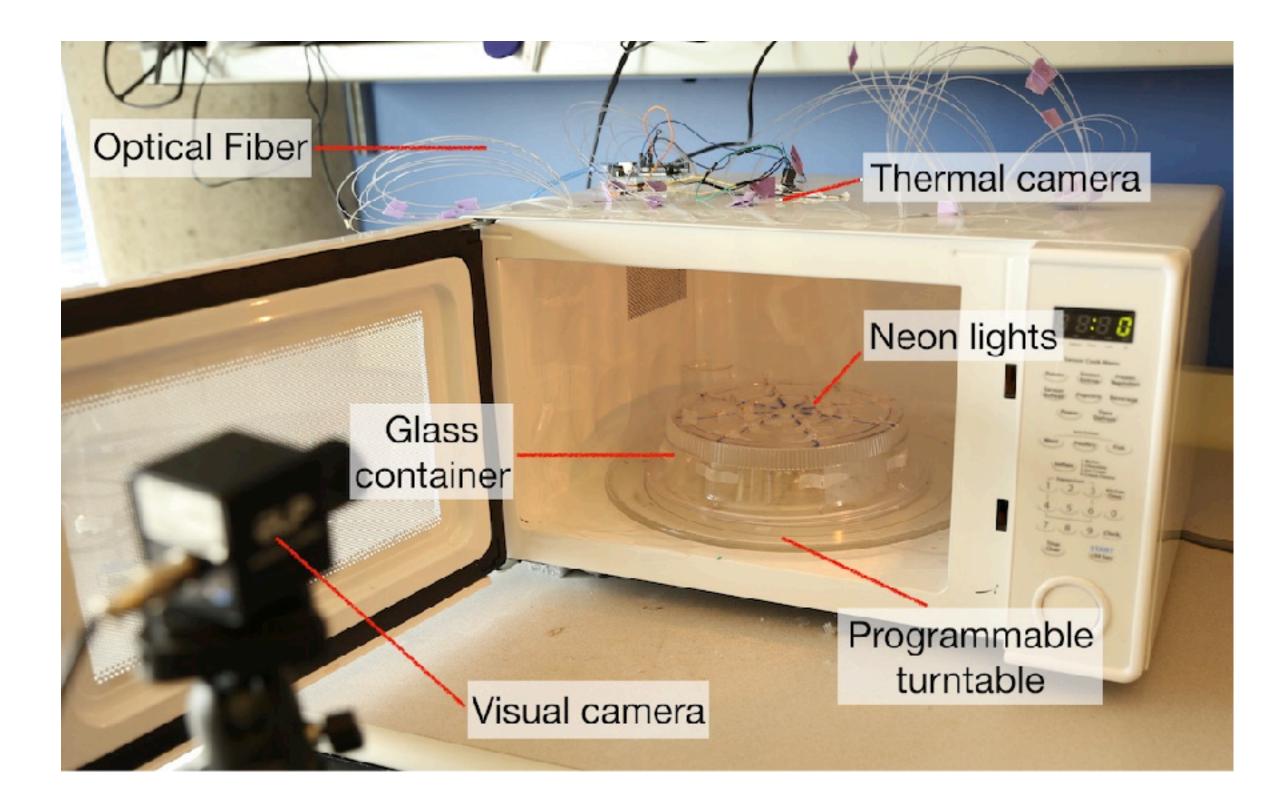
- 1. Auditability
- 2. App development
- 3. Security





>
-
 \rightarrow
 →
-

A story behind Modular Privacy Flows Software-defined Cooking



No Turntable



SDC Uniform Heating



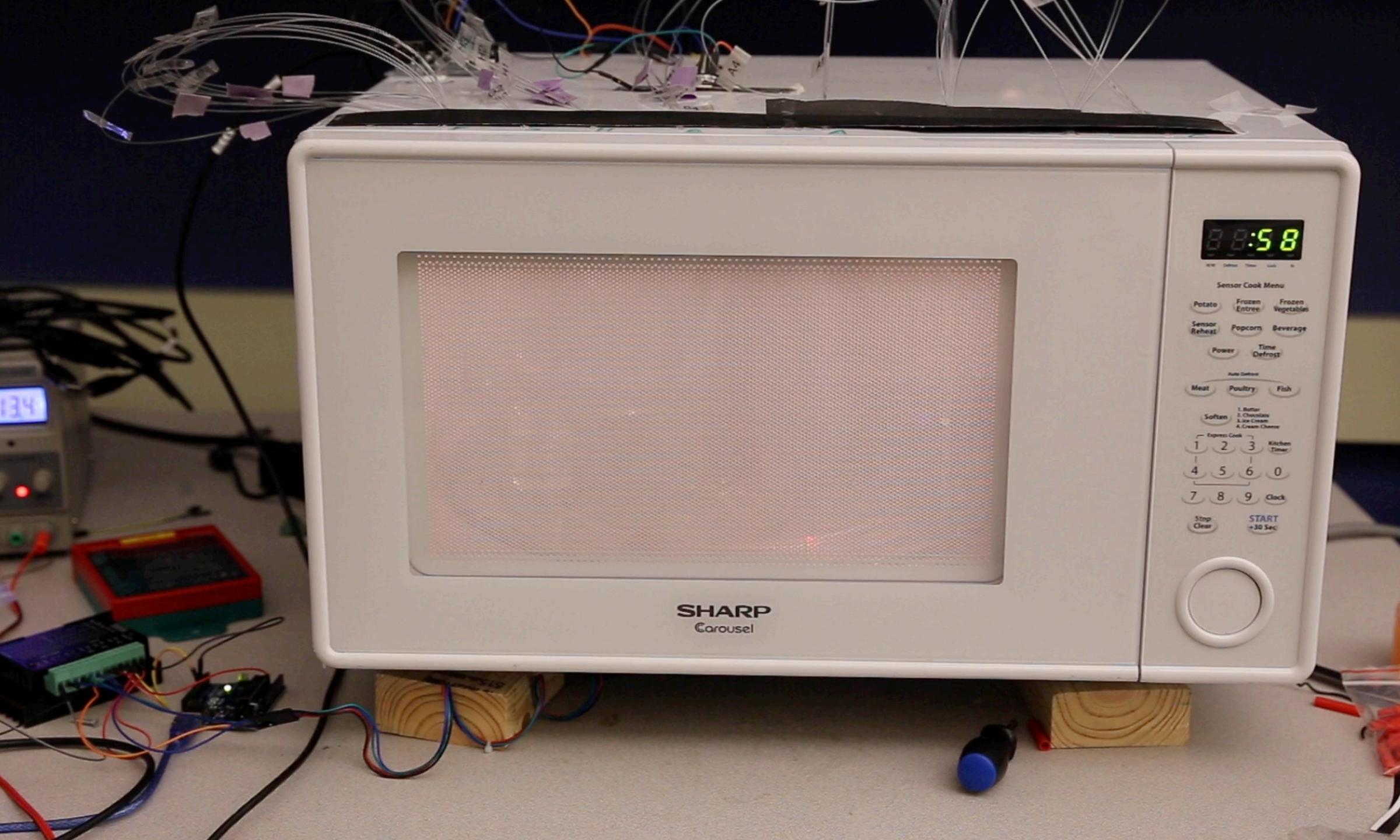
Default Turntable



SDC Arbitrary Heating



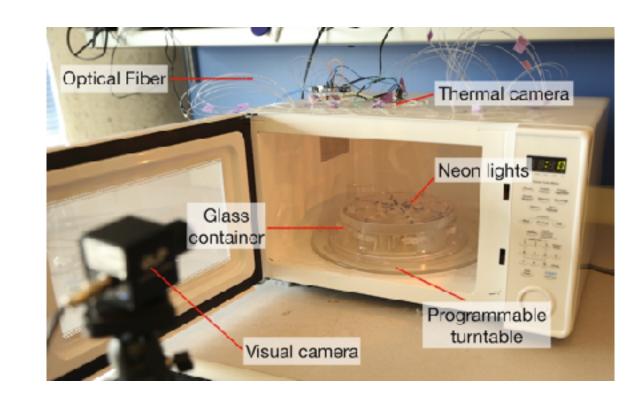






A story behind Modular Privacy Flows

Software Defined Hardware



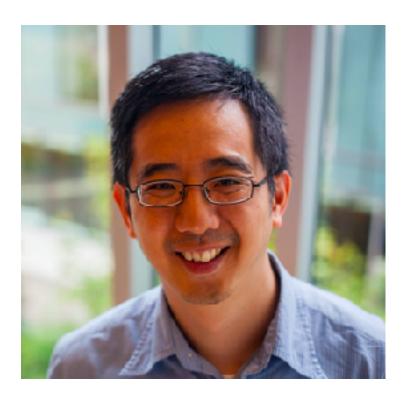


+Operators





+Runtime



+Manifest

Talk outline

1. Modular Privacy Flows (MPF) in a Nutshell 2. Why MPF

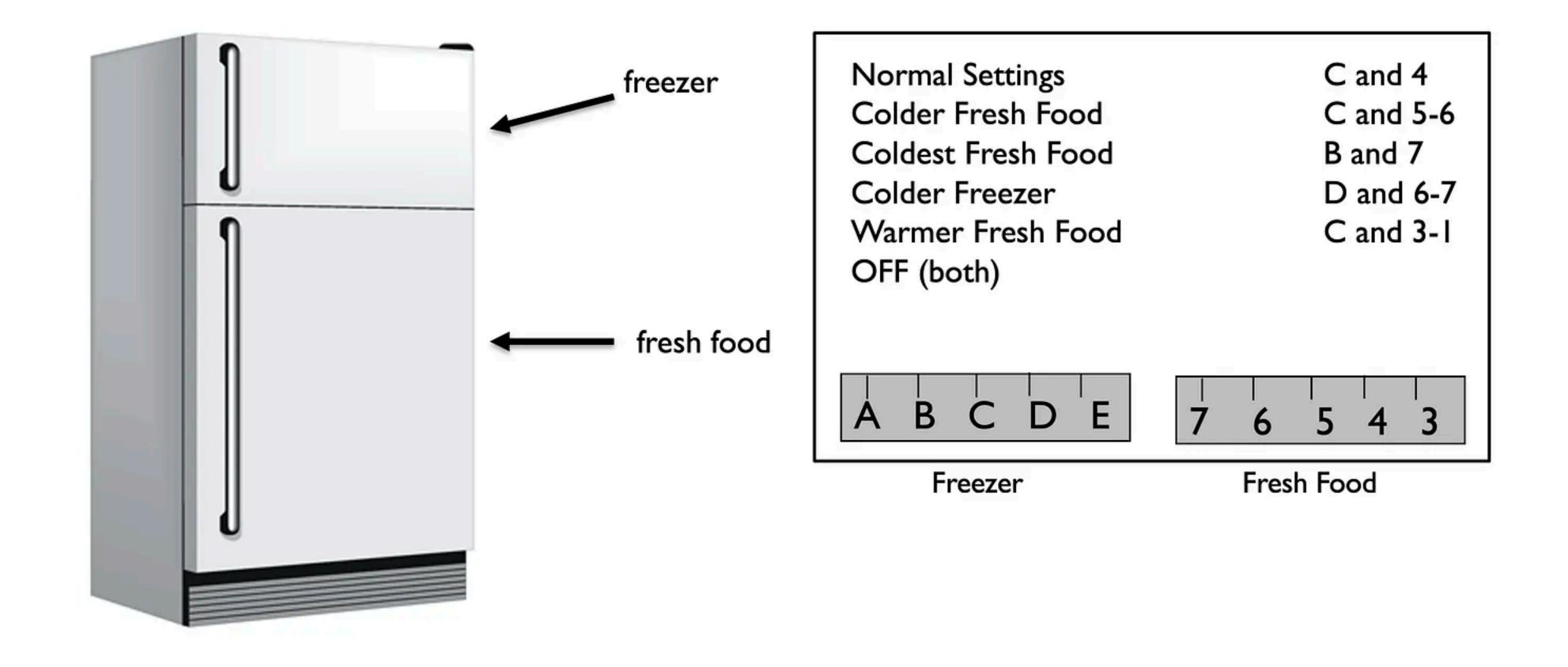
3. How MPF

4. When and when not MPF

5. Future Work

Computing systems are increasingly complex, we need something on the order of a single page demonstrating that the system will work as intended.

Mental models

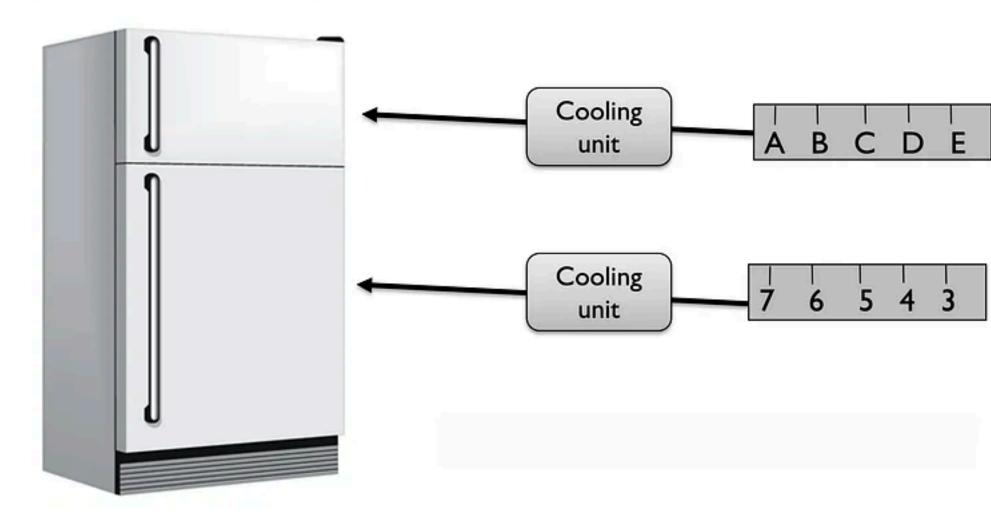


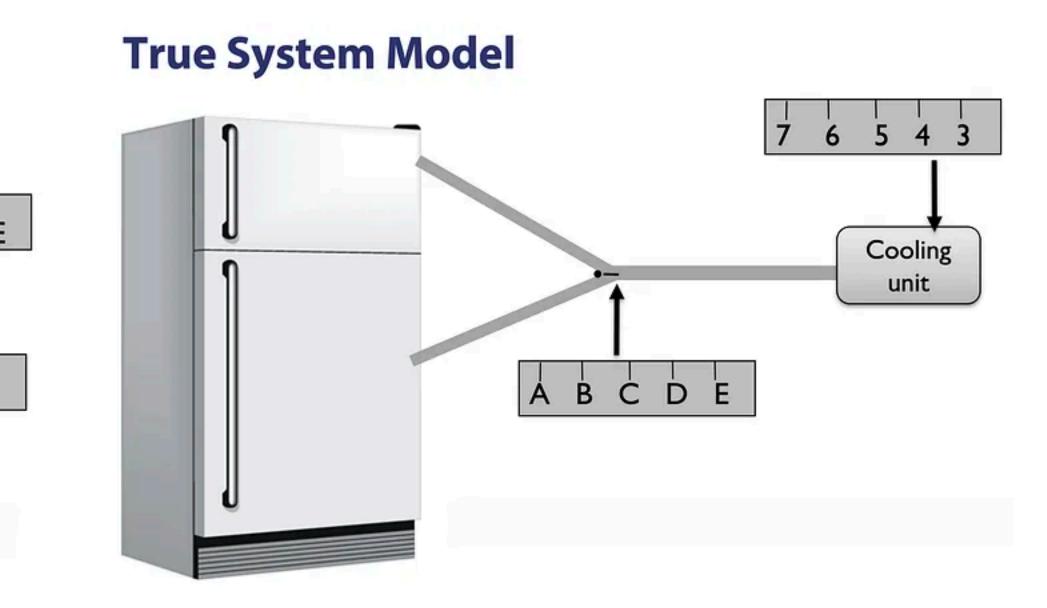
The design of everyday things



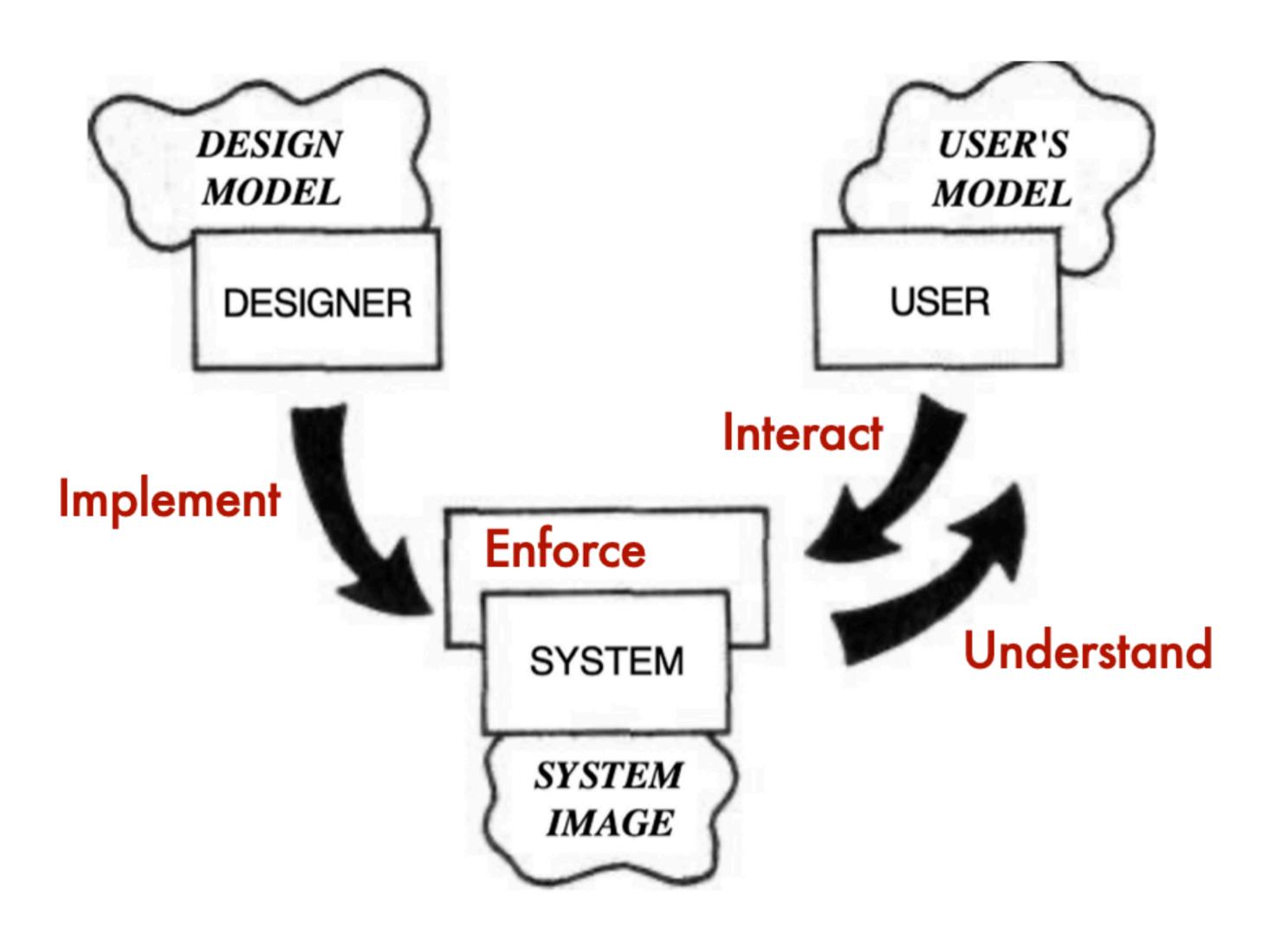
The Gap

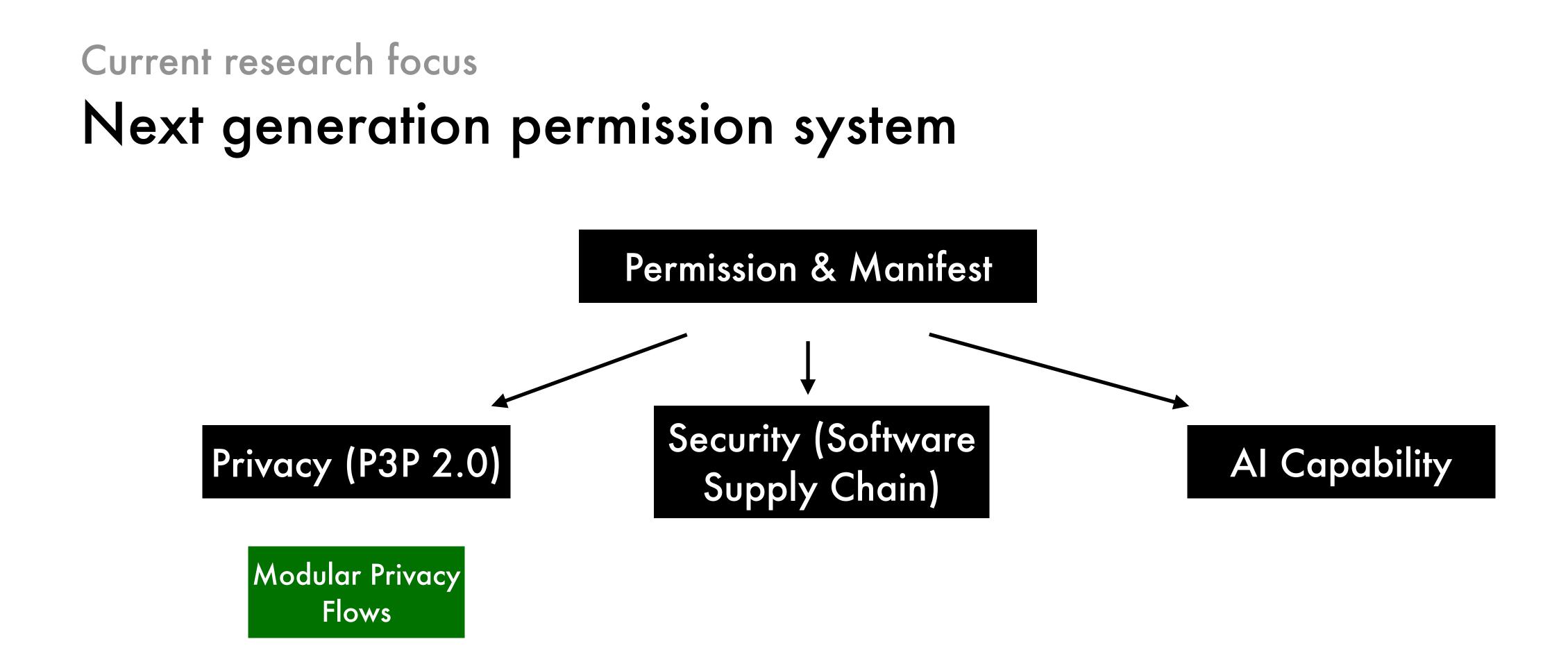
User's Conceptual Model





Our approach





Data Smith Lab is recruiting!

We study the security and privacy of data systems by researching the people who design, implement, and use these systems.

Contact: <u>haojian@ucsd.edu</u> http://haojianj.in/

Acknowledgment







