Operator-based Data Access for Protecting Users' Privacy

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

Haojian Jin
haojian@ucsd.edu

Data science is on fire.

Privacy is giving a bad name to us.

Posted by u/Daridla 3 years ago

Shame on you Google its sad how evil you've become

Beyond sick of the Google spy network. I talk about something with someone in real life and I see an ad five minutes later. It's happened three times now. Tommorow I'm going to be calling consumer affairs or something this is seriously just ridiculous at this point.

https://www.reddit.com/r/privacy/comments/8sym9v/shame_on_you_google_its_sad_how_evil_youve_become/

Our discipline is on fire.

Privacy is giving a bad name to us.

Public Complaints

144,376

complaints from EU citizens

Google

£50M

French Data Protection Authority, Jan 2019

Marriott INTERNATIONAL

$123M

UK Data Protection Authority, June 2019

BRITISH AIRWAYS

£183M

UK Data Protection Agency, June 2019

https://www.advisory21.com.mt/statistics-behind-the-1st-year-of-gdpr/
https://www.bankinfosecurity.com/privacy-fines-total-gdpr-sanctions-reach-331-million-a-15790
https://idbbn.com/story/gdpr-one-year-old-and-144376-complaints/
https://www.bbc.com/news/business-48905907

Research mission:

Help developers, users, auditors to protect users' privacy

Developers

Users

GDPR

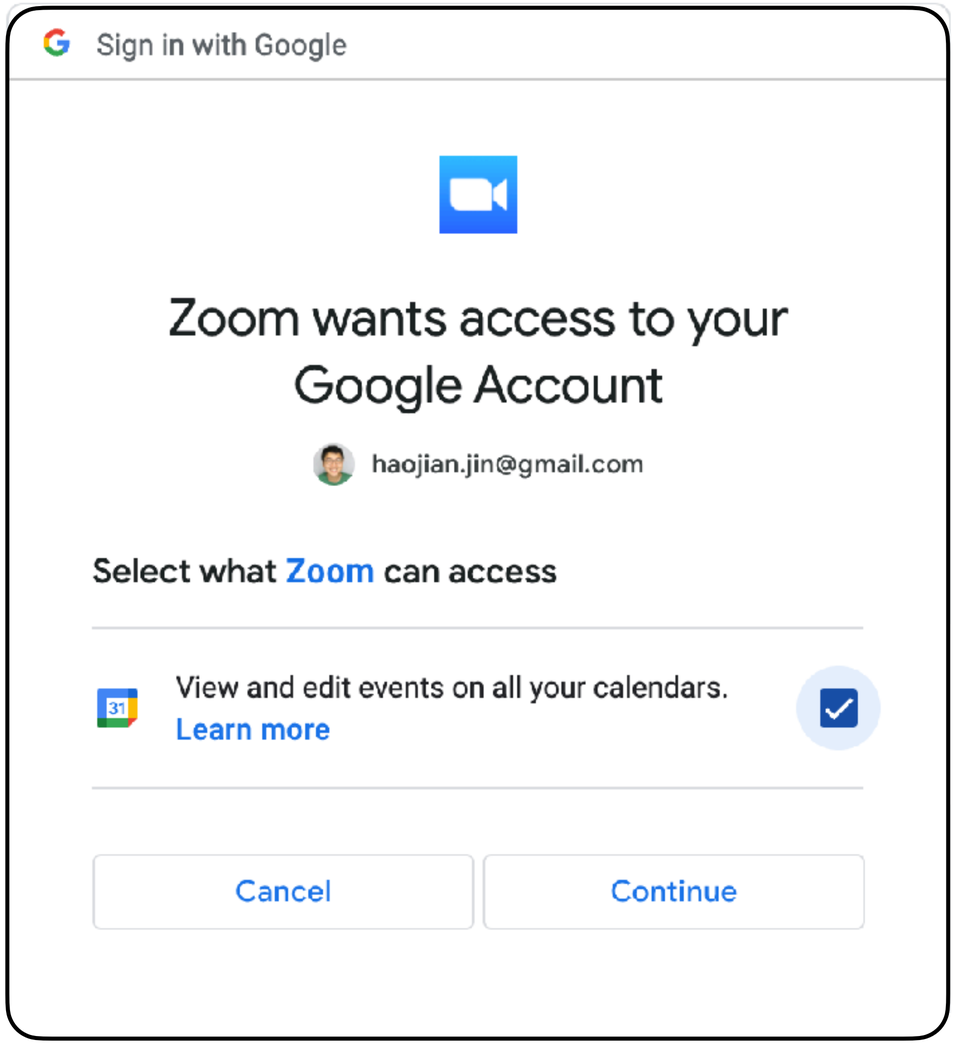
CCPA

FTC

COPPA

Auditors

Zoom accesses **all** your calendar events **continuously**!



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

Google APIs - All-or-nothing binary permissions

Scope	Meaning
`https://www.googleapis.com/auth/calendar`	read/write access to Calendars
`https://www.googleapis.com/auth/calendar.readonly`	read-only access to Calendars
`https://www.googleapis.com/auth/calendar.events`	read/write access to Events
`https://www.googleapis.com/auth/calendar.events.readonly`	read-only access to Events
`https://www.googleapis.com/auth/calendar.settings.readonly`	read-only access to Settings
`https://www.googleapis.com/auth/calendar.addons.execute`	run as a Calendar add-on

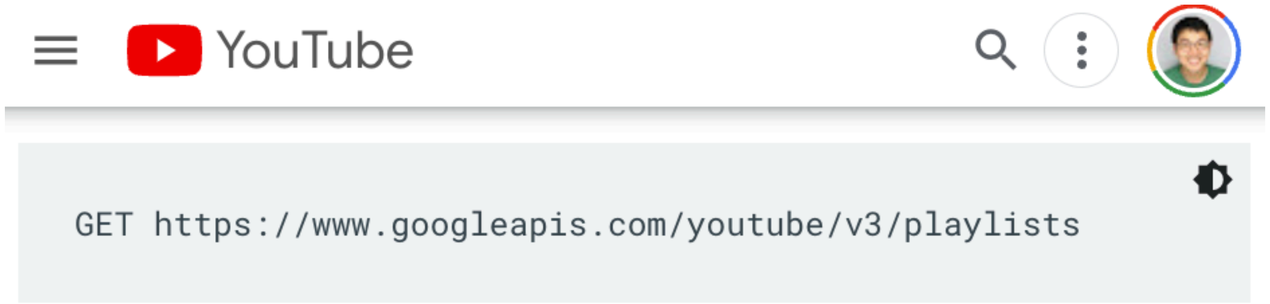
https://developers.google.com/calendar/api/guides/auth

Most applications do not need raw data!

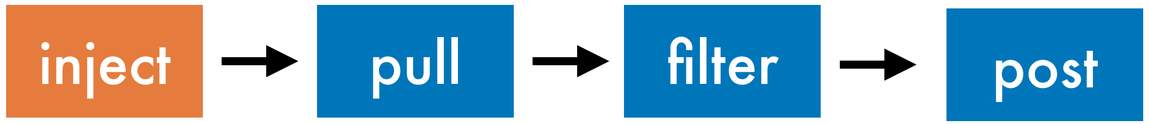
	Question	*Needed data*
Productivity tracking	How busy are you?	# of business meetings
Meeting scheduling	When are you available?	Time blocks

Program data transformation functions using chainable *operators*

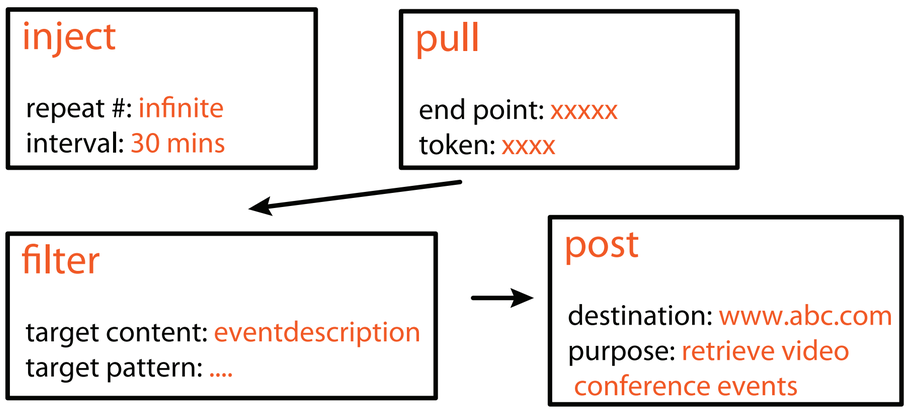
URL-based APIs



Operator-based APIs



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



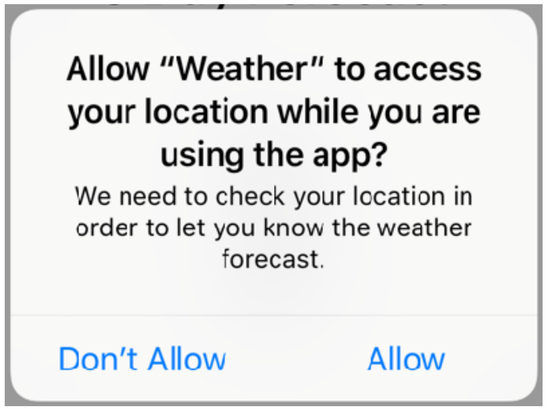

```
@purpose: The app can access calendar events
which contains a zoom link.
ZoomCalendarIntegration{
  // operator [properties]
  inject[...] -> pull Calendar[...] ->
  filter [Zoom join link pattern] ->
  post [Zoom events]
}
```

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

MPF v.s. Binary permissions

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

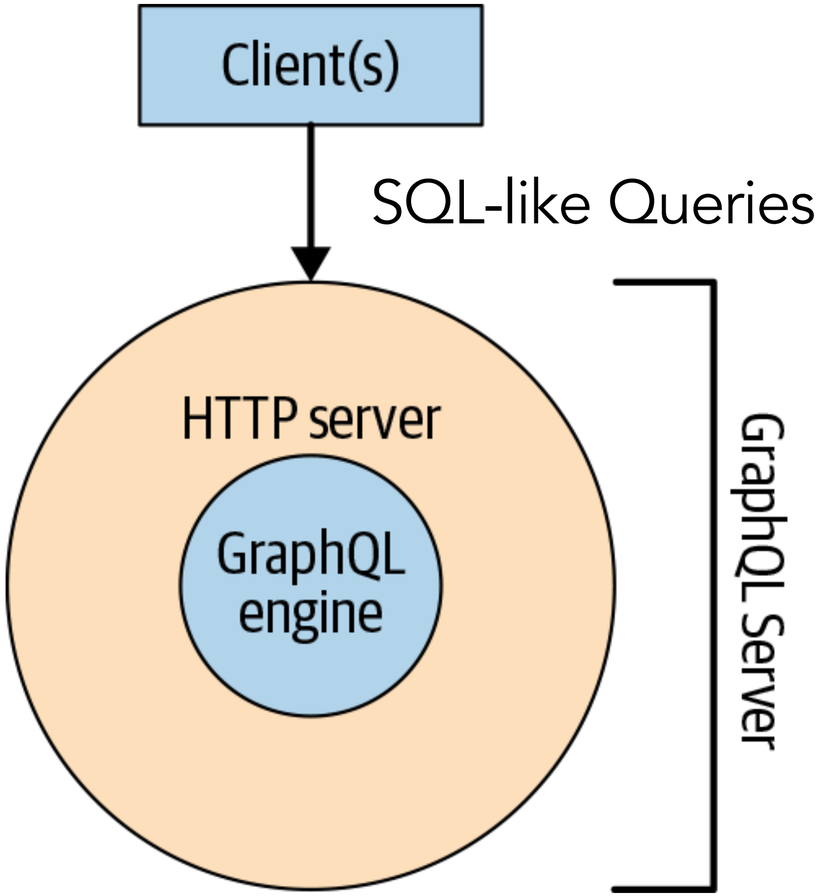
Android Permission Manifest



Popup window

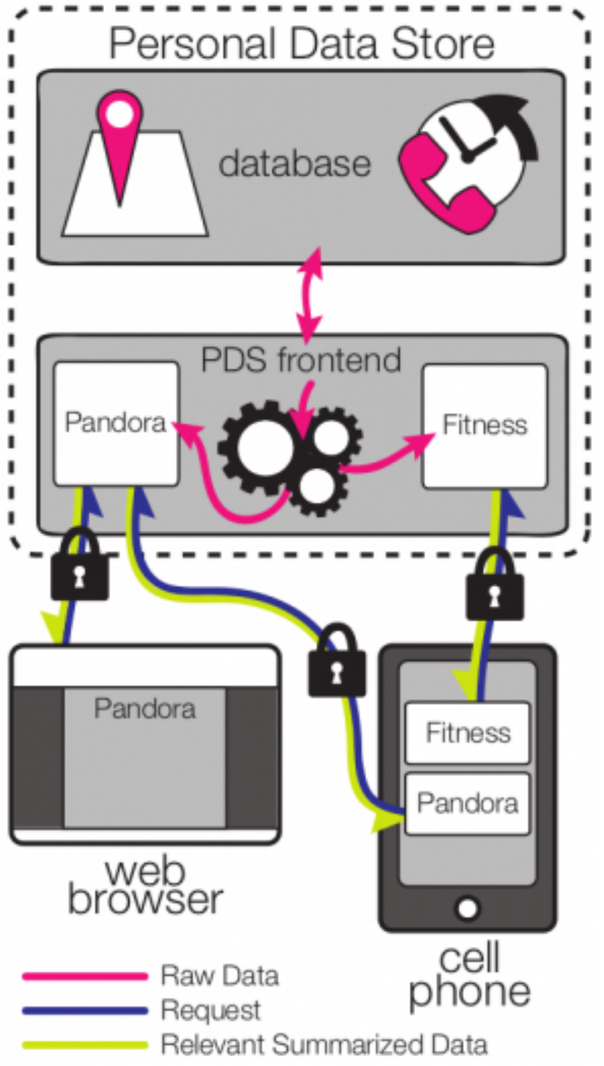
1. System implementation
2. API complexity
3. End-user management

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



1. Flexibility/Extendability
2. Auditability

MPF v.s. Remote Code Execution

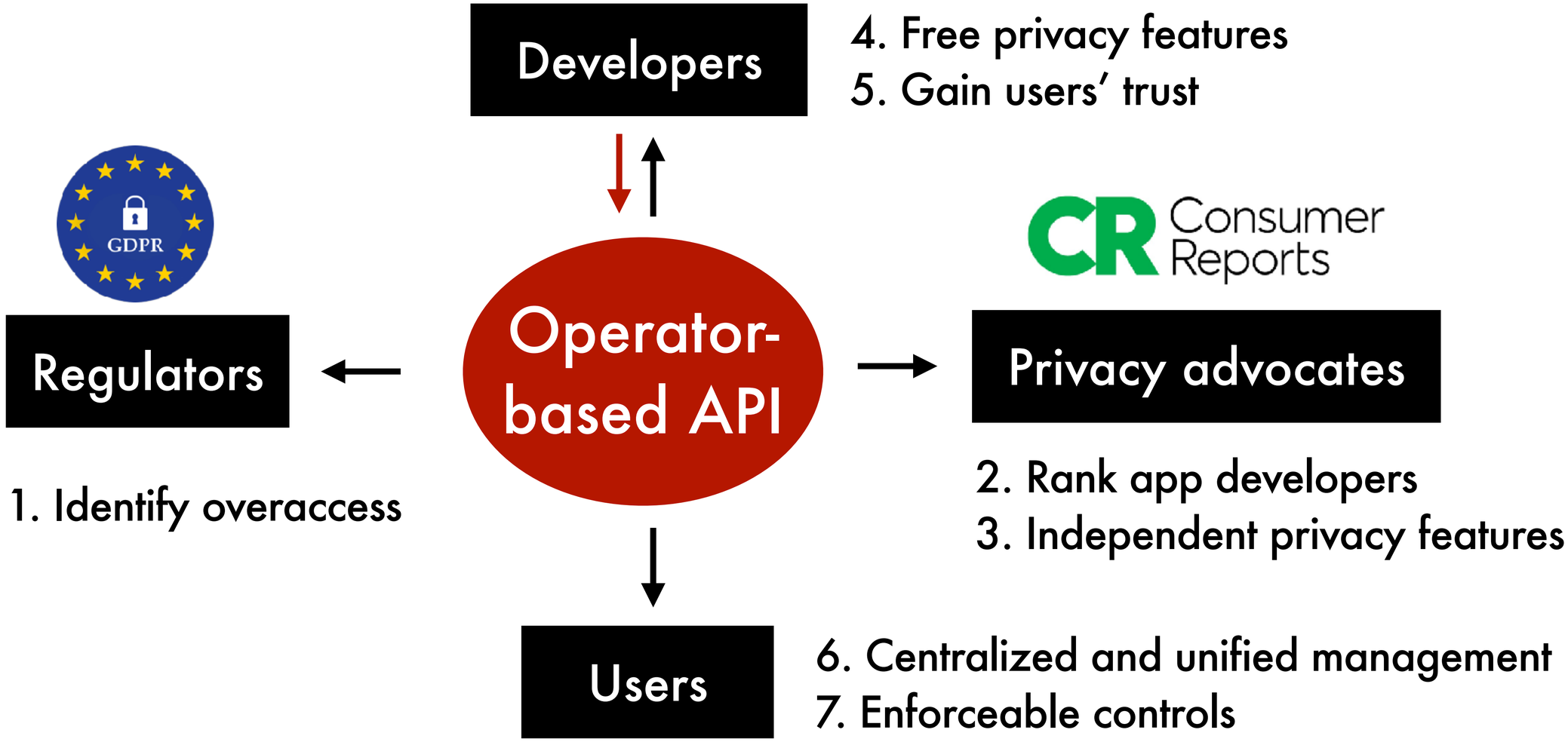


Developer-uploaded arbitrary code

1. Auditability
2. App development
3. Security

https://openpds.media.mit.edu/

Let the good privacy drive out the bad privacy



Broader application domains

Smart Home

Smart City

Mobile apps?

Vox recode

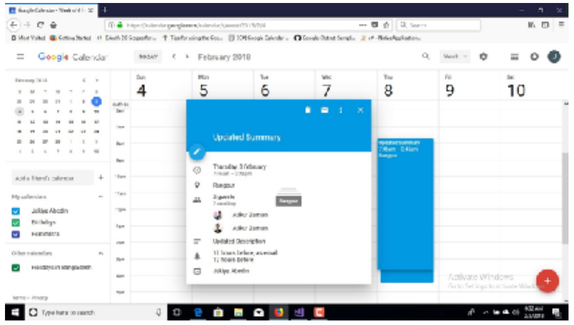
Here's how Facebook allowed Cambridge Analytica to get data for 50 million users

Facebook says it isn't at fault.

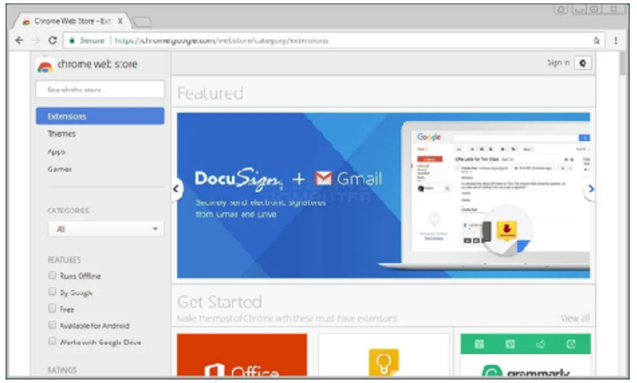
By Kurt Wagner | Mar 17, 2018, 3:47pm EDT

SHARE

Social network?



Personal data API?



Browser extensions?