Session 8: Toward a More Private Web with Browser Advertising APIs

Mark Chen (lecturer on Zoom), assisted by
CU CS B.S. student and CS3 researcher
Giorgio Cavicchioli (in Person)
CU CS M.S. student and CS3 researcher

Session Goals

- Review what Cookie does. How do advertisers and publishers rely on Cookie today (clicks, conversions, performance)?
- How to accomplish what Cookie allows ad-techs to do without enabling individual user-level tracking.
  - What is privacy-preserving attribution (PPA)?
  - What is Cookie Monster?
  - What is differential privacy?

Outline

00:00–00:15 — Warm-up: Reversing the Problem

00:15–00:45 — Mini-Lecture: Privacy-Preserving Advertising: Draft Standard and Our Contributions to It

00:45–01:15 — Activity: “Explain Cookie Monster to a Non-Techie”

01:15–01:30 — Break

01:30–02:10 — Design Activity: The Ideal Private Web

02:10–02:40 — Share-Out + Discussion

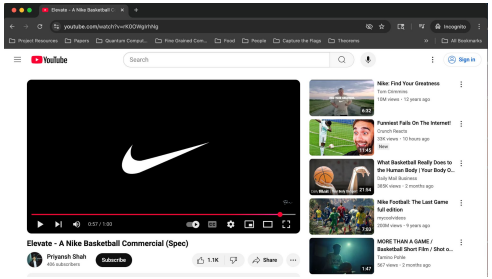
02:40–03:00 — Wrap-Up + Project Reflection

Mini-Lecture:

Privacy-Preserving Advertising: Draft Standard and Our Contributions to It

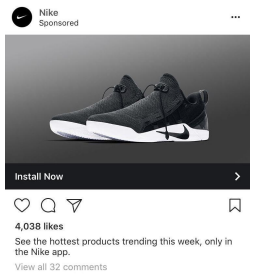
Ads measurement - Canonical example

Consider a user journey:

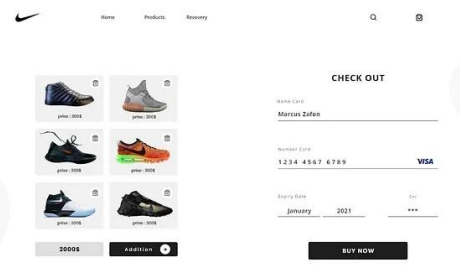


t_0: YouTube Ad

$t_{1-(i-1)}$: Unrelated actions



t_i: Instagram Ad



t_{i+1}: Make Purchase

How does Nike know what ad(s), which we call "**impression**(s)," contributed to the purchase, which we call an "**conversion** event" (can be clicks, conversions, or performances, etc.)?

- **Cookie**: Keep track of your identifier across sites
- **Unique extension**: When you click on link from t_i to trigger the purchase in t_{i+1}, the link could have a unique extension that identifies what directed you to the purchase
- **Questionnaire at the end of purchase**: "How did you find out about this product?"

Ads measurement – Short discussion

What are the pros and cons of each of the methods from previous slide?

Method	**Description**	**Pros**	**Cons**
Cookie	Keep track of your identifier across sites	Everything you need to make the most accurate measurement	They track everything
Unique extension	When you click on link from ti to trigger the purchase in t_{i+1}, the link could have a unique extension that identifies what directed you to the purchase	Direct without over-the-board tracking	Really hard to know the significance of an impression from a long time ago. In our example, the impression from t_0.
Question-naire	At the end of each purchase, “How did you find out about this product?”	No tracking needed at all	Very subjective, biased, and not standardized

The big question:
How to imitate Cookie with the complete, exact tracking?

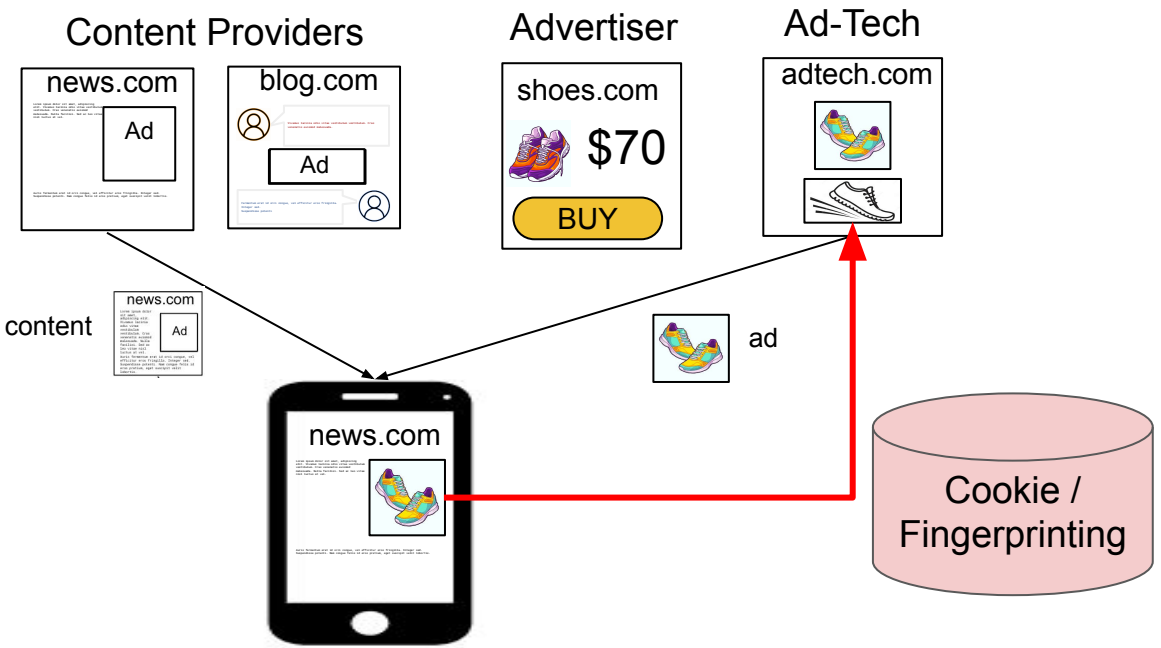
That is what Privacy-Preserving Attribution (PPA) wants to do: “Aggregate-level answers without per-user tracking”

Our attempts at this:

- Cookie Monster → Privacy-Preserving Attribution (PPA)
- Big Bird
- More works to come using similar techniques!!

Overview of how Cookie works in the Nike example

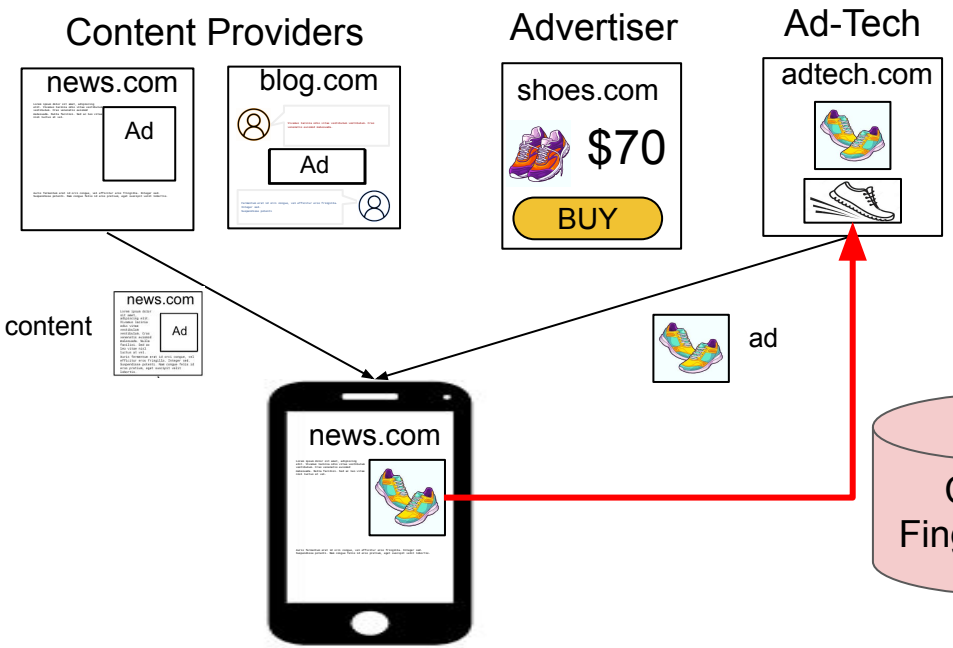
- Recall from lecture 6 how Cookie works in general
- Now, let's think about the Nike example with the Cookie as a "black box"



- Cookie gives ad-techs the complete knowledge of user interactions, on which Ad-Tech can run any analyses it wants to…

Overview of how Cookie works in the Nike example

- But really, what Ad-Tech needs for its honest application, which is to measure ads effectiveness using Cookie, is the following **attribution analysis**:

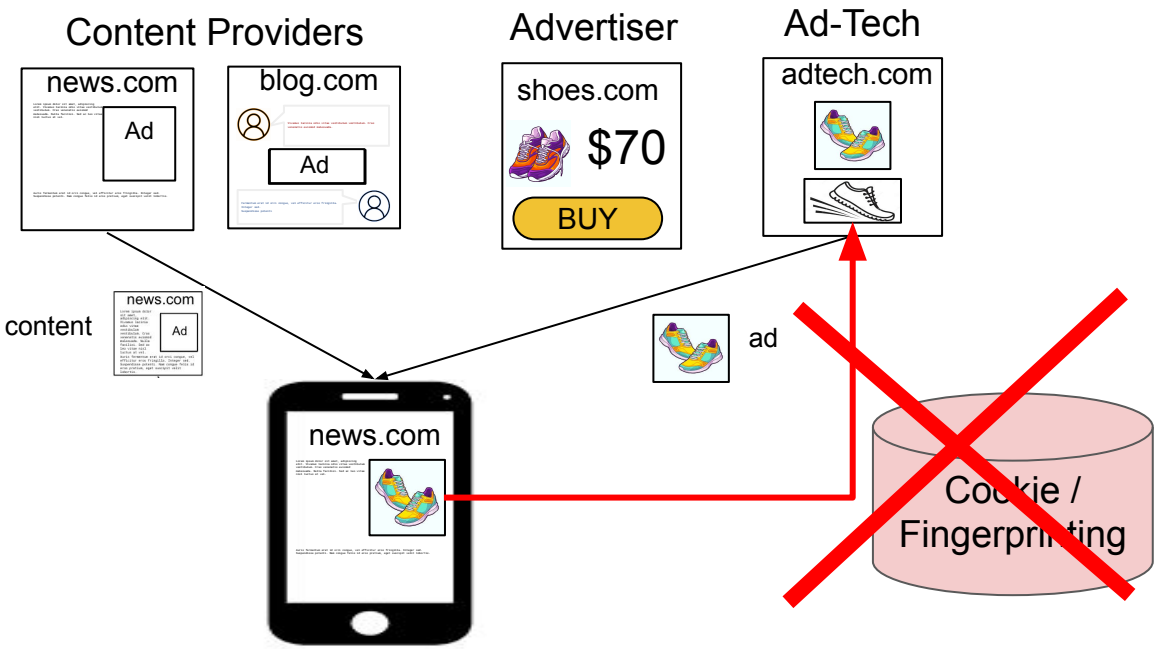


Attribution analysis of a conversion event happens at here, after full Cookie tracking record is provided!

- But with the knowledge that Ad-Tech is given by Cookie, its ability is **not necessarily** restricted to ads analyses. Particularly, these knowledge enables **per-user tracking**!

Overview of how Cookie works in the Nike example

- But really, what Ad-Tech needs for its honest application, which is to measure ads effectiveness using Cookie, is the following **attribution analysis**:



- So let's make it less powerful. What can we do?

Differential privacy (DP)

- What we want to guarantee after removing Cookie? The trade-off:
  - Ad-Tech providers: to have enough information to derive useful ads attribution findings
  - Individuals: No matter what Ad-Tech providers do with the data they are provided, they cannot gain any significant knowledge about an individual
- This fits into the guarantees of DP
  - Data provided with DP guarantees limit *individual-level information* leakage
  - While analyses run on such data has *rigorously formulated error bounds*
  - It's like compare:
    - tracking to a camera following each shopper vs.
    - a sensor at checkout that just counts how many used a coupon
- There are several variants to DP guarantees. One popular one literally translates to the following sentence: "For any individual x in a dataset, whether or not x is contained in the dataset, the outcome of analysis change by at most ϵ with (1-δ) probability." This is what's known as the (ϵ, δ)-DP.

Good and bad use cases of DP

Based on the high-level definition of DP, what are and what are not good use cases of DP?

- Publishing census-style counts / histograms while limiting **re-identification** risk ✅
- Location density maps: Sharing aggregate flows (e.g., crowd movement, traffic) without exposing any individual's path ✅ (A project of ours, CityOS in session 11, looks into this perspective!)
- A/B experiment analytics: Releasing conversion / click-through aggregates across many users without leaking a single user's behavior ✅
- Tiny or sparse datasets: Too little redundancy—DP noise swamps signal or utility collapses ⚠️ (even the US government falls into this pitfall: next slide)
- Individual level learning, like personalized recommendations ⚠️
- Already public / non-sensitive data: Applying DP to openly published weather readings or stock prices (adds noise, no extra safety) ⚠️ (We model such information as **public information** not protected by DP)

DP doesn't prevent drawing statistical conclusions. E.g. does smoking lead to lung cancer? Think about what is guaranteed and what is not guaranteed!

Anecdote about DP

Recall one caveat when using DP we mentioned: “Tiny or sparse datasets: Too little redundancy—DP noise swamps signal or utility collapses”

Monowi, Nebraska

Article Talk

From Wikipedia, the free encyclopedia

Monowi (/məˈnoʊwaɪ/ *MON-oh-wye*) is the least populous incorporated village in the United States by population. It is in Boyd County, Nebraska, United States, and received national and international[4] attention after the 2010 United States census recorded only one resident in the village, Elsie Eiler, who serves as its mayor, librarian, clerk, and treasurer.[5] Although the 2020 census listed Monowi's population as two,[6] this was confirmed to be an example of differential privacy in the census data; Eiler remains the town's sole resident.[7]

What’s wrong? Privacy parameters in this case are astronomical, making the data almost not meaningful.

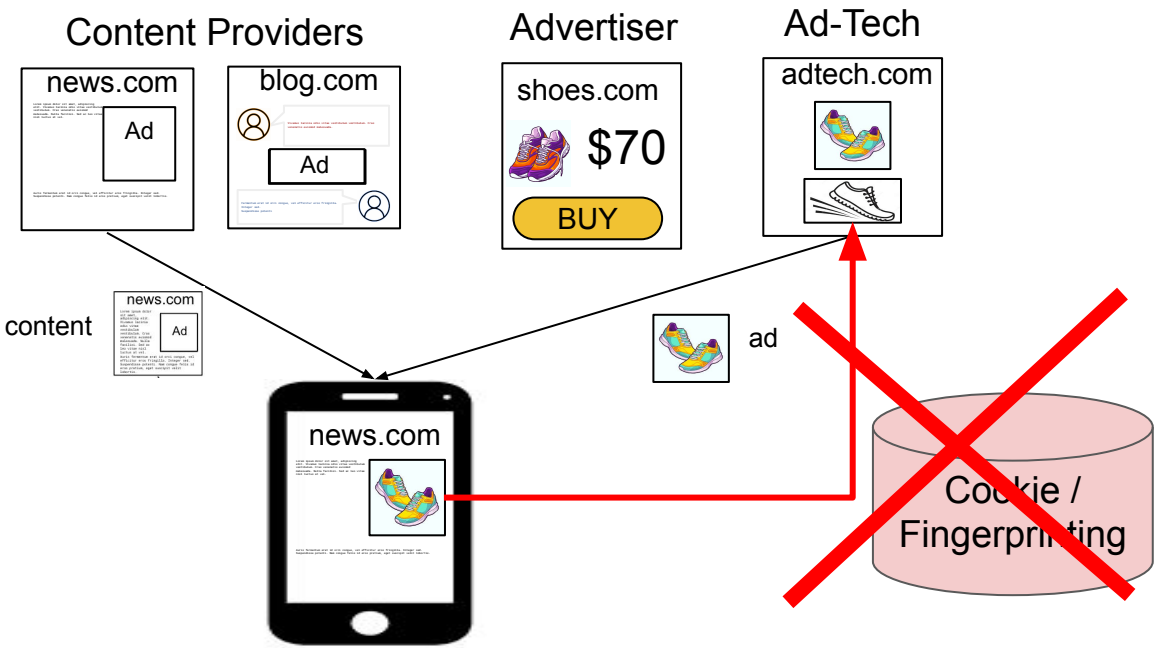
- **Good news:** Many important organizations are adopting DP in their practices, including the U.S. census!
- **Bad news:** There’s still a long way to go for deployments in many cases…

Learn more about DP?

- This class is how I first learned about DP: http://www.gautamkamath.com/courses/CS860-fa2022.html
  - It recommends two conical textbooks in the field
  - The professor is a very productive and active researcher in the theoretical aspects of DP
  - The class link posts very comprehensive materials:
    - PDF typed notes of each class, along with written notes
    - Class recordings
    - Useful external references etc
  - Do not stress about watching all of it! Take it slow. For starters, finishing watching the first 9 lectures already give almost all you need to get started.

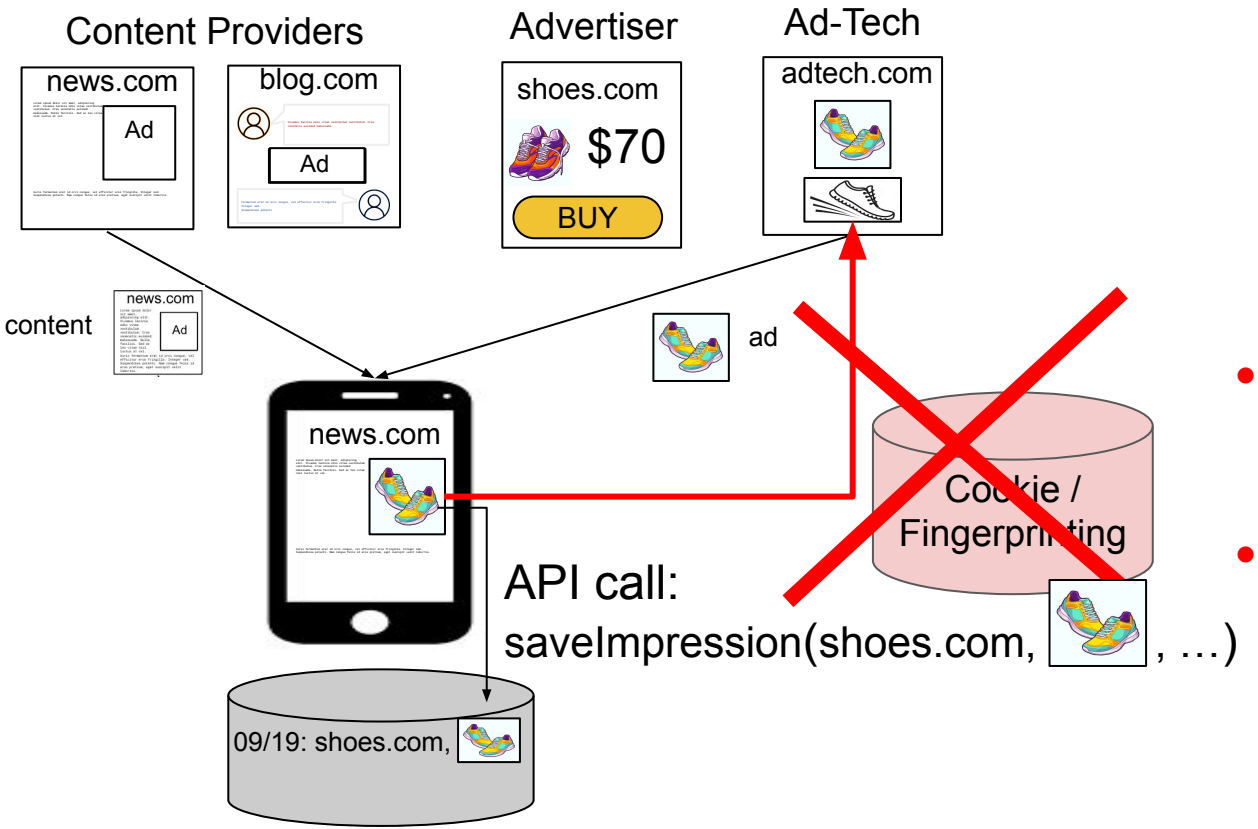
How? Cookie Monster! — Review our goal

- Back to our goal of replacing Cookie while keeping its abilities to enable ads attribution analysis:



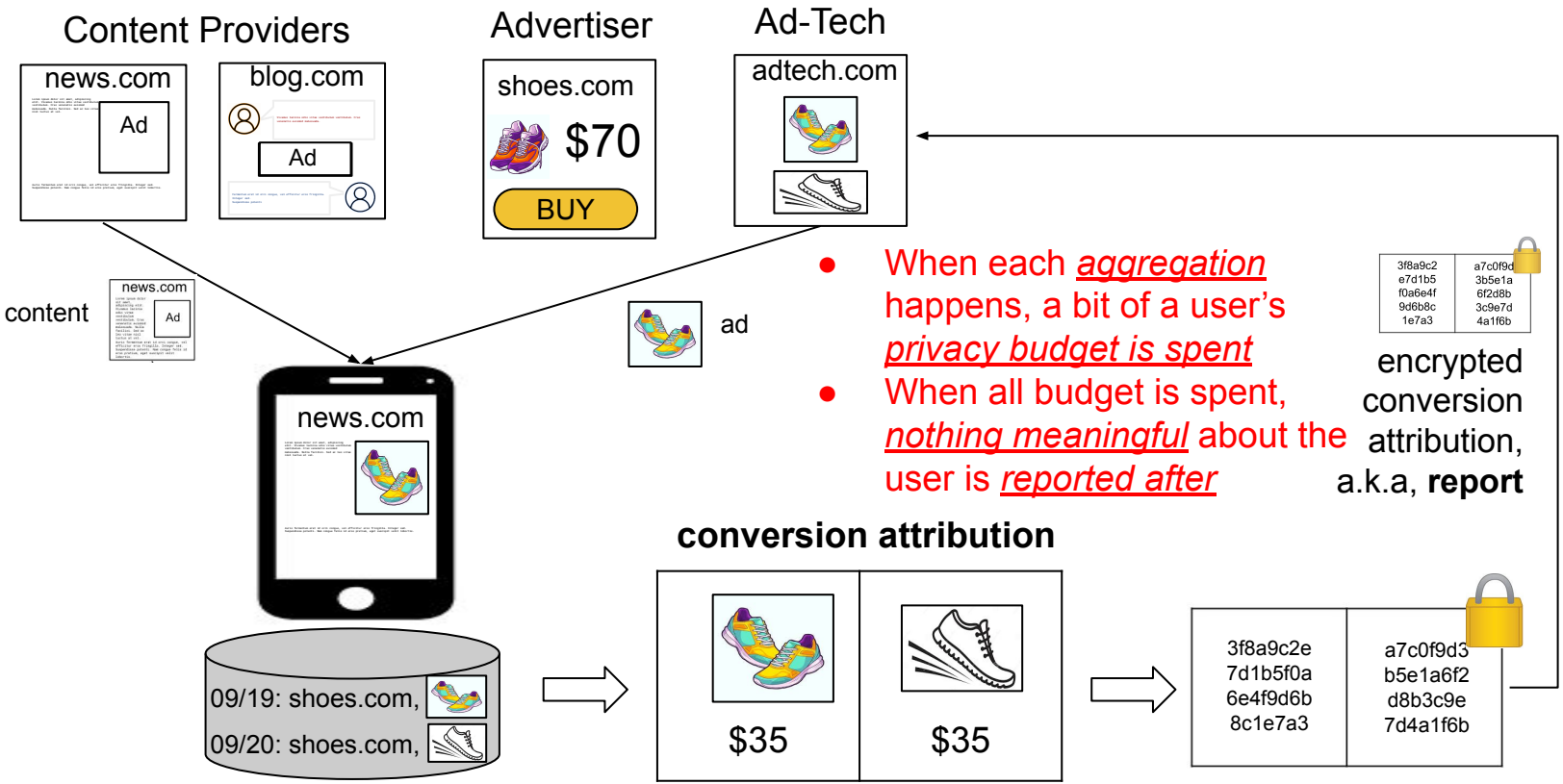
- So let's make it less powerful. What can we do?

How? Cookie Monster! — Impression history



- We can repeat this a few times for each impression up to a conversion
- Say we saved the following impressions on various content providers:

How? Cookie Monster! — Aggregation at conversion



- When each *aggregation* happens, a bit of a user's *privacy budget is spent*
- When all budget is spent, *nothing meaningful* about the user is *reported after*

How? Cookie Monster! — Our guarantees

Our API features (mentioned in red texts in the last slide)	**Guarantees**
“Aggregation”	APIs give just enough info to measure ads effectiveness
“Privacy budget is spent” per aggregation	Privacy budget
“Nothing meaningful reported” after privacy budget is depleted	No per-user tracking

Demo of PPA on Firefox

Giorgio's pdslib Firefox demo:
https://drive.google.com/file/d/1yjEImNjSxdvNOtEfM5OrHz_1jxIB7ivt/view

- Background: Part of our recent work Cookie Monster → Big Bird

- Ask Giorgio anything about this demo. Sample questions are like:
  - Why Firefox (in particular, why not Chrome, Safari, Opera, Edge etc)?
  - How is developing with the Firefox interface like?
  - Is this a good replacement for Cookies?
  - How close is this to being deployed in the real world?

What's Next?

How could advertisers measure ads effectiveness of this type of ads without having a camera literally following you around?

The similarities and differences between these ads and the browser ads seem difficult to articulate, right? This is what we will touch on in session 11 with CityOS!

Activity: “Explain Cookie Monster to a Non-Techie”

Activity: “Explain Cookie Monster to a Non-Techie”

- Imagine you’re trying to explain Cookie Monster/PPA to:
  - Your parent
  - A community member who’s worried about ads
  - A tech-savvy friend who doesn’t trust companies
- Task:
  - Pick one persona
  - Draft a short, 3-sentence explainer or sketch a visual analogy
  - Emphasis: Keep it non-technical, but concrete
- Example: “It’s like a system that lets advertisers ask: Did this kind of ad help at all? -- without ever knowing what you did.”

Design Activity: The Ideal Private Web

Small group brainstorm (whiteboards or slide deck):

- Prompt: “If the web could be rebuilt from scratch for privacy, what would it look like?”
- Incorporate ideas from PPA and Cookie Monster (maybe privacy tools from Session 7)
- Prompt questions:
  - What kinds of APIs would exist?
  - How would companies measure success?
  - What could users still do or see?
- Deliverable:
  - Each group sketches 2–3 rules or features of their ideal web
  - Optionally sketch a new “Ad API” or browser design

Share-Out + Discussion

Small group brainstorm (whiteboards or slide deck):

- Each group presents:
  - Their “explainer” from Activity 1
  - A few features of their “ideal web” from Activity 2
- Whole class discussion:
  - What parts of today’s web could actually work this way?
  - What would be hard to change?

Wrap-Up + Project Reflection

Small group brainstorm (whiteboards or slide deck):

- Prompt:
  - “How would you describe the reality of tracking on the web right now to someone in your community?”
  - “If you had to make a survey question about privacy-preserving APIs for your community, what would it ask?”
- Students write:
  - One potential survey question
  - One sentence they might use in a pamphlet or slideshow
  - Save both to the shared project folder