W3C: Read All About It!

- Code of Ethics and Professional Conduct
- Antitrust and Competition Guidance

Minutes Doc:

https://tinyurl.com/mtzmeubz

Individual Differential Privacy for High-Utility Private Web Measurements

Roxana Geambasu

Associate Professor of Computer Science
Columbia University

Temporarily with Meta

Benjamin Case

Research Scientist
Meta

Session Overview

- We developed a privacy framework for cross-site Web measurements with improved privacy-utility tradeoff, user control, and transparency.
- We designed it for Web advertising measurement, as part of emerging ad-measurement APIs. Mozilla's Martin Thompson will be presenting an API proposal that incorporates it at the PATCG meeting on Friday.
- However, we believe that our framework has broader potential for other Web and mobility measurements.
- We're seeking community input on:
  - Expanding to non-advertising domains
  - Addressing remaining challenges

Outline

- Background on ad measurements and emerging APIs
- Our privacy framework: Cookie Monster
- Discussion on broader applications and bias mitigation

Outline

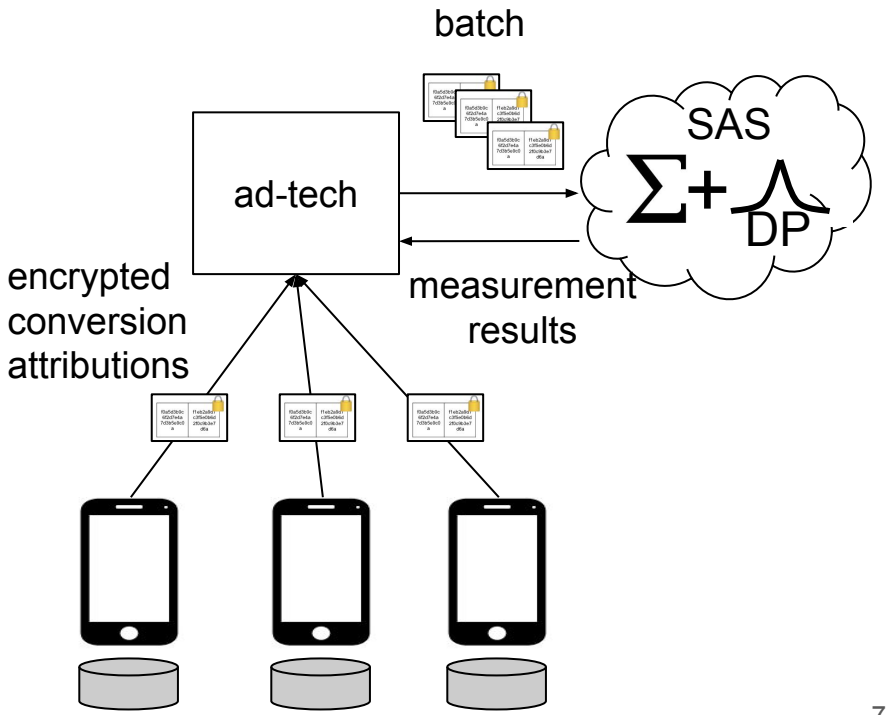
- **Background on ad measurements and emerging APIs**
- Our privacy framework: Cookie Monster
- Discussion on broader applications and bias mitigation

Web Ad Measurements

- Online ads outperform traditional ads by enabling performance measurements, such as **conversion attribution measurements**.
- Historically, this relied on third-party cookies and remote fingerprinting to track users across sites.
- Browsers are seeking more privacy-preserving alternatives: APIs that permit conversion attribution and other measurements without compromising user privacy.
- Example APIs: Google ARA, Apple PAM, and two from Meta/Mozilla--IPA and Hybrid. We focus on **ARA/PAM/Hybrid**.

Browser API Overview

- Browsers record impressions locally and send encrypted conversion attributions to the advertiser/ad-tech.
- Ad-tech forwards a batch of these to a Secure Aggregation Service (SAS), which aggregates them and adds noise for differential privacy (DP).
- SAS uses TEEs or secure multiparty computation to minimize trust in single entities.



Example

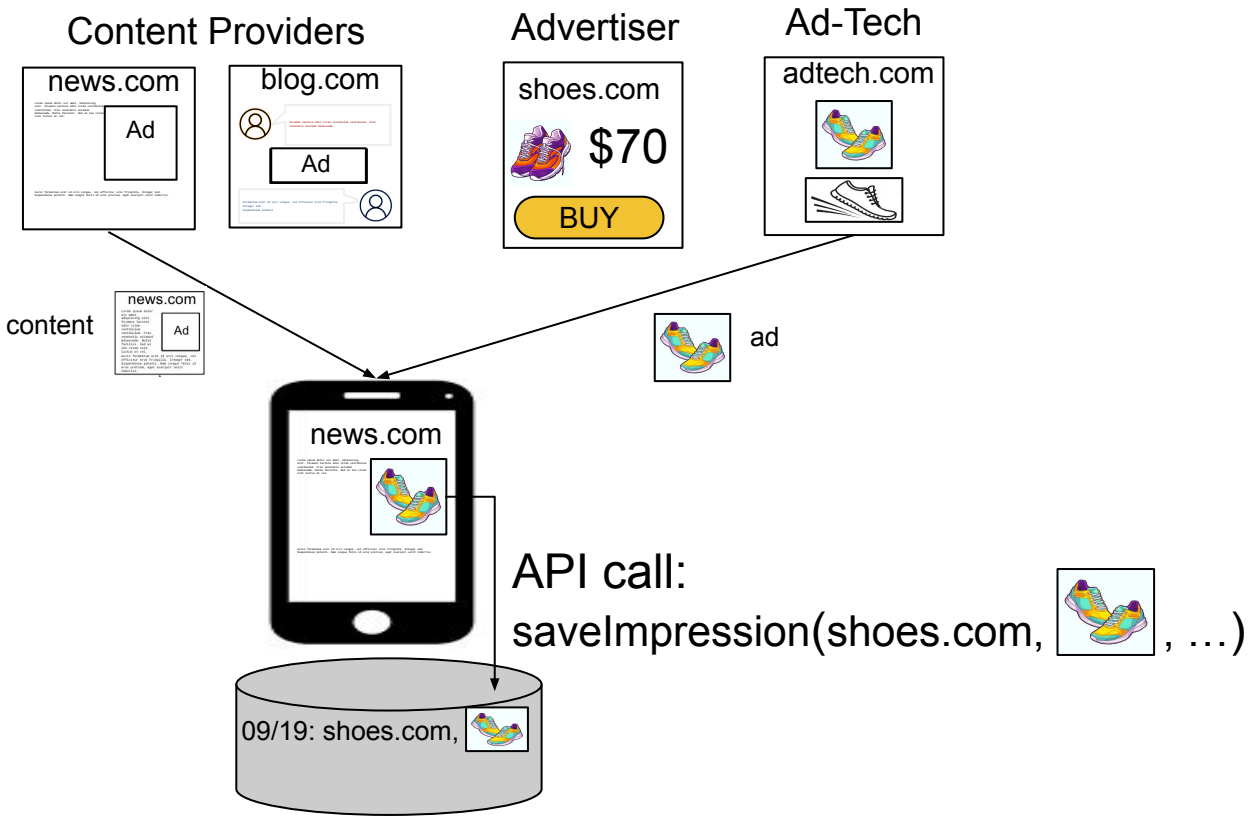
- Shoes.com launches two ad campaigns for a new running shoe.
- Wants to determine which campaign drives more sales.
- That is what conversion attribution measurement does.
- Typically, advertisers delegate these measurements to ad-techs.

Campaign 1

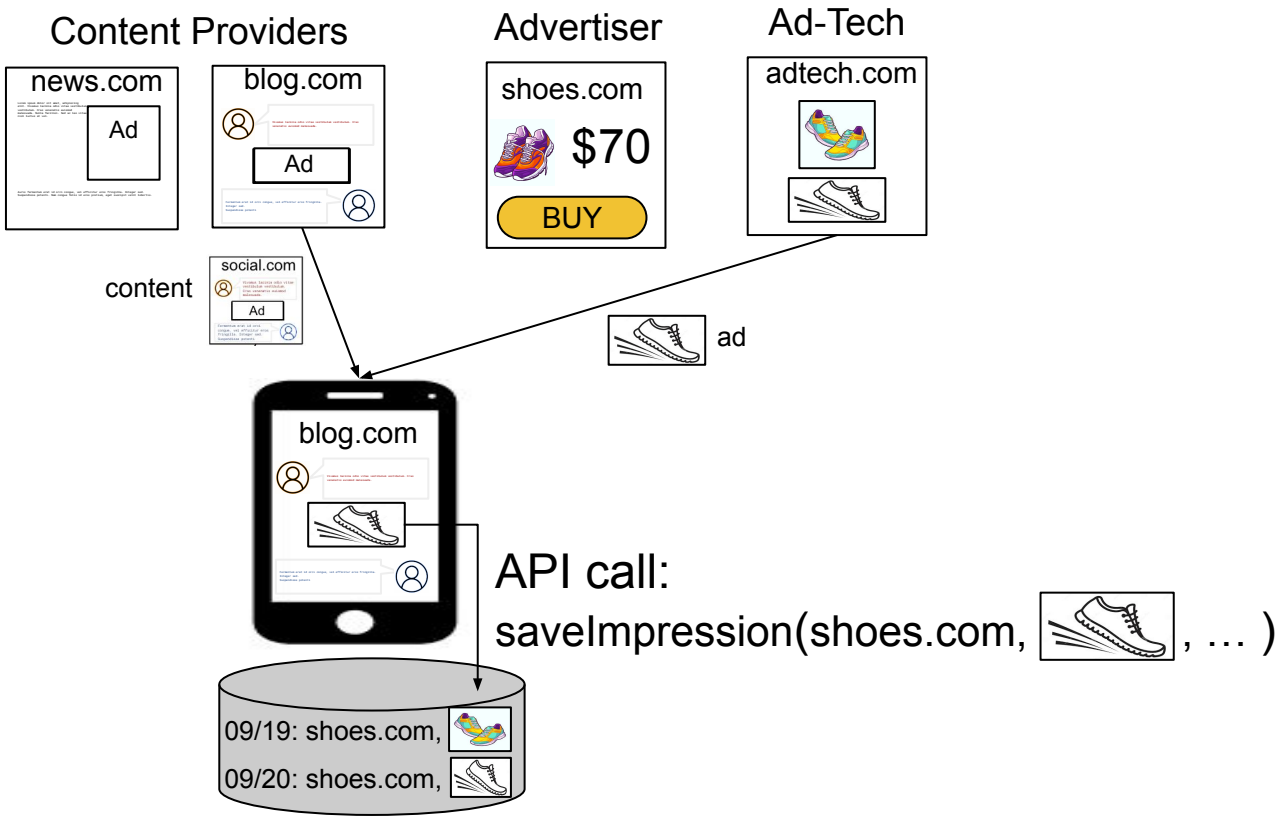
Campaign 2

How Browser APIs Work

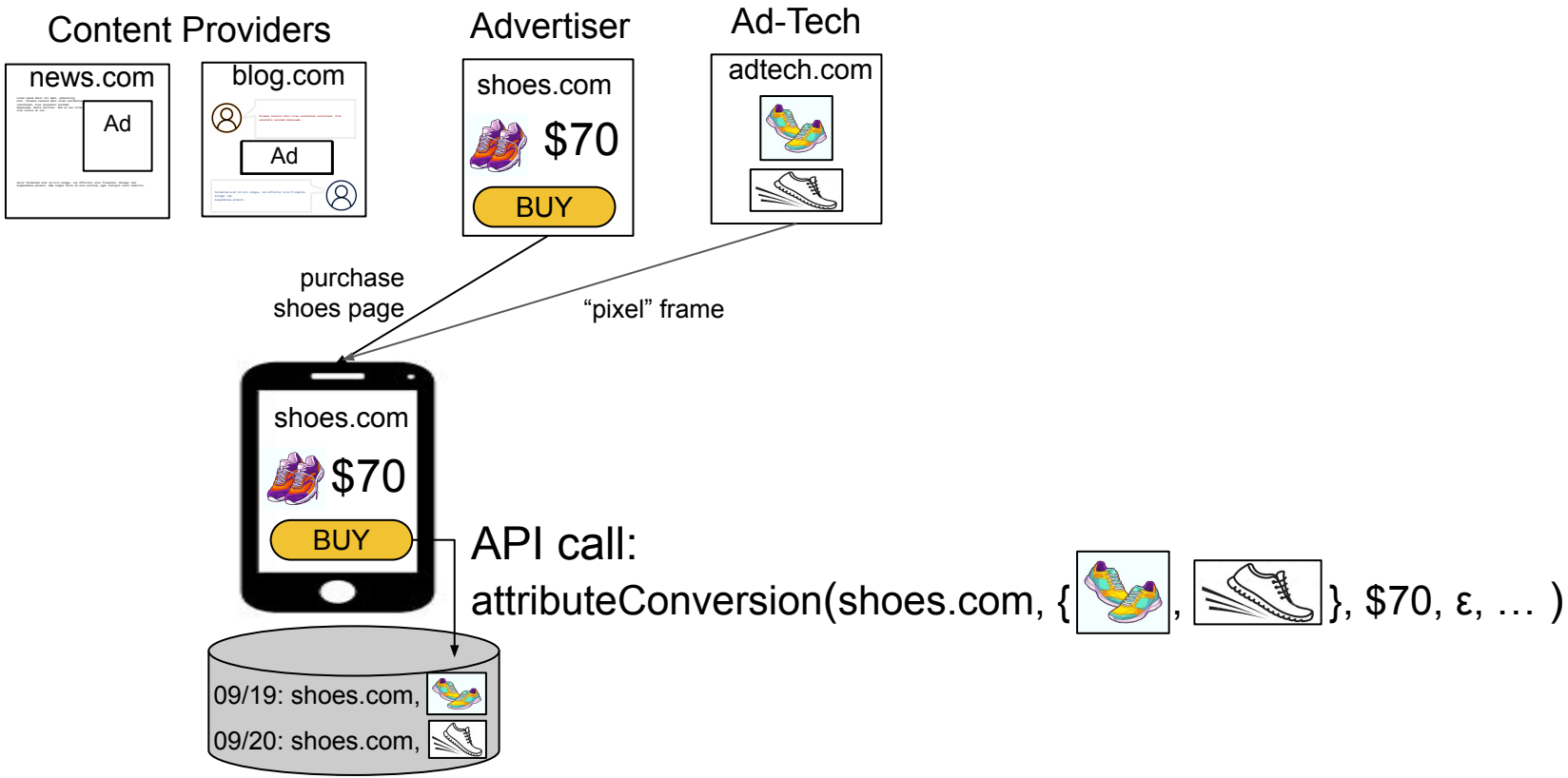
How Browser APIs Work



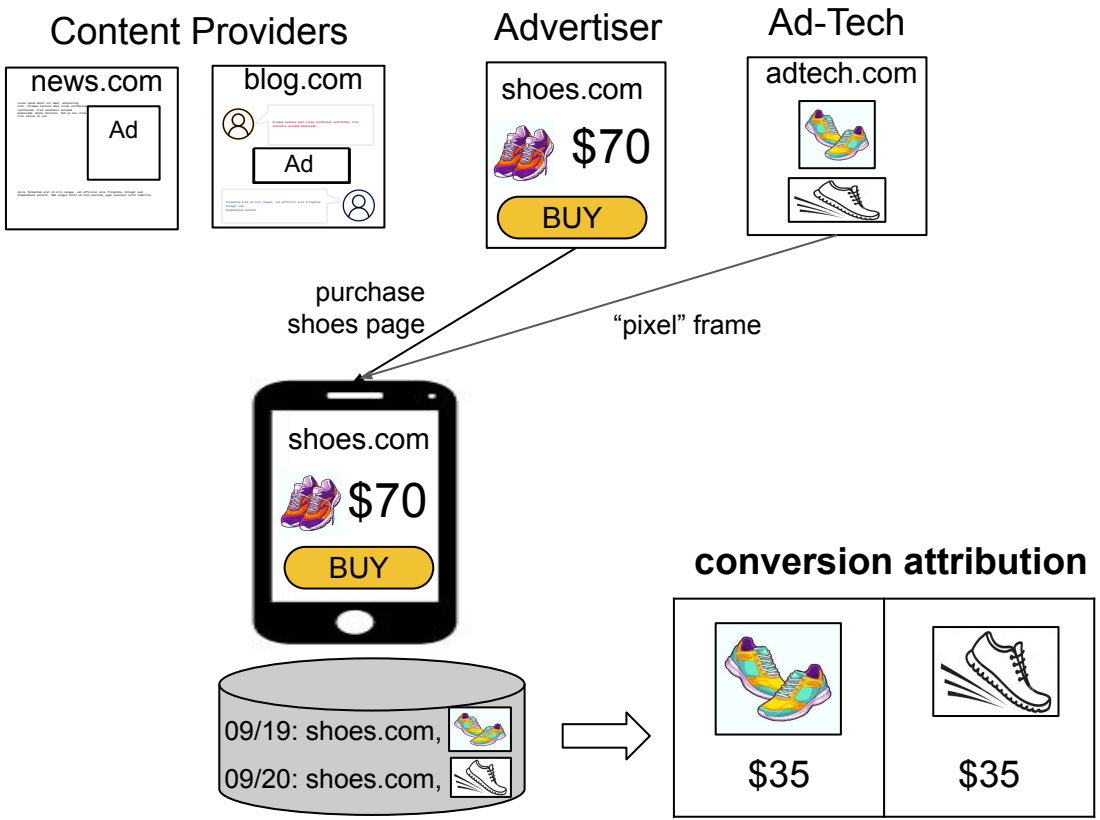
How Browser APIs Work



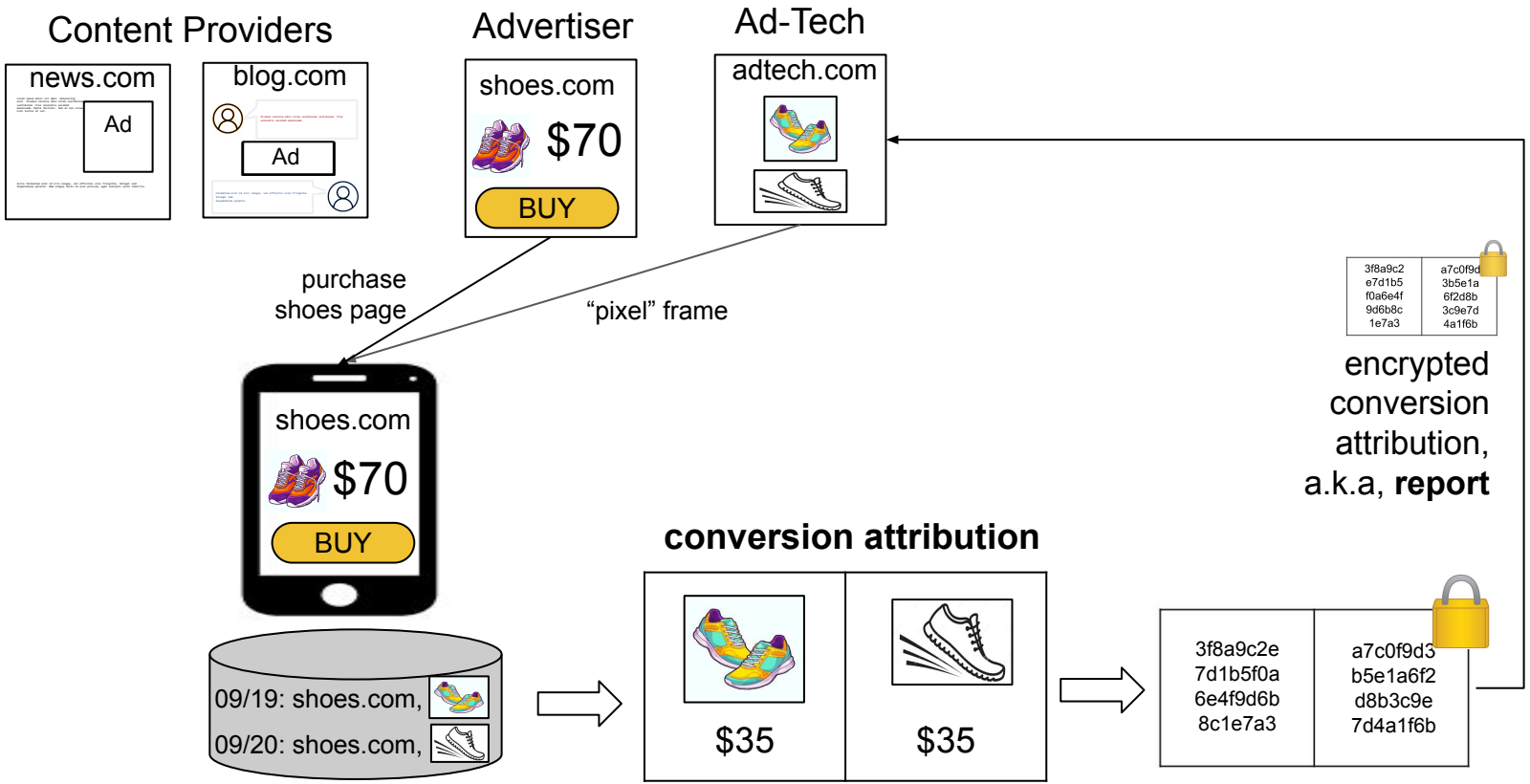
How Browser APIs Work



How Browser APIs Work

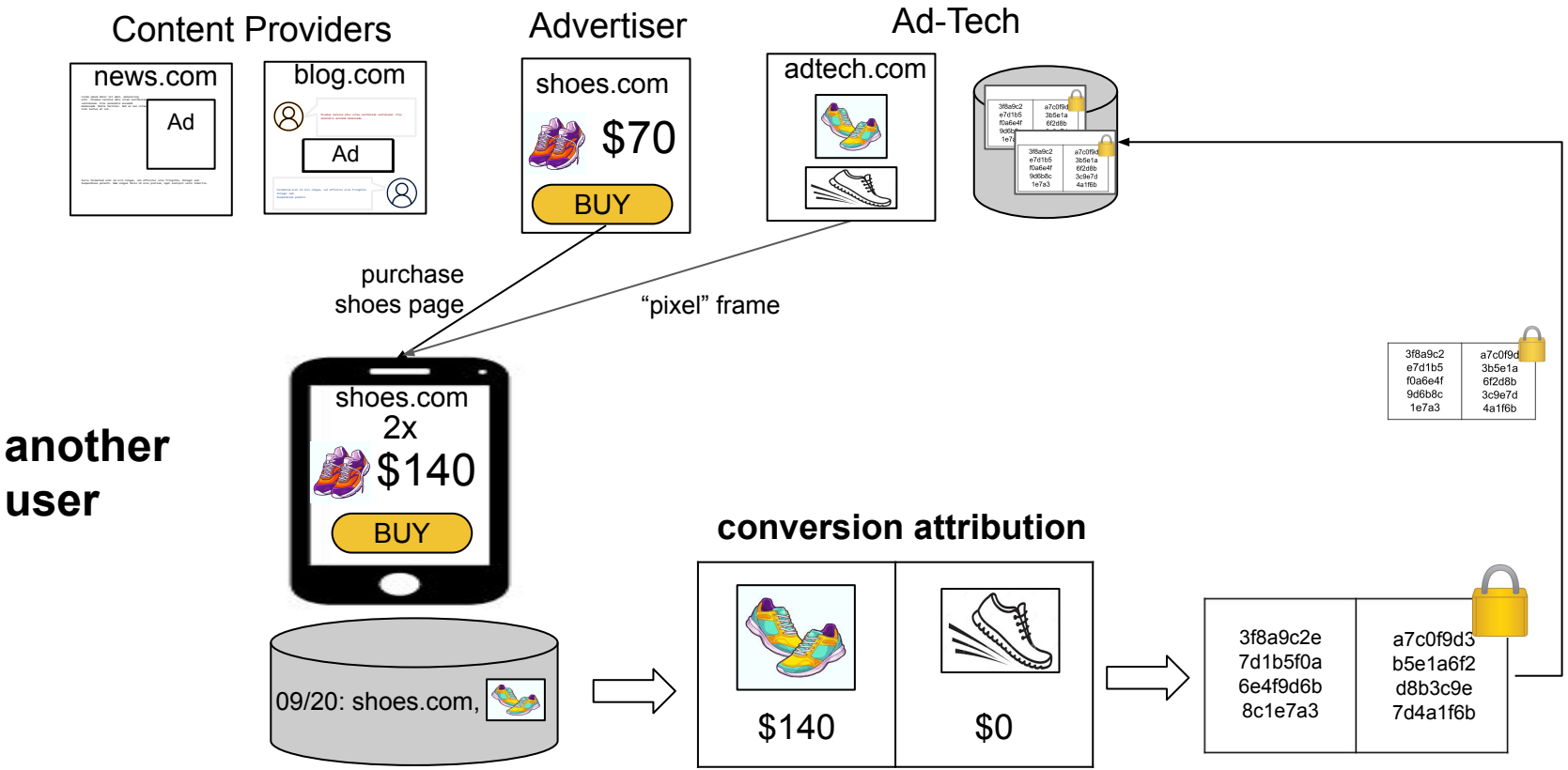


How Browser APIs Work

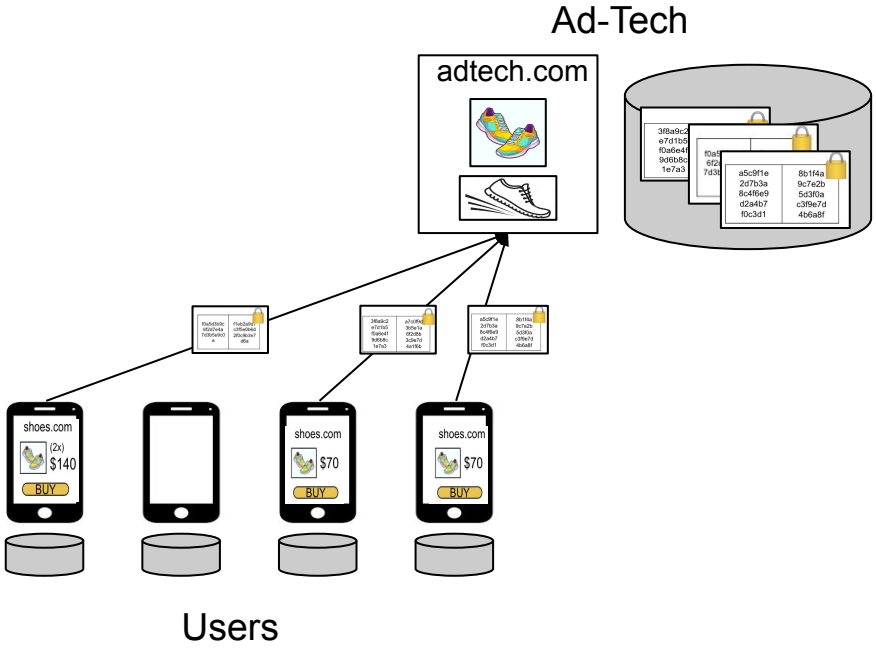


How Browser APIs Work

How Browser APIs Work



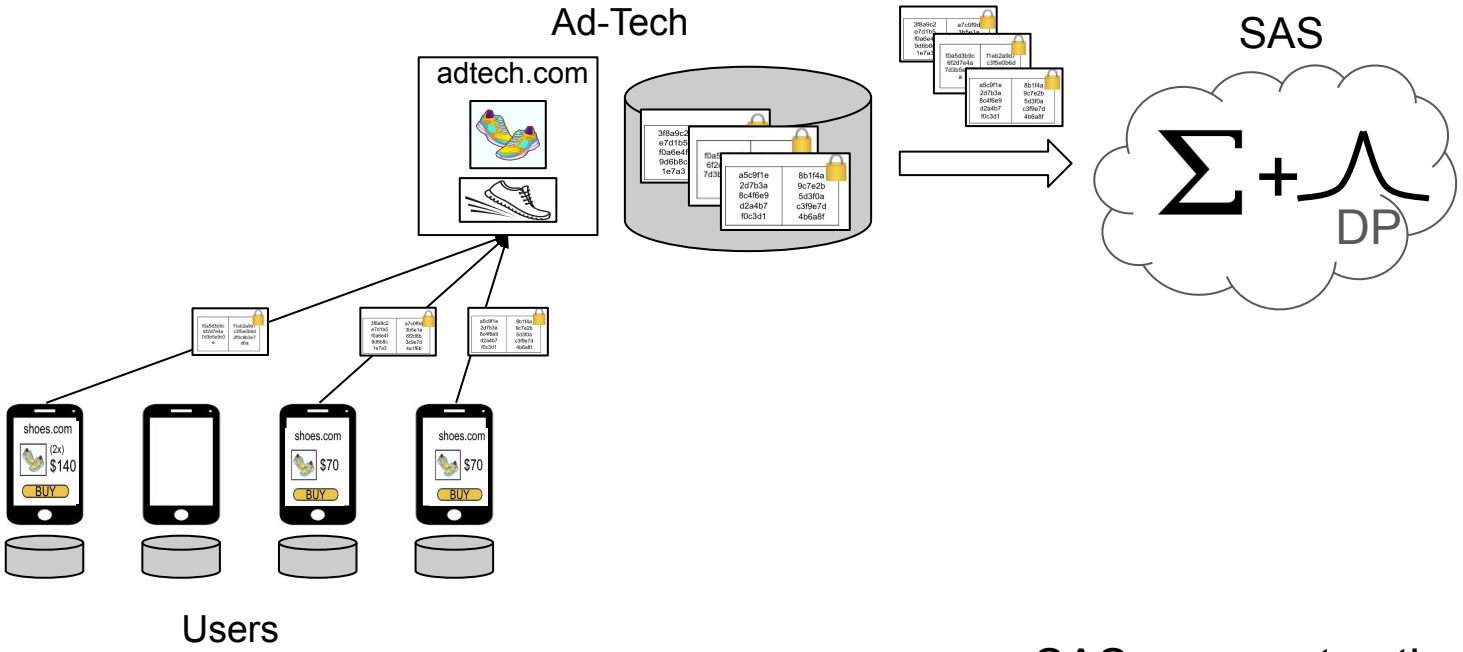
How Browser APIs Work



encrypted attributions sent only with conversions

ad tech can't decrypt them

How Browser APIs Work

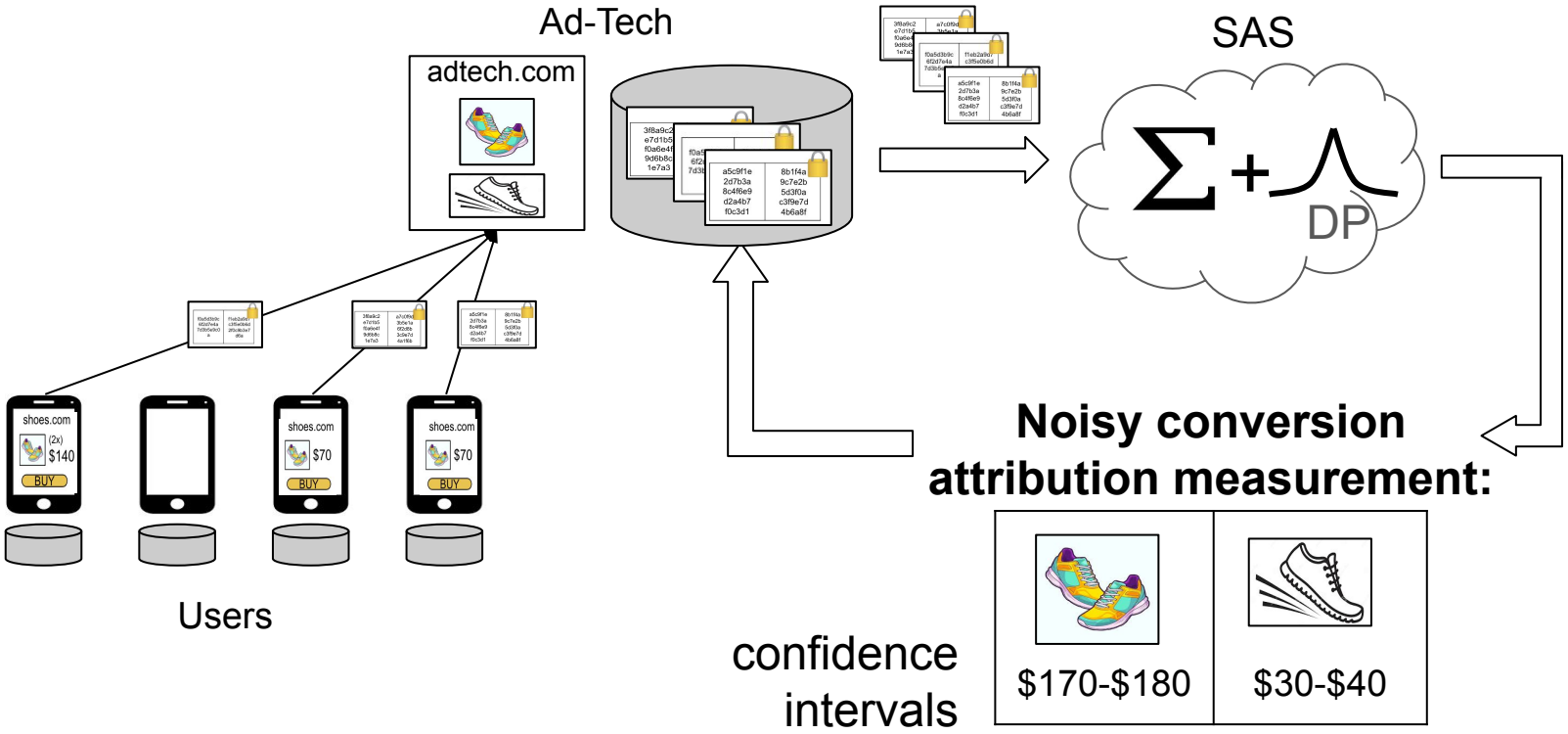


encrypted attributions sent only with conversions

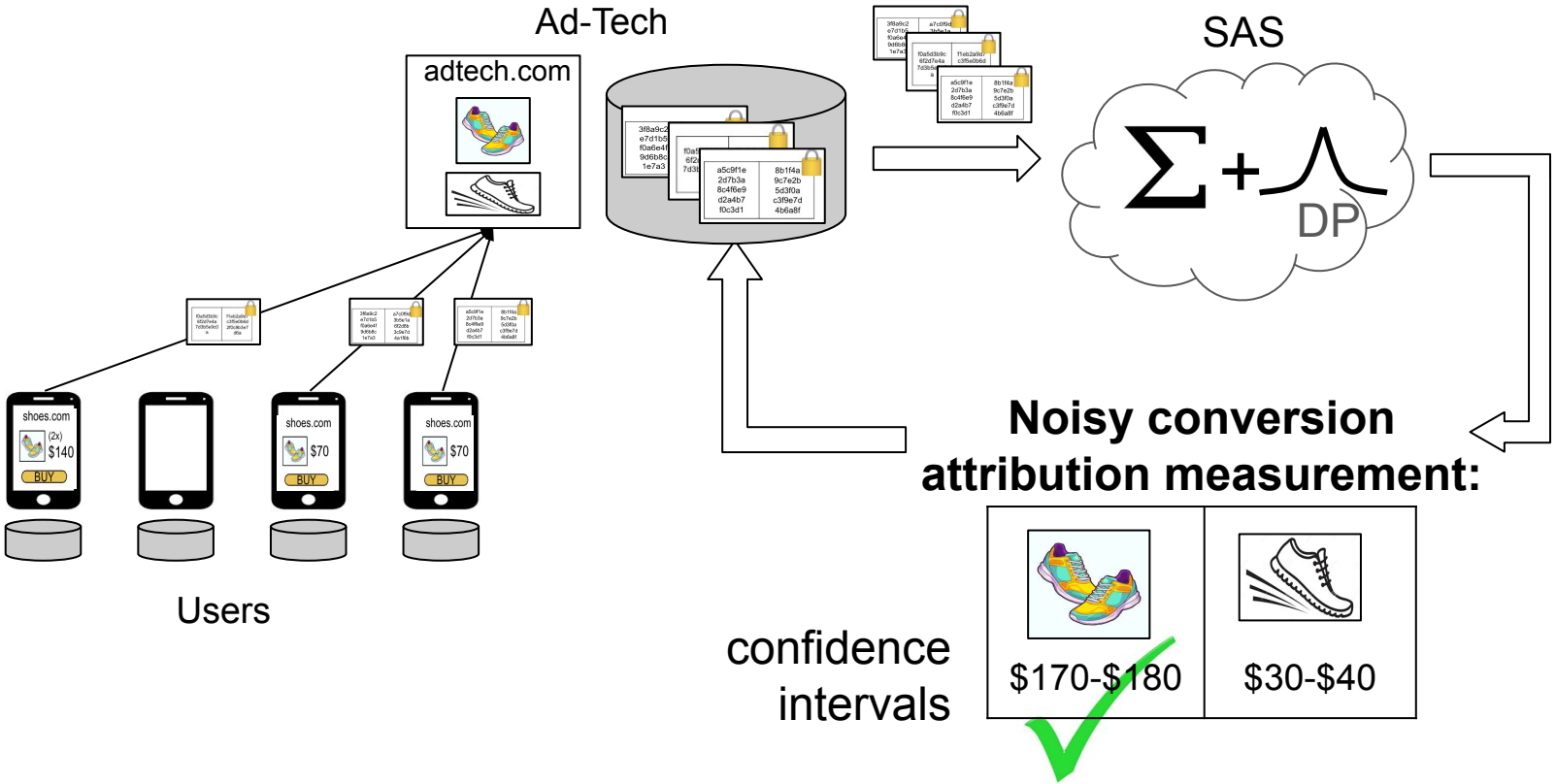
ad tech can't decrypt them

SAS aggregates them in a trusted TEE/MPC and adds noise for DP

How Browser APIs Work

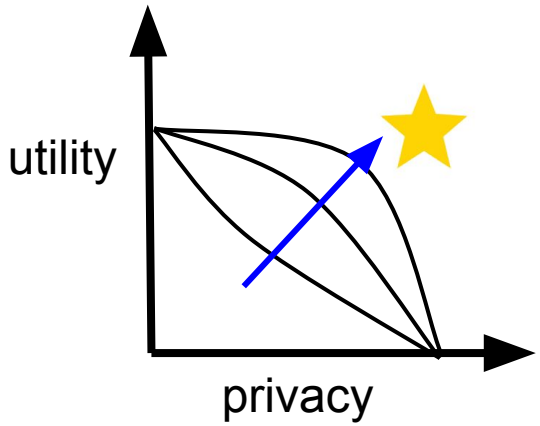


How Browser APIs Work



Challenge: Privacy-Utility Tradeoff

- Differential privacy (DP) adds noise to limit what ad-techs can learn about individual users, known as **privacy loss**.
- More noise reduces privacy loss but lowers accuracy (a.k.a. utility).
- Privacy loss accumulates across measurements, so DP systems track it and bound it.
- This is the privacy-utility tradeoff in DP, and the goal is to push the tradeoff curve toward the high-privacy/high-utility point. ★

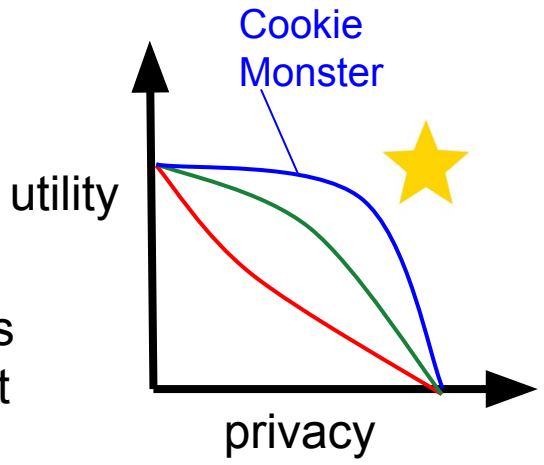


Outline

- Background on ad measurements and emerging APIs
- **Our privacy framework: Cookie Monster**
- Discussion on broader applications and bias mitigation

Cookie Monster

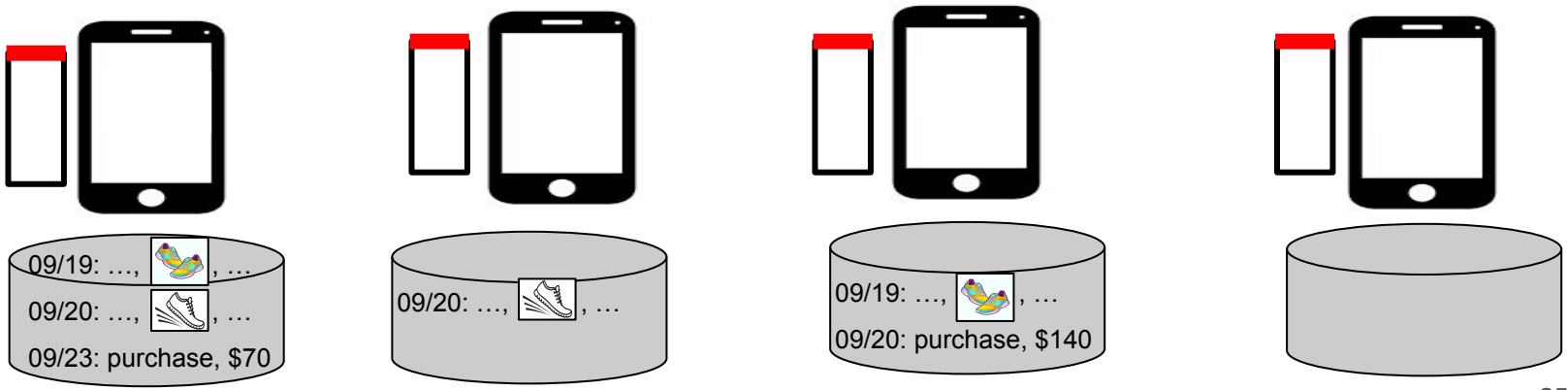
- Theoretically-justified, privacy loss accounting module for ARA/PAM/Hybrid, that enhances privacy-utility tradeoffs and offers added benefits like user transparency and control.
- Based on **Individual Differential Privacy (IDP)** [POPL15],[1] Cookie Monster lets browsers track their own privacy guarantees and account for privacy loss based on their contribution (or lack thereof) to each measurement.



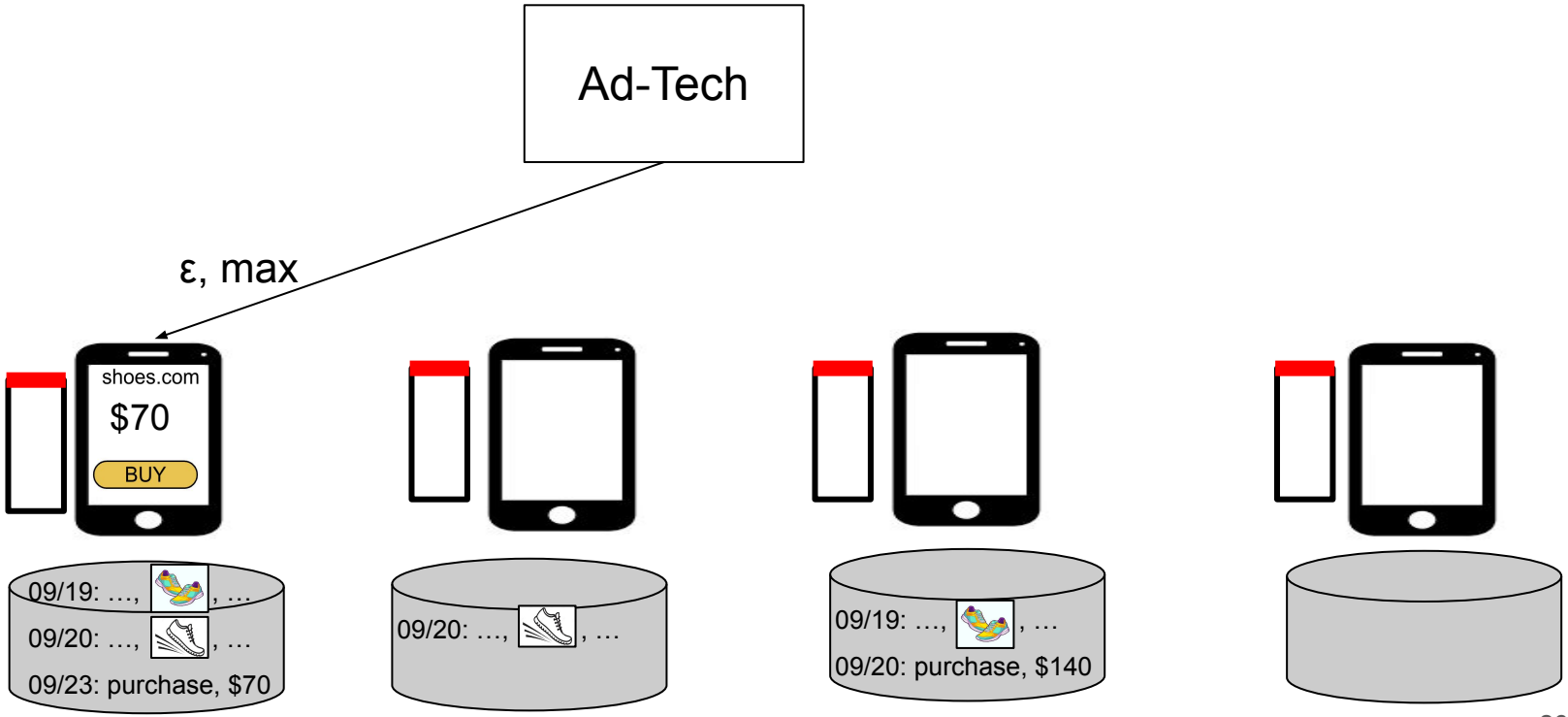
- Our peer-reviewed paper gives a detailed description and formal analysis of our design, and evaluates our Chrome ARA-based prototype.

How Cookie Monster Works

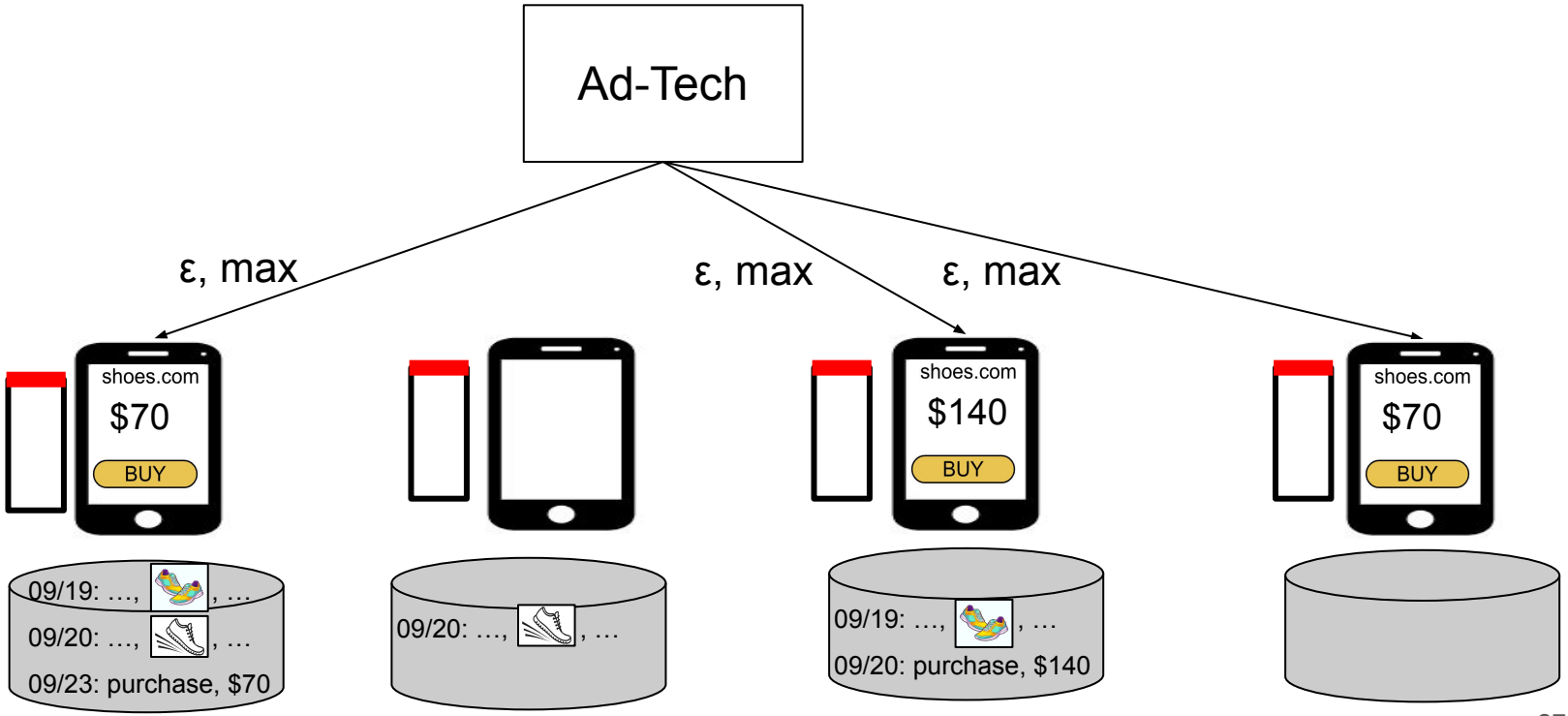
- Each device tracks its own privacy loss for each ad-tech and enforces a **cap** (a.k.a., **total privacy budget**).



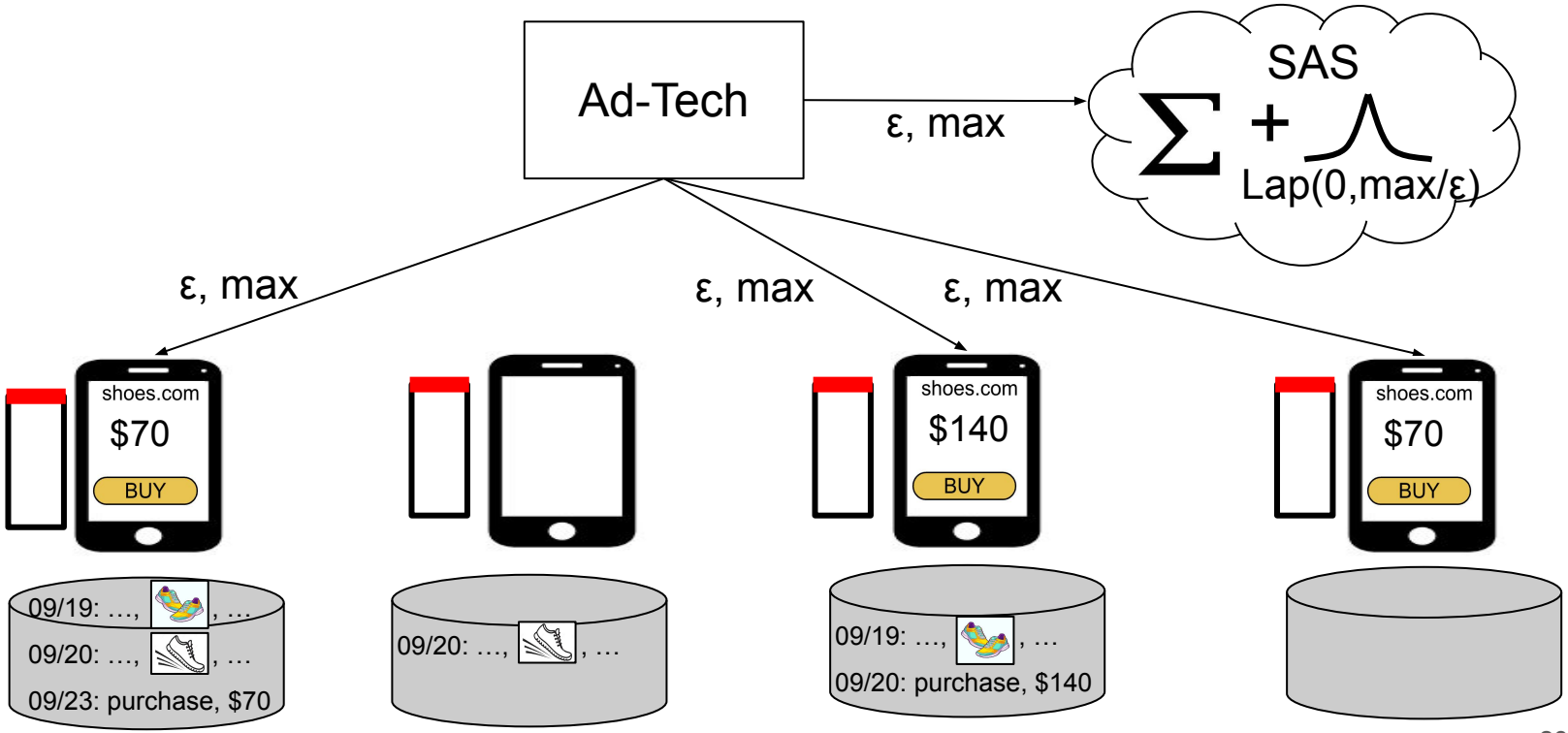
- Ad-tech specifies privacy loss (ε) and maximum conversion value (max), and the SAS calibrates DP noise accordingly.



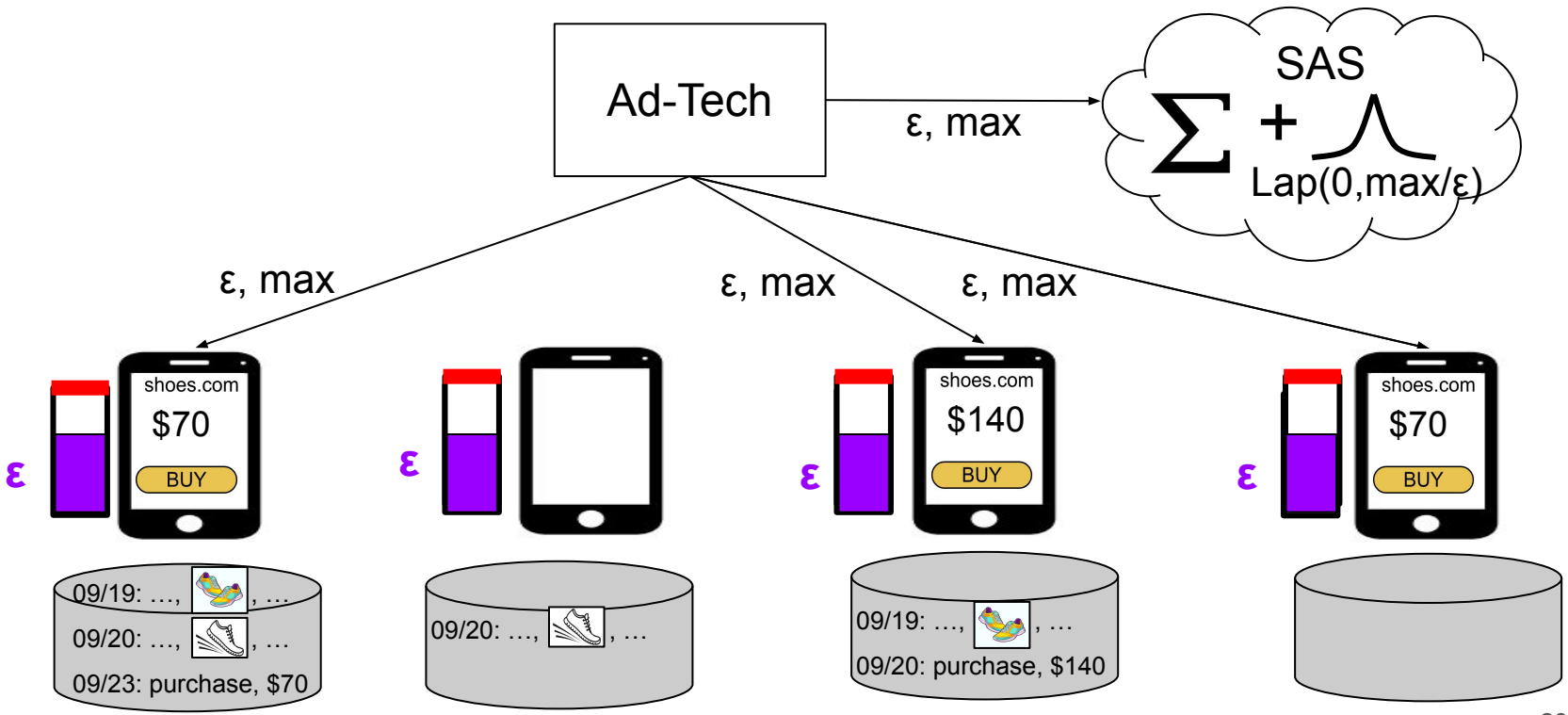
- Ad-tech specifies privacy loss (ε) and maximum conversion value (max), and the SAS calibrates DP noise accordingly.



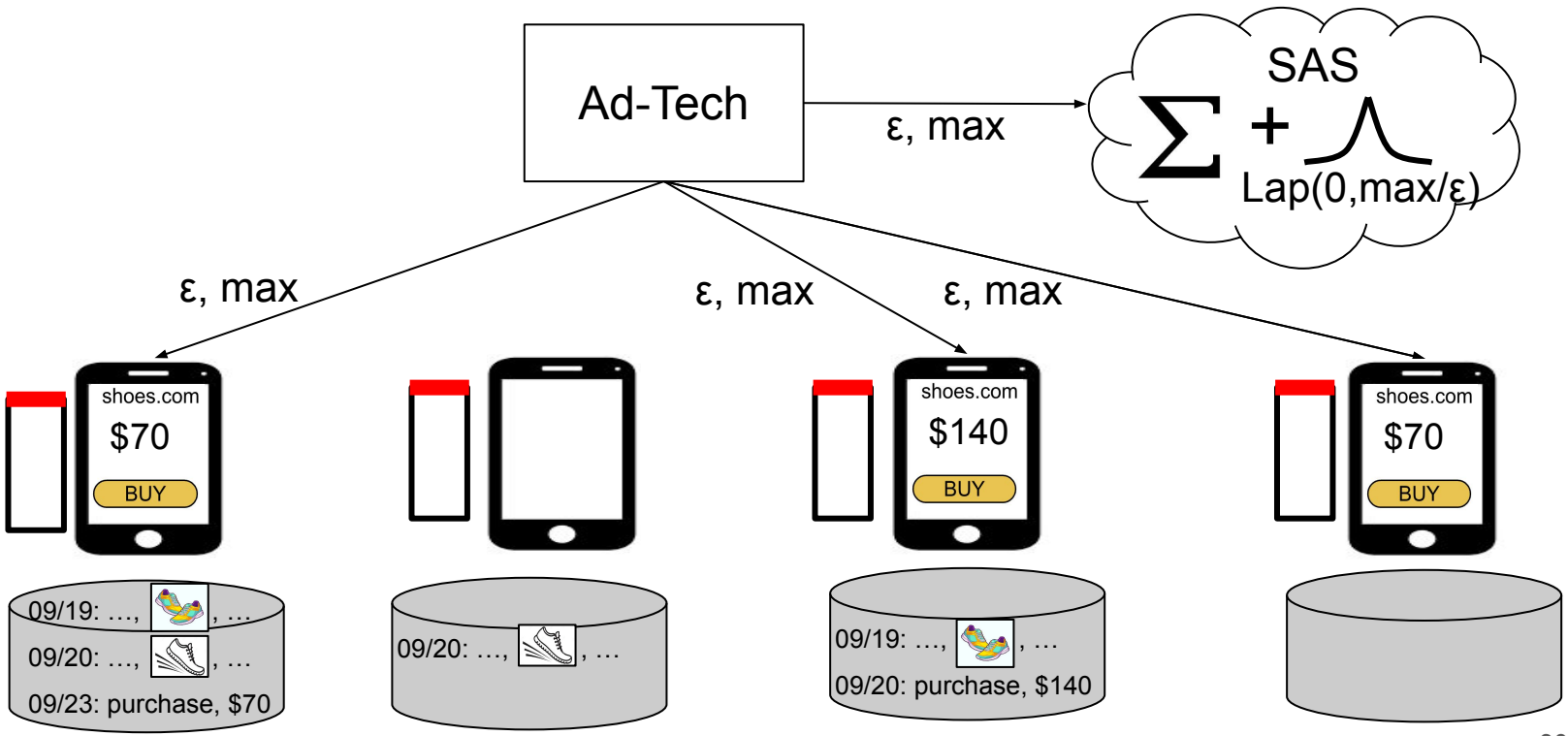
- Ad-tech specifies privacy loss (ε) and maximum conversion value (max), and the SAS calibrates DP noise accordingly.



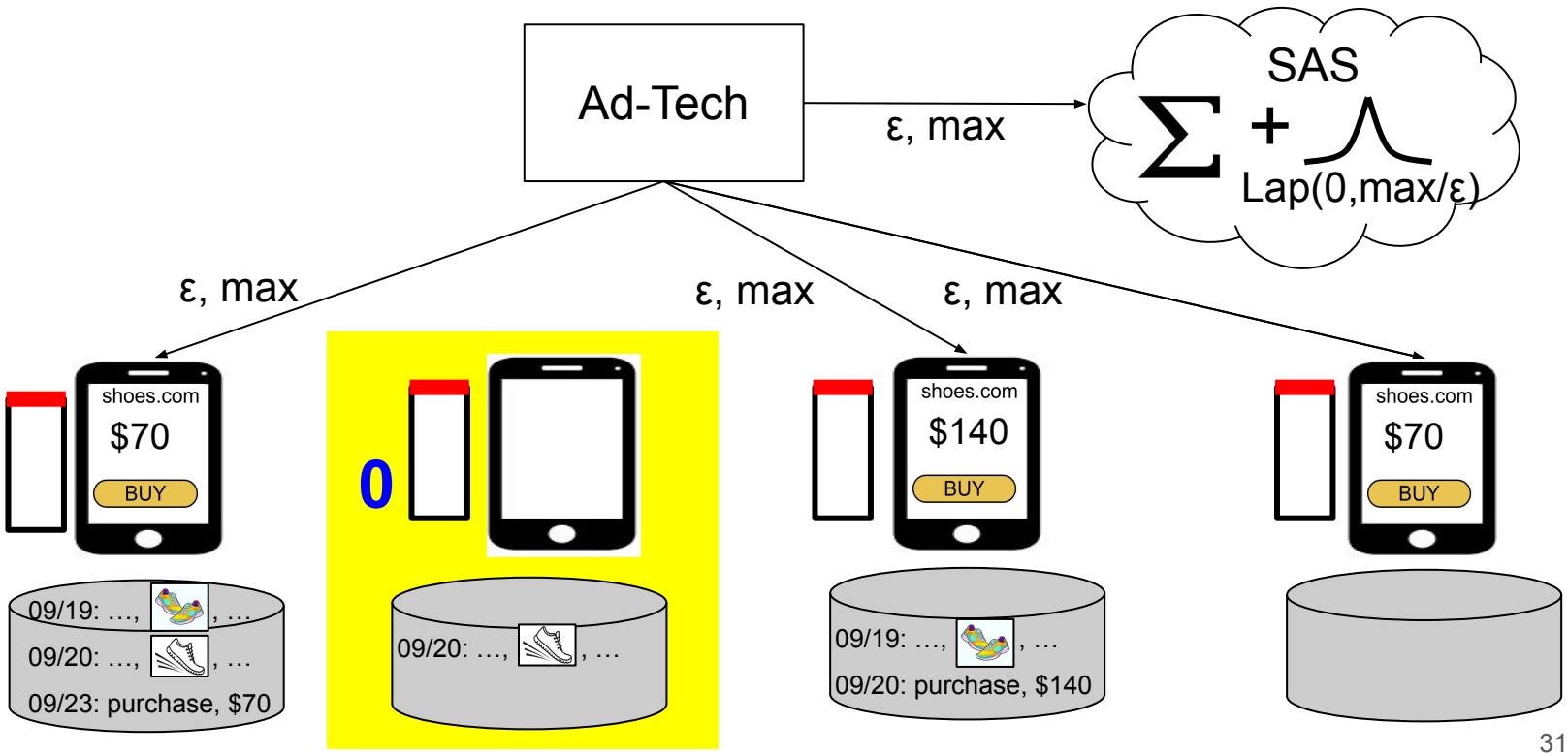
- **With traditional DP**, all devices would have to account for the same **global privacy loss ε**, including those not participating in the measurement.



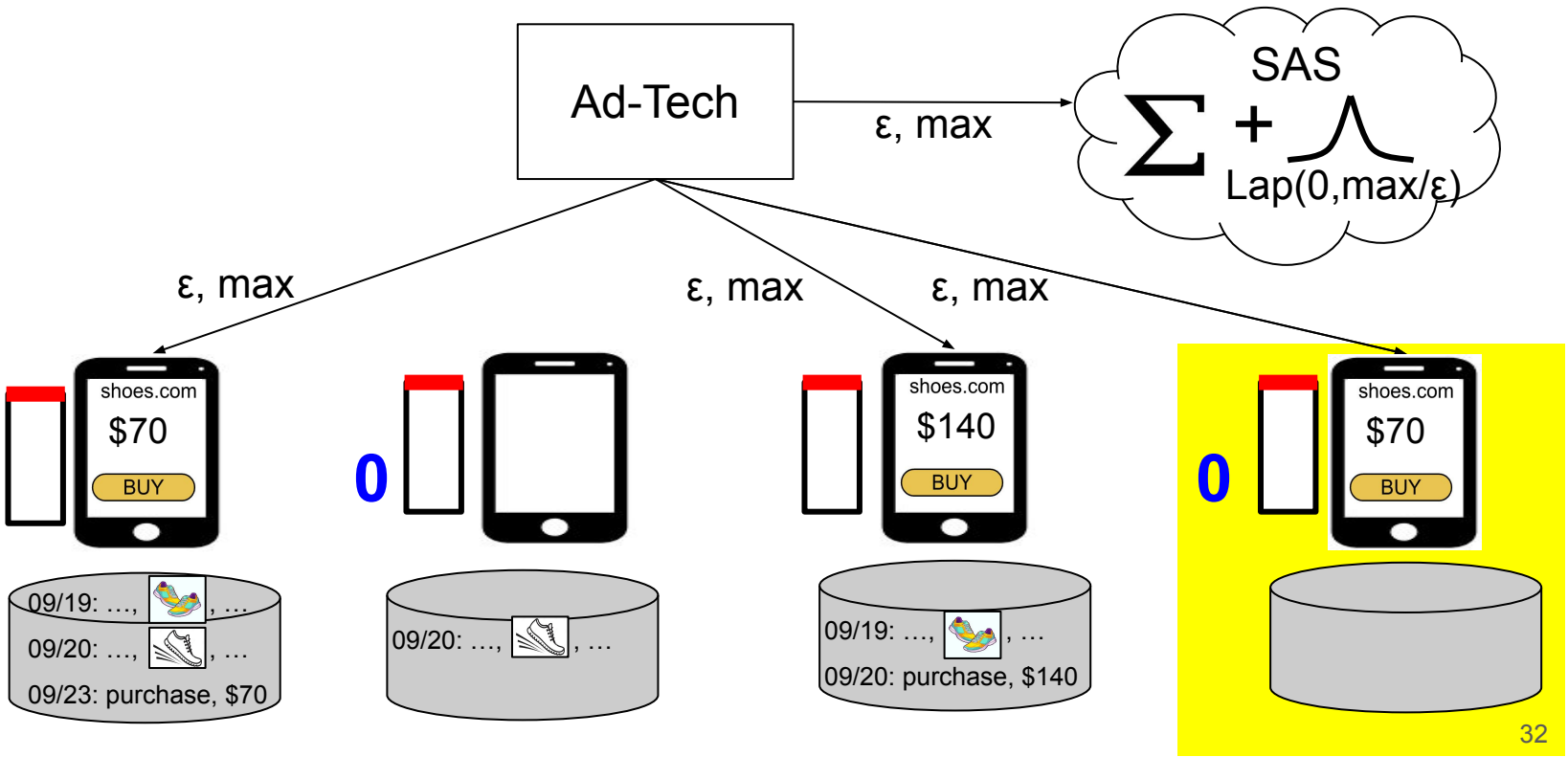
- In Cookie Monster, **with individual DP**, each device tracks its **individual privacy loss**, based on its own maximum contribution to the measurement.



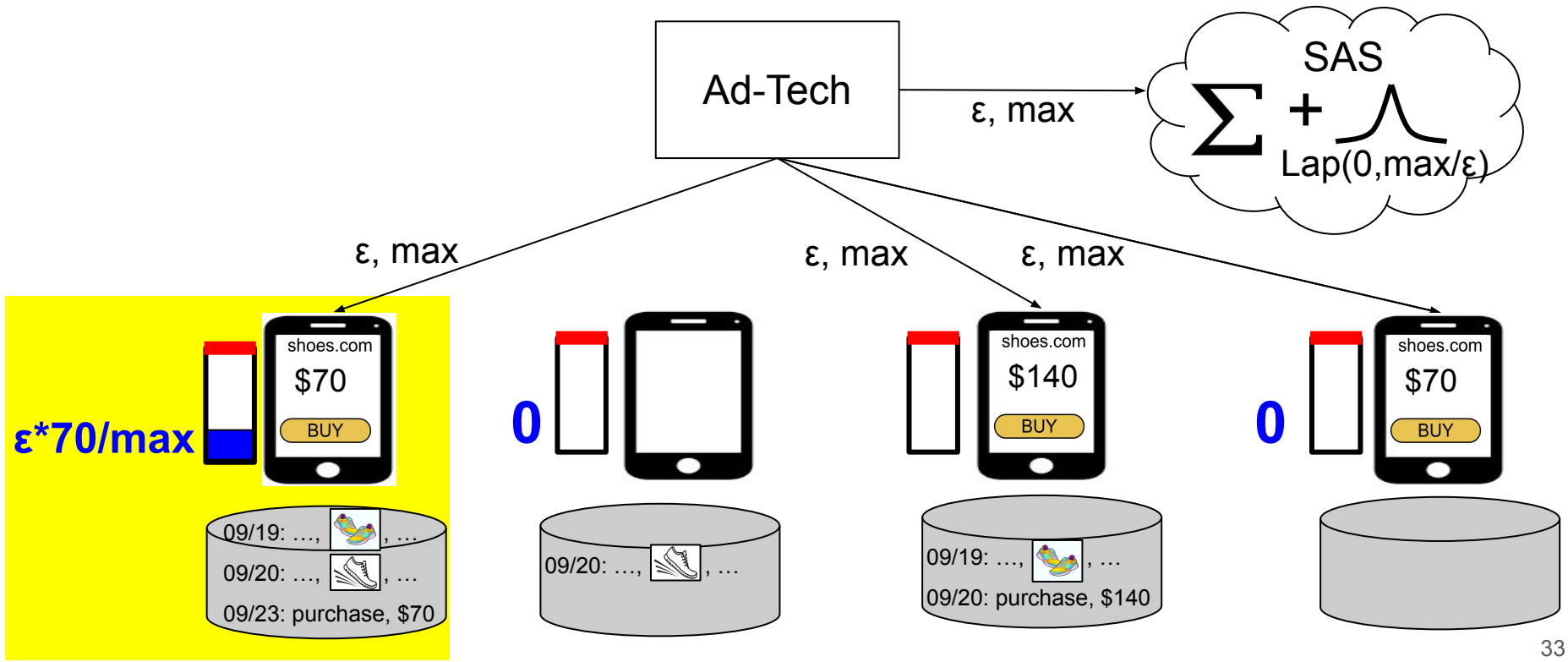
- In Cookie Monster, **with individual DP**, each device tracks its **individual privacy loss**, based on its own maximum contribution to the measurement.



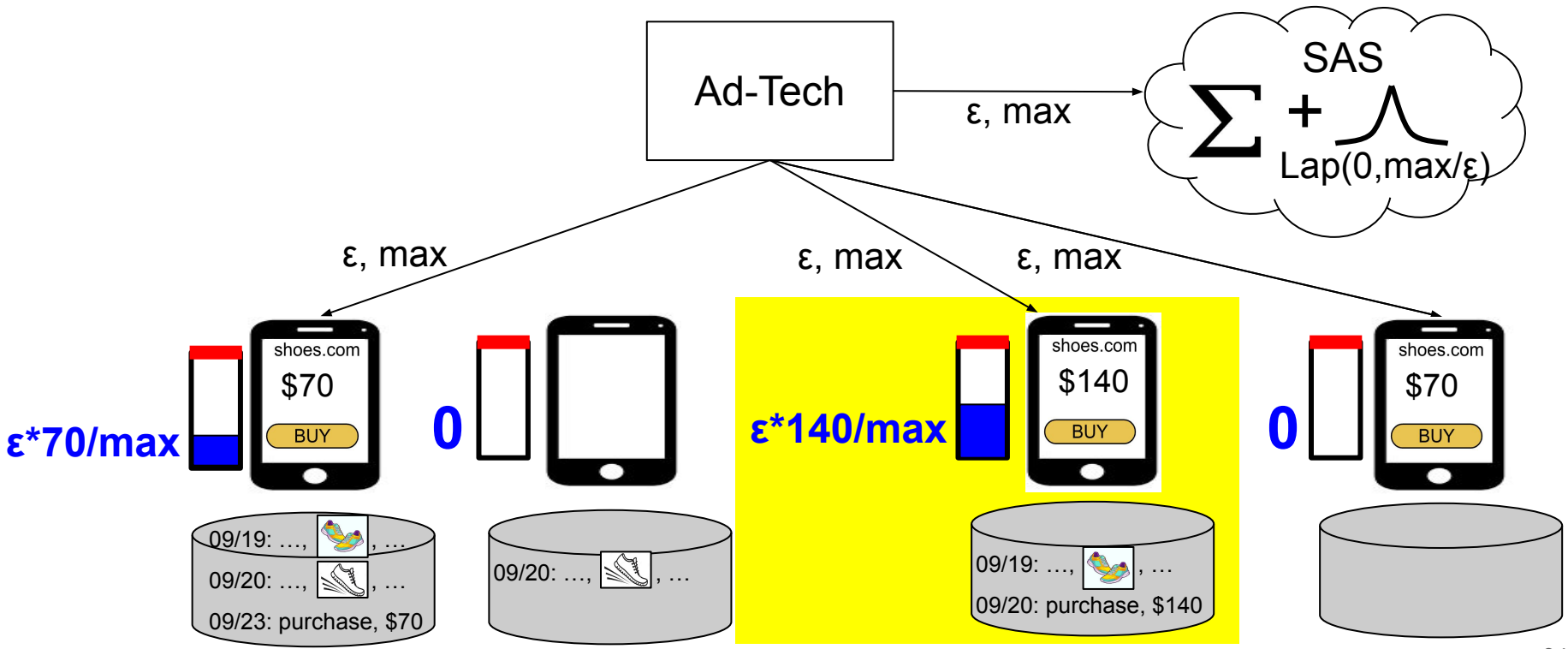
- In Cookie Monster, **with individual DP**, each device tracks its **individual privacy loss**, based on its own maximum contribution to the measurement.



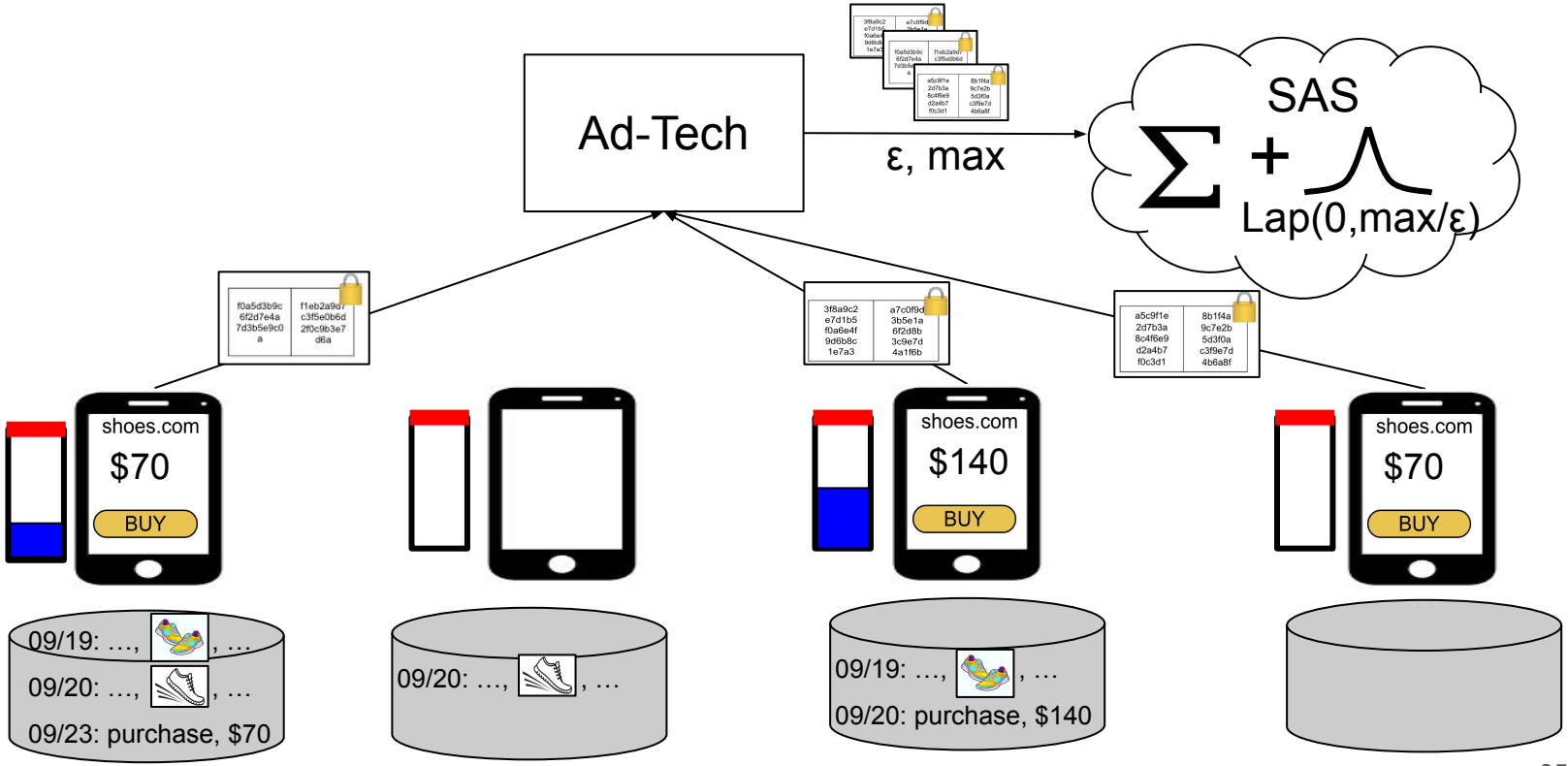
- In Cookie Monster, **with individual DP**, each device tracks its **individual privacy loss**, based on its own maximum contribution to the measurement.



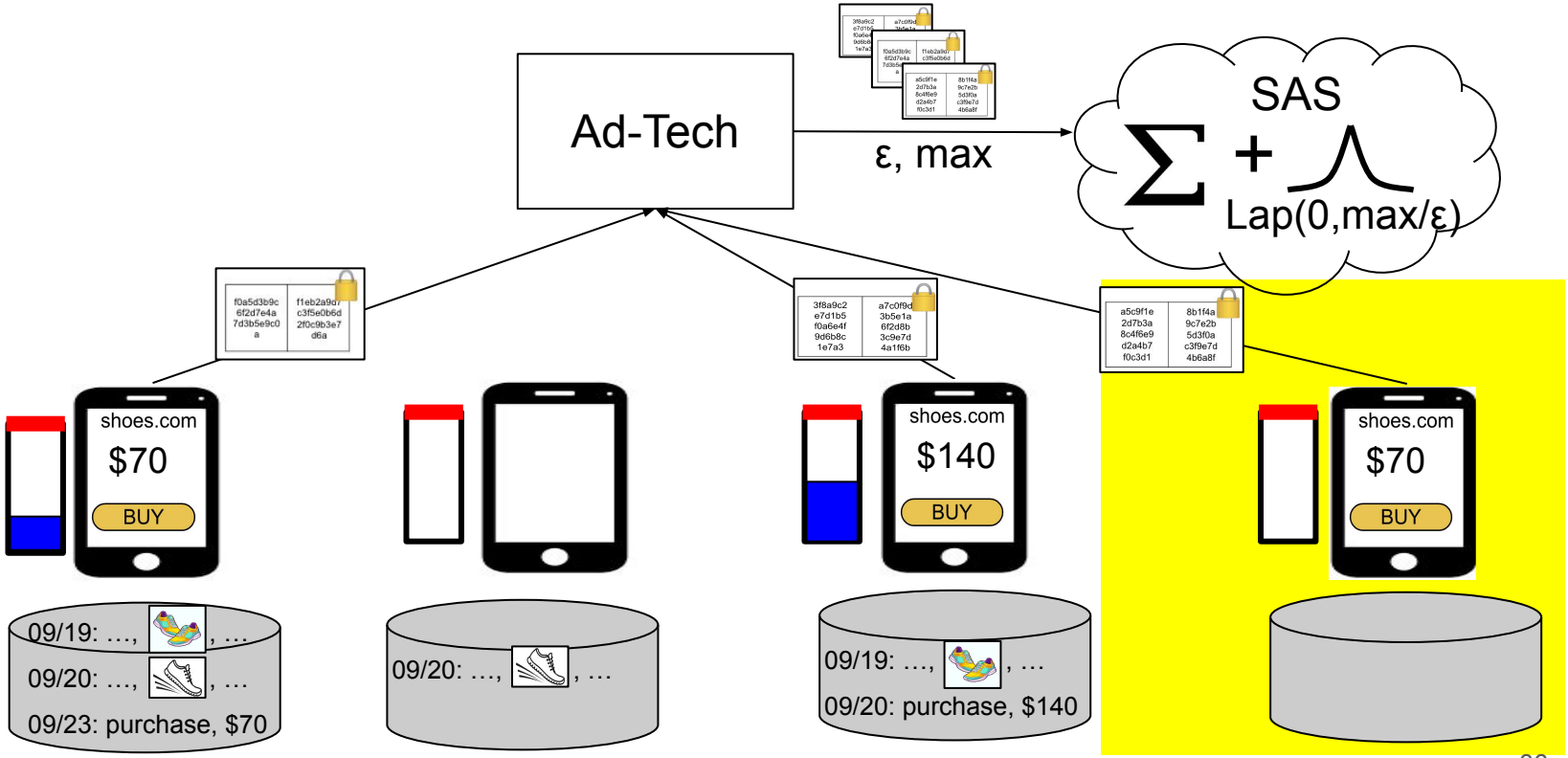
- In Cookie Monster, **with individual DP**, each device tracks its **individual privacy loss**, based on its own maximum contribution to the measurement.



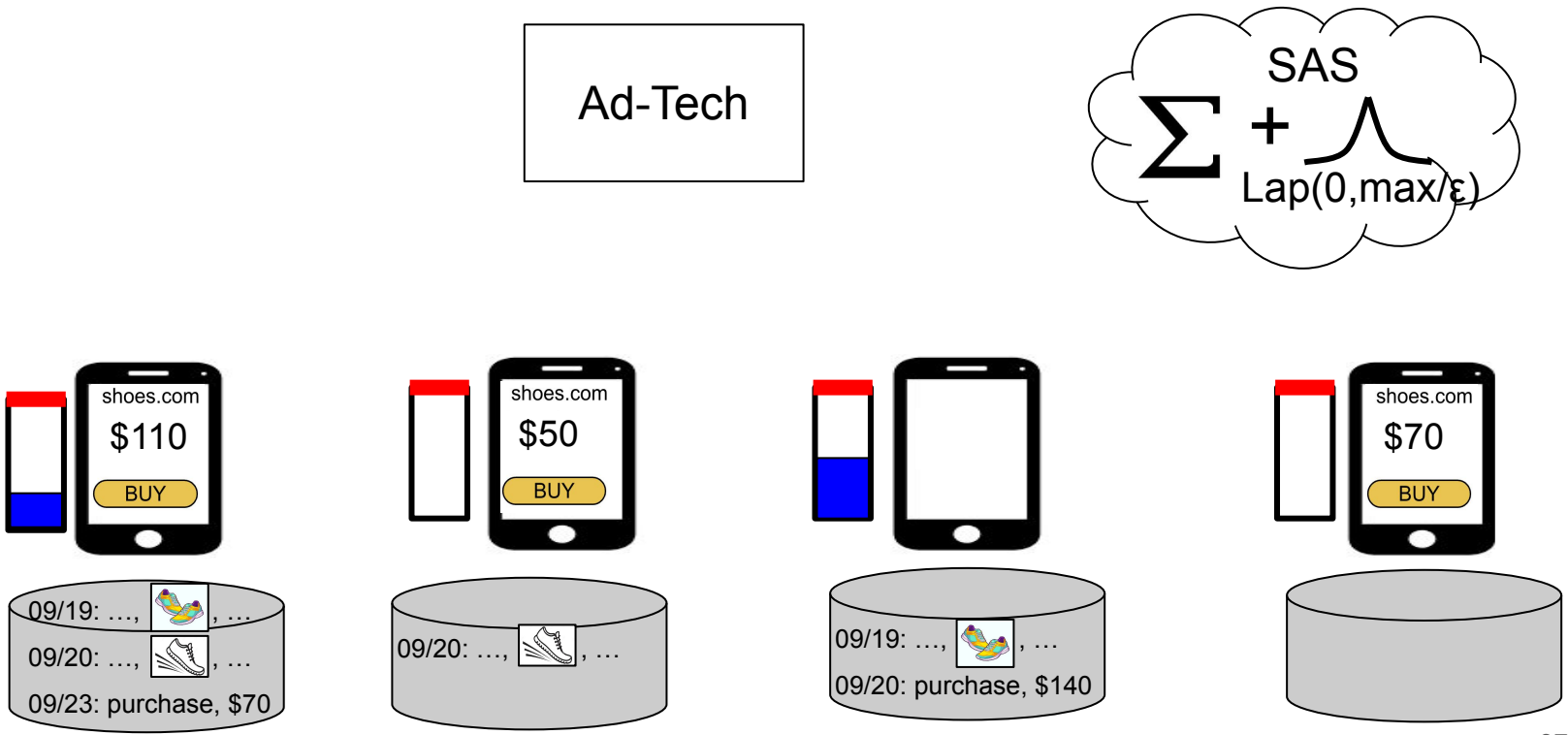
- Device includes ε, max in each report as authenticated data.
- SAS ensures ε, max consistency across reports, and that each report is used only once.



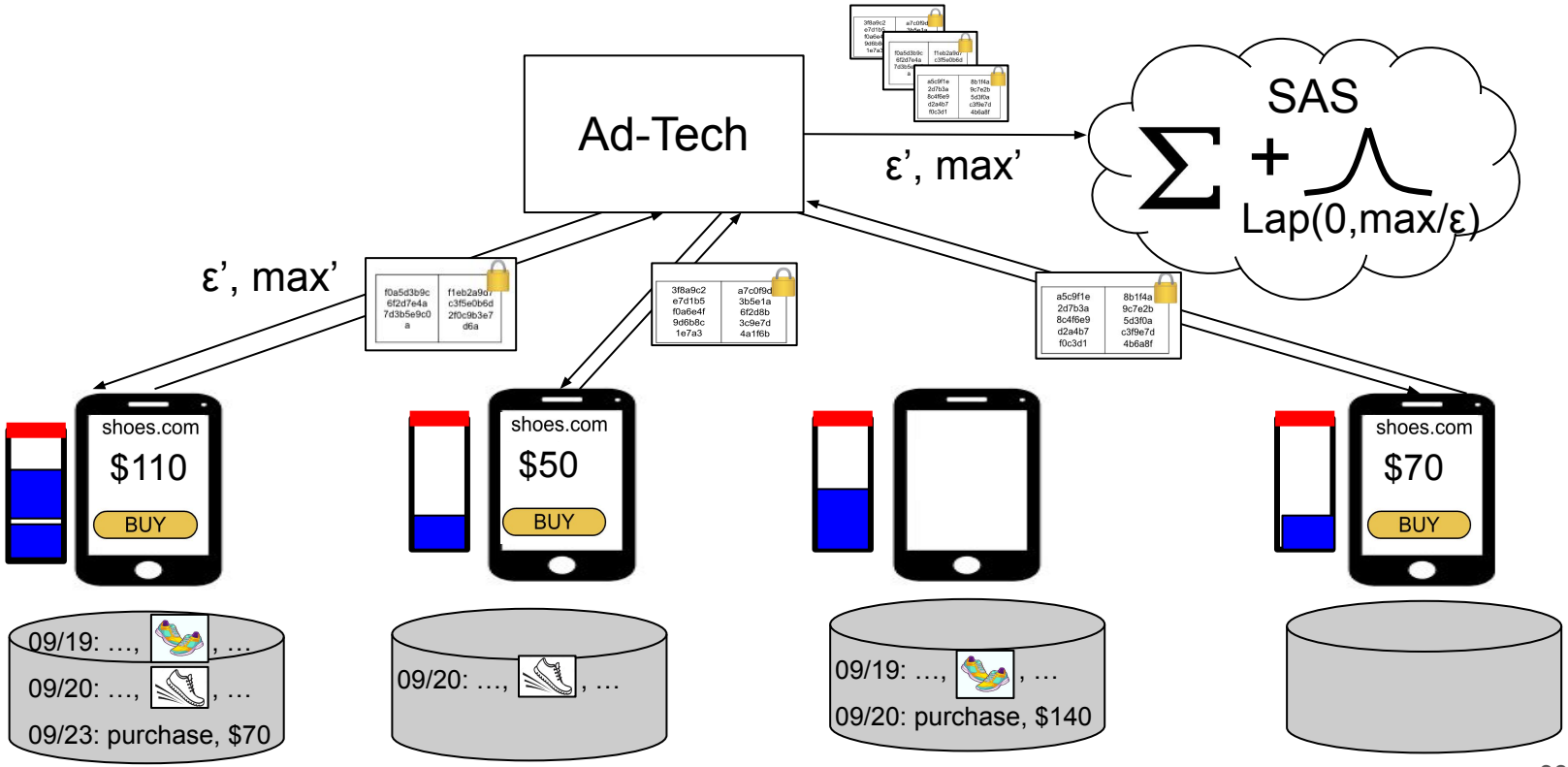
- Device includes ε, max in each report as authenticated data.
- SAS ensures ε, max consistency across reports, and that each report is used only once.



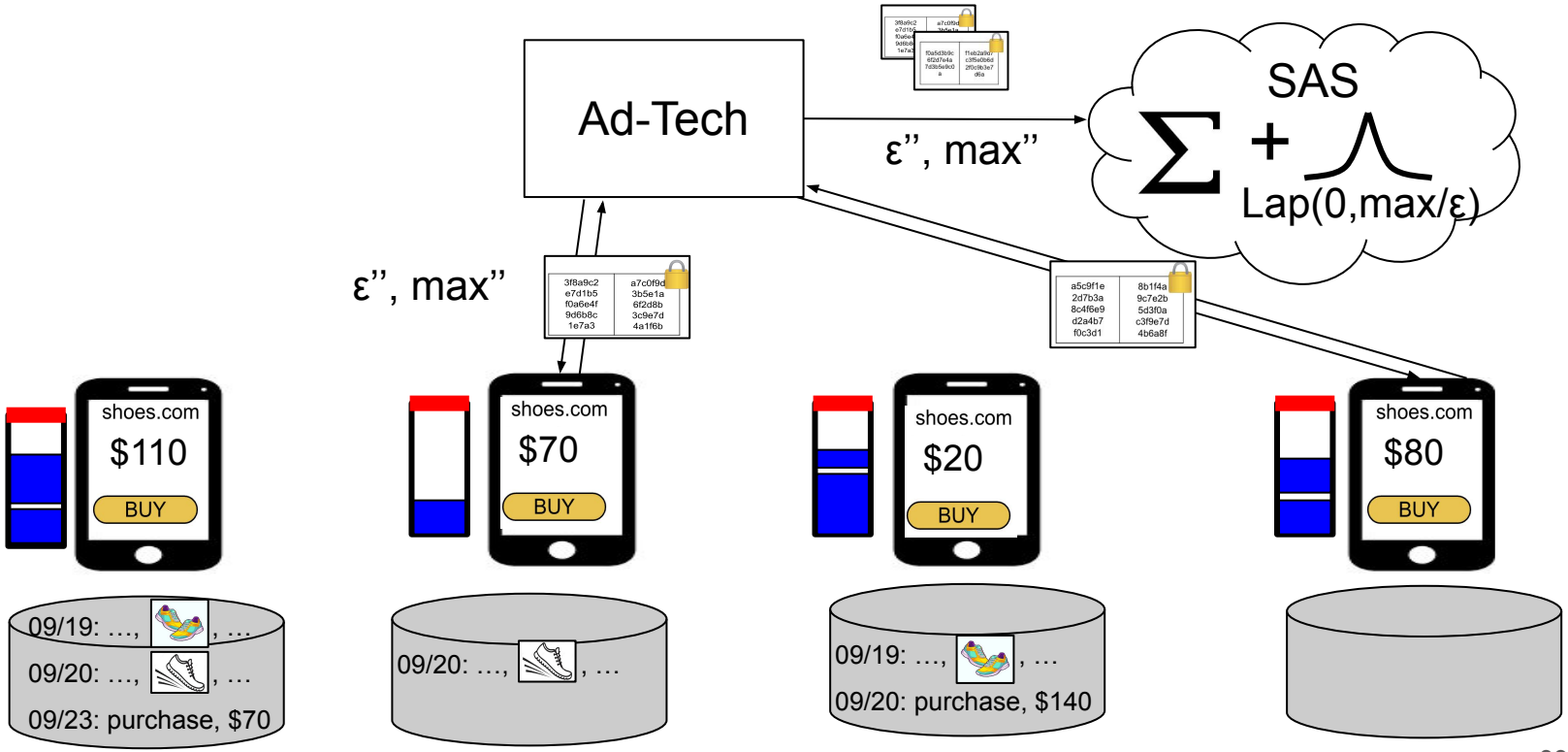
- This tighter individual privacy loss accounting lets the ad-tech execute **many more measurements** compared with traditional DP accounting.



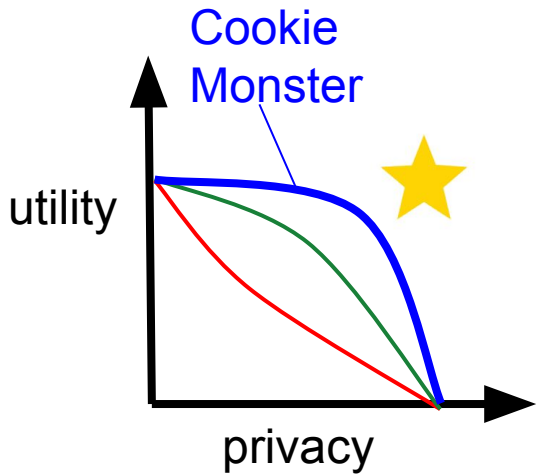
- This tighter individual privacy loss accounting lets the ad-tech execute **many more measurements** compared with traditional DP accounting.



- This tighter individual privacy loss accounting lets the advertiser execute **many more measurements** compared with traditional DP accounting.

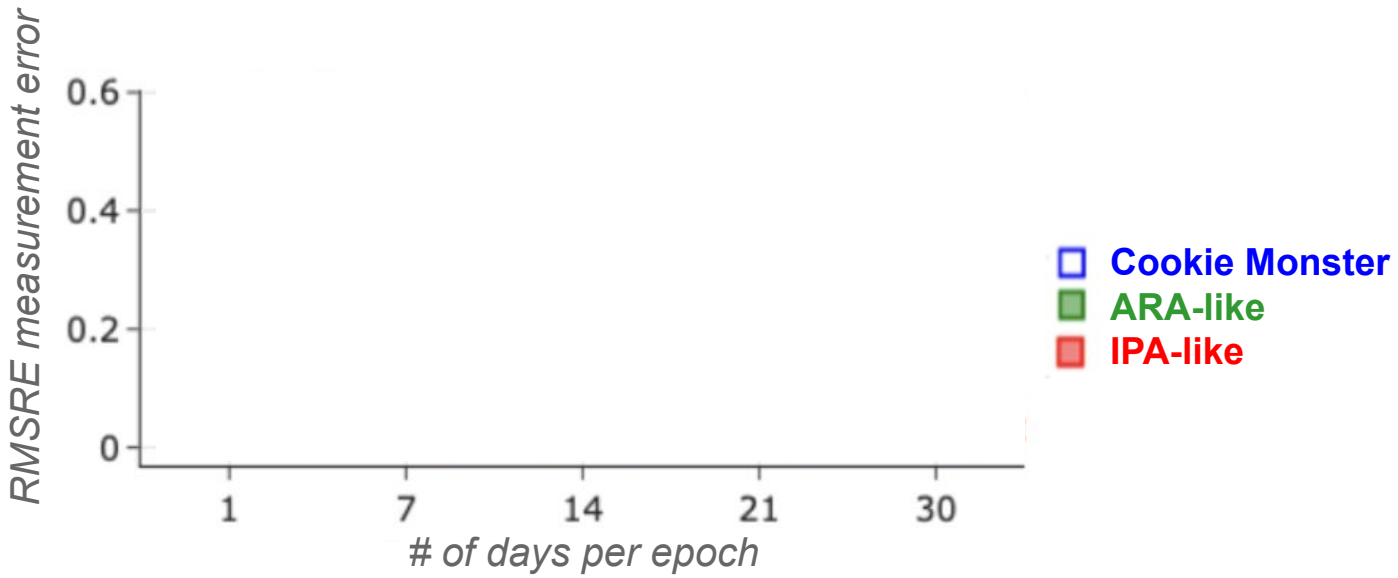


Cookie Monster Improves Privacy-Utility Tradeoff



Evaluation on PATCG Dataset

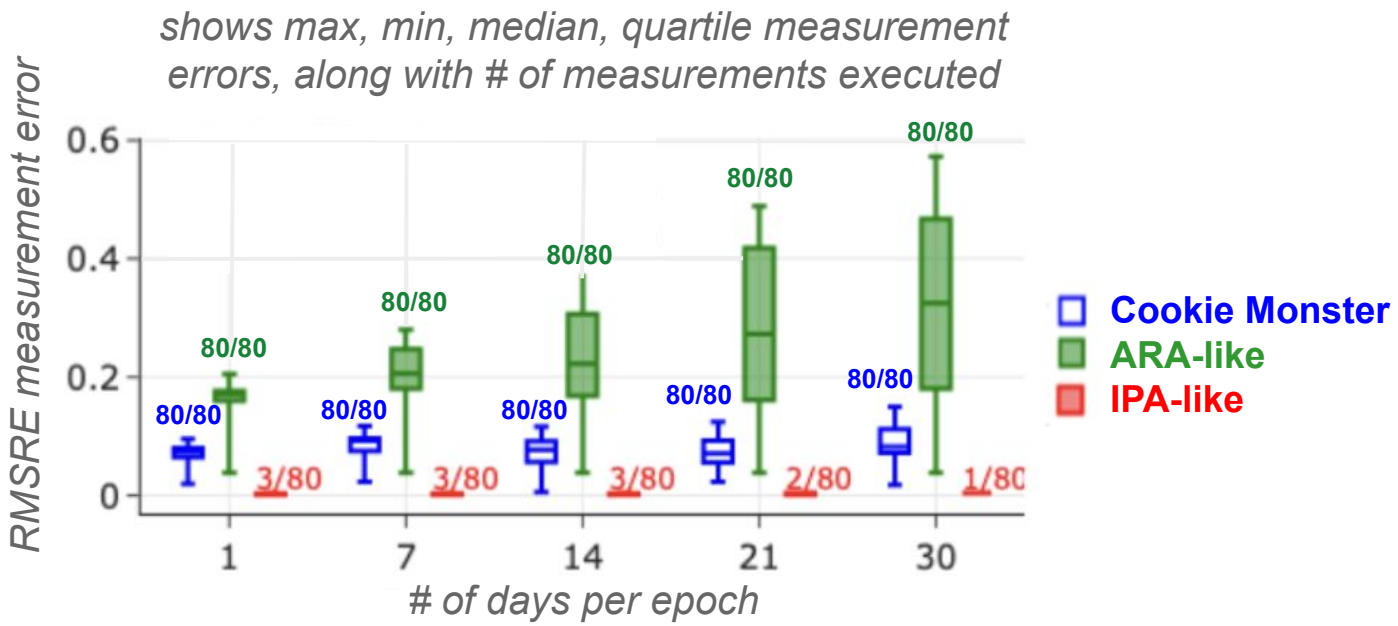
Utility metric
(lower is better at high numbers of measurements executed)



Privacy metric
(higher is better)

Evaluation on PATCG Dataset

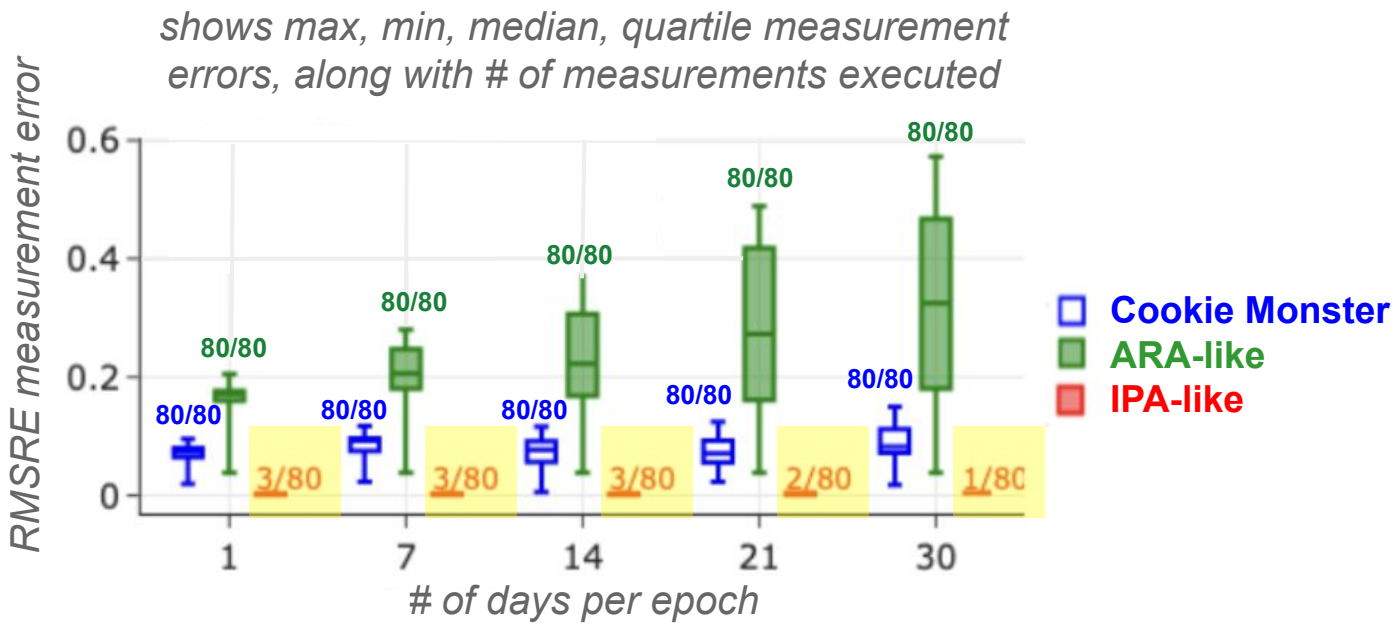
Utility metric
(lower is better at high numbers of measurements executed)



Privacy metric
(higher is better)

Evaluation on PATCG Dataset

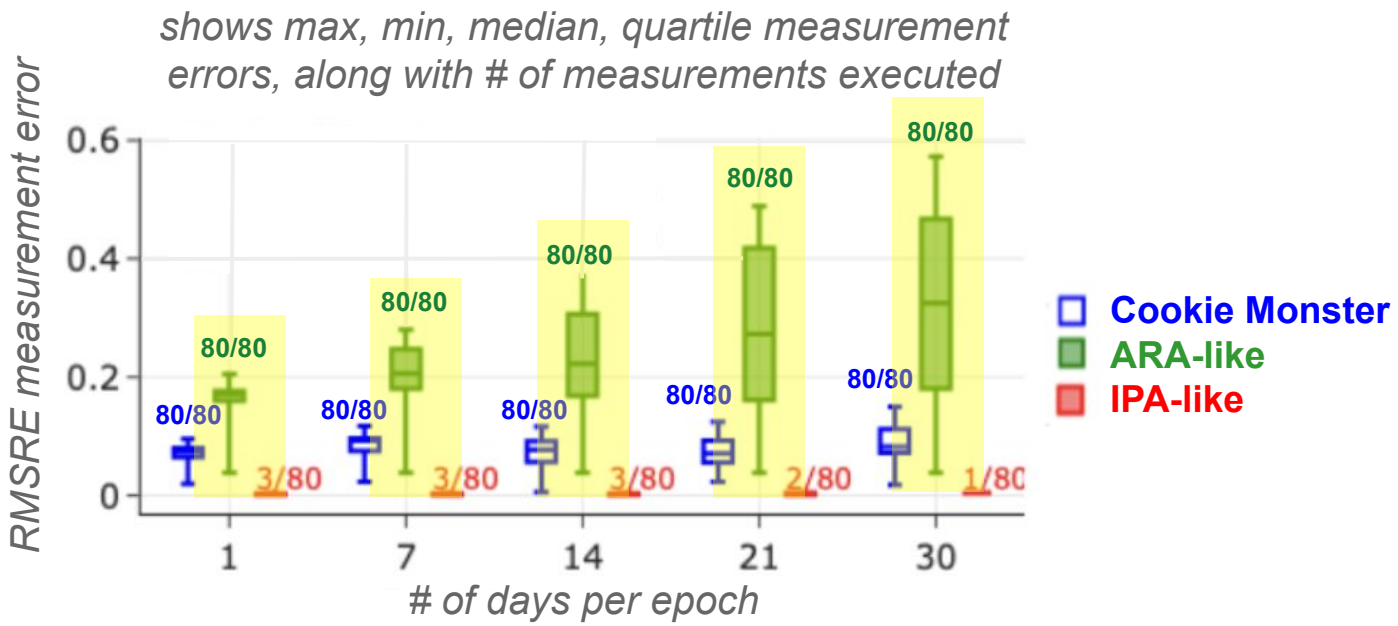
Utility metric
(lower is better at high numbers of measurements executed)



Privacy metric
(higher is better)

Evaluation on PATCG Dataset

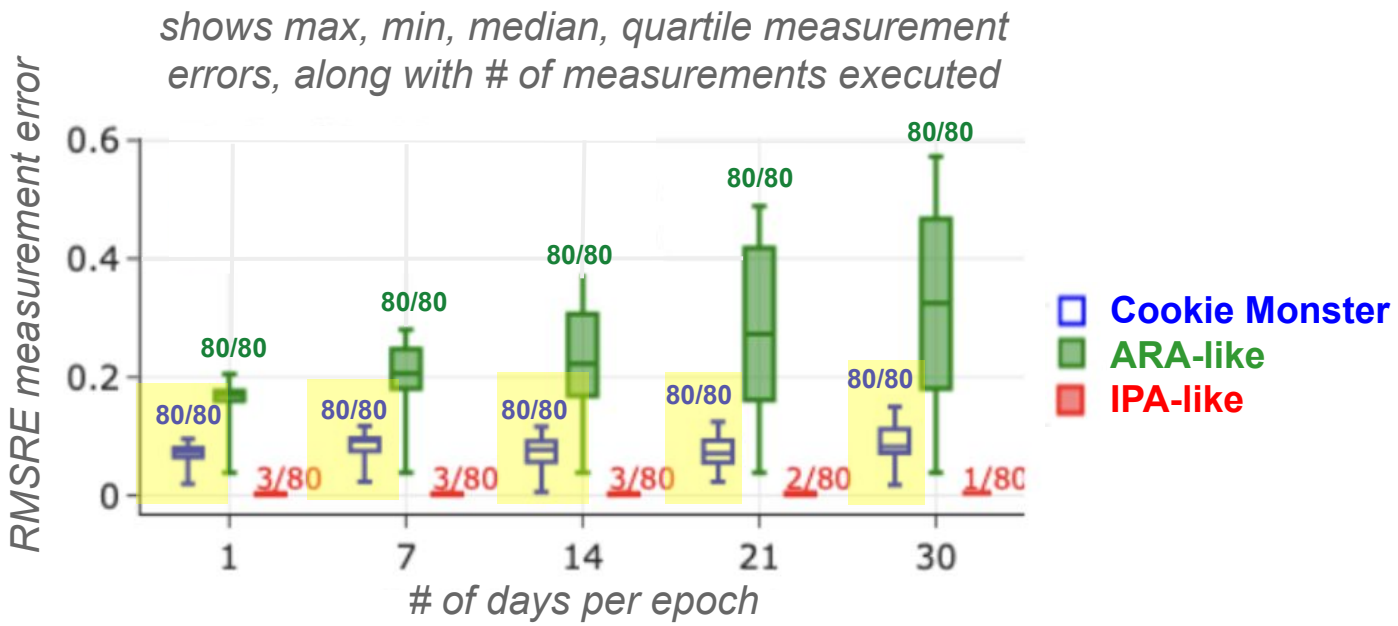
Utility metric
(lower is better at high numbers of measurements executed)



Privacy metric
(higher is better)

Evaluation on PATCG Dataset

Utility metric
(lower is better at high numbers of measurements executed)

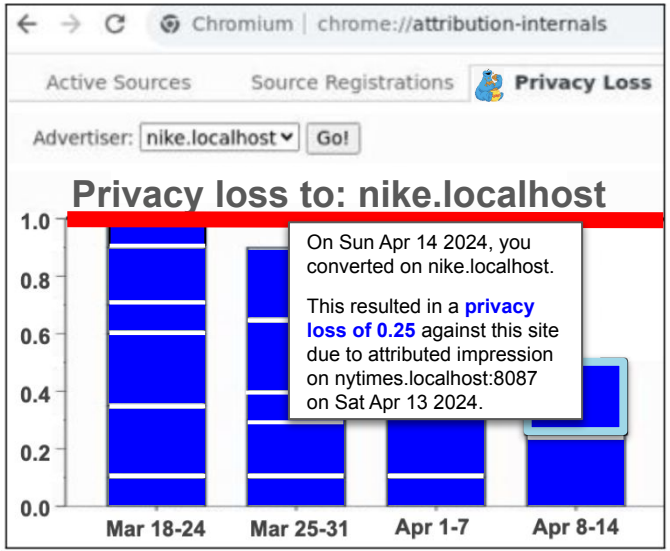


Privacy metric
(higher is better)

Other Compelling Properties

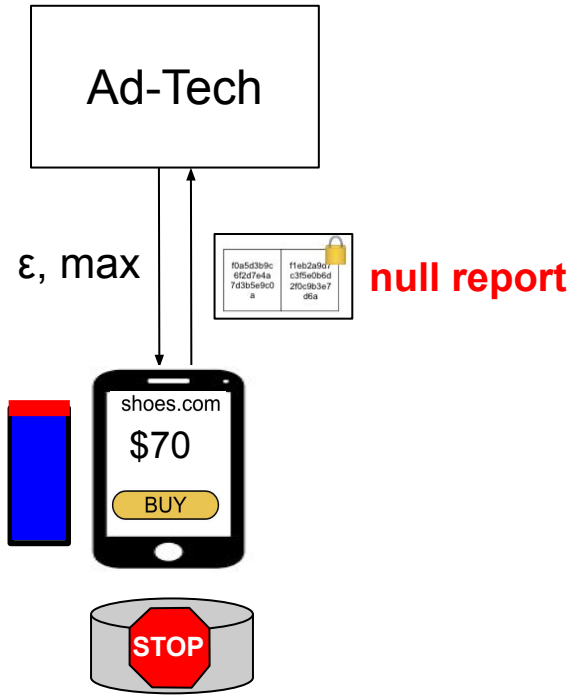
- **Transparency:** user can track their own privacy loss against different parties (see screenshot).
- **Control:** user's device is in charge of tracking and capping the user's own privacy loss to a value that can be tuned by the user.
- All come from our **IDP formulation**, which lets each device maintain its own DP guarantee instead of putting all devices under the same umbrella, as traditional DP would.

Screenshot of Cookie Monster's Privacy Loss Dashboard in Chrome



Downside: Can Add Bias

- Individual privacy loss depends on user data, so it **must be kept private**.
- Thus, when the browser reaches the privacy loss cap for the ad-tech, it stops real-data attributions, sending **null reports** instead.
- This can introduce **bias in query results** (although our experiments show that total error is still much smaller than ARA's, b/c our accounting is so efficient).

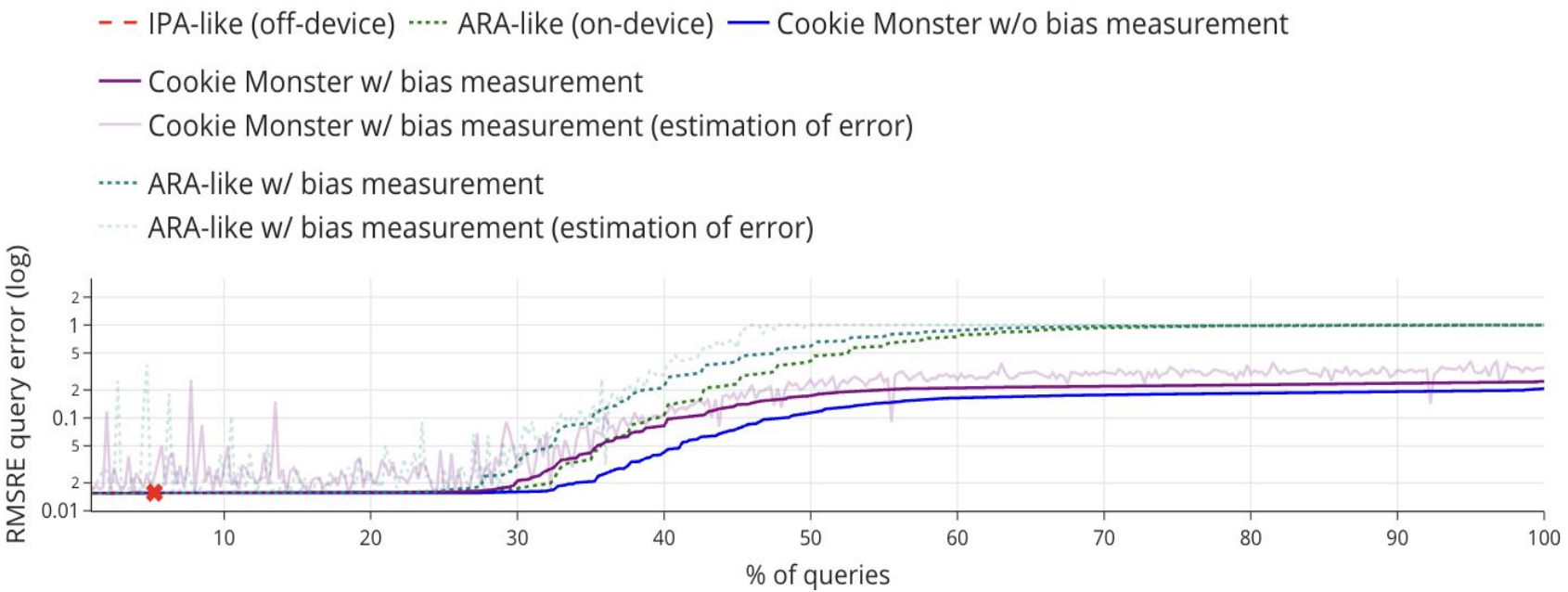


Preliminary Solution

Measure bias just like user data, using MPC, DP, etc.

- Alongside each true measurement, run a **side-measurement** that tracks reports potentially affected by running out of budget.
- Each device also accounts for privacy loss from this side-measurement.
- The device will include in each report both the original attribution and a 0/1 flag indicating if budget exhaustion affected it, all encrypted for the MPC/TEE.
- The ad-tech submits batches of these two-piece reports to SAS, which returns, along with the attribution measurement results, a DP count of affected reports, allowing for a bound on error.
- Preliminary results show Cookie Monster with bias measurement still delivers privacy-utility improvements, but we need tighter bounds on bias estimation.

Bias Evaluation on Microbenchmark



Outline

- Background on ad measurements and emerging APIs
- Our privacy framework: Cookie Monster
- **Discussion on broader applications and bias mitigation**

Questions

Broader Applications:

- What other cross-site measurements use cases exist on the Web beyond advertising?
- Is architecture also applicable to in-app advertising too? What are the diffs?
- Could similar APIs be suitable to replace location tracking?

Bias Mitigation:

- Does measurement make sense, or is prevention necessary?
- How should we quantify bias when there are multiple possible answers?
- For last-touch (for example), if the last impression is unavailable, it preferable to attribute to an older impression, or report null?

Footnotes

[1] While referred to as Personalized Differential Privacy (PDP) in some papers, we use the term Individual Differential Privacy (IDP), as it better reflects the concept and aligns with individual sensitivity, the basis of the definition (what we informally call “each device’s contribution” in this presentation). This recent paper also uses IDP terminology.