Privacy-Preserving Systems (a.k.a., Private Systems)

CU Graduate Seminar

Instructor: Roxana Geambasu

Private Web Advertising

NOTE: A lot of this lecture is speculative in nature, as private advertising tech is not yet well established. You get your instructor's perspective, which should be thought as "somewhat informed hypothesis" as opposed to "known fact."

Web advertising is changing

- In the past, mostly done through tracking via third-party cookies, IP tracking, device fingerprinting
- Lots of companies tracked users' moves on the web (with or without direct identification), built profiles of them, targeted them with ads, and optimized/measured the effectiveness of their ad campaigns
- This is changing, in two ways:
  1. Major browsers are disabling third-party cookies and raking up defenses against IP and device fingerprinting
  2. New APIs for (i) targeting and (ii) measurement/optimization of ads are being introduced, which involve the browser and "advanced privacy technologies" to ensure privacy protection while permitting advertising

Example APIs and Proposals

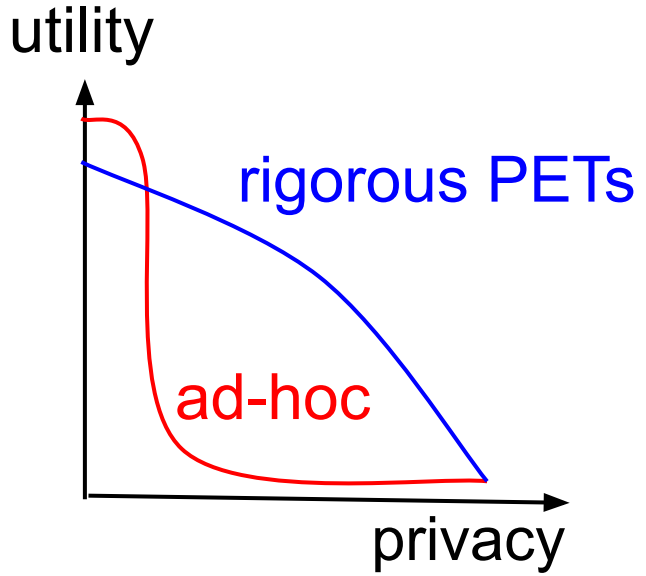
- Google's Privacy Sandbox contains multiple related APIs
  - Protected audience API (used for user targeting)
  - Attribution measurement API (used for ad effectiveness optimization, measurement)
- Mozilla/Meta's Interoperable Private Attribution (IPA)
  - Used for optimization/measurement
- Apple's Private Attribution Measurement (PAM)
  - Also for optimization/measurement
- Mozilla/Meta's "Hybrid" proposal
  - Also for optimization/measurement

- These APIs are still very much in flux, and there are discussions within W3C's PATCG community group about potential merges into inter-operable standards

Our opportunity/role

- Particularly promising about these changes is the focus on **rigorous privacy technologies** as building blocks, which in this instructor's view, open the door for **serious, qualitative improvements** of privacy on the web
- Academics and graduate students have a significant opportunity to contribute to this change: get involved in PATCG, learn about these APIs, systems, and their requirements, and help improve their designs so they can achieve the **best privacy-utility tradeoff**
- Unfortunately, without both privacy and utility, qualitative changes to the web are unlikely

typical privacy-utility curves



Our work at Columbia

- Collaborate with Meta and Mozilla on privacy components of an inter-operable ad-measurement API proposal
- Substantiate our proposals through scientific studies of the proposed privacy designs
- Ongoing engagement in PATCG to communicate our progress/designs, get feedback, and revise

- Main proposal to date: DP budgeting component for Meta/Mozilla's "Hybrid" proposal (design doc, scientific paper forthcoming)

Today

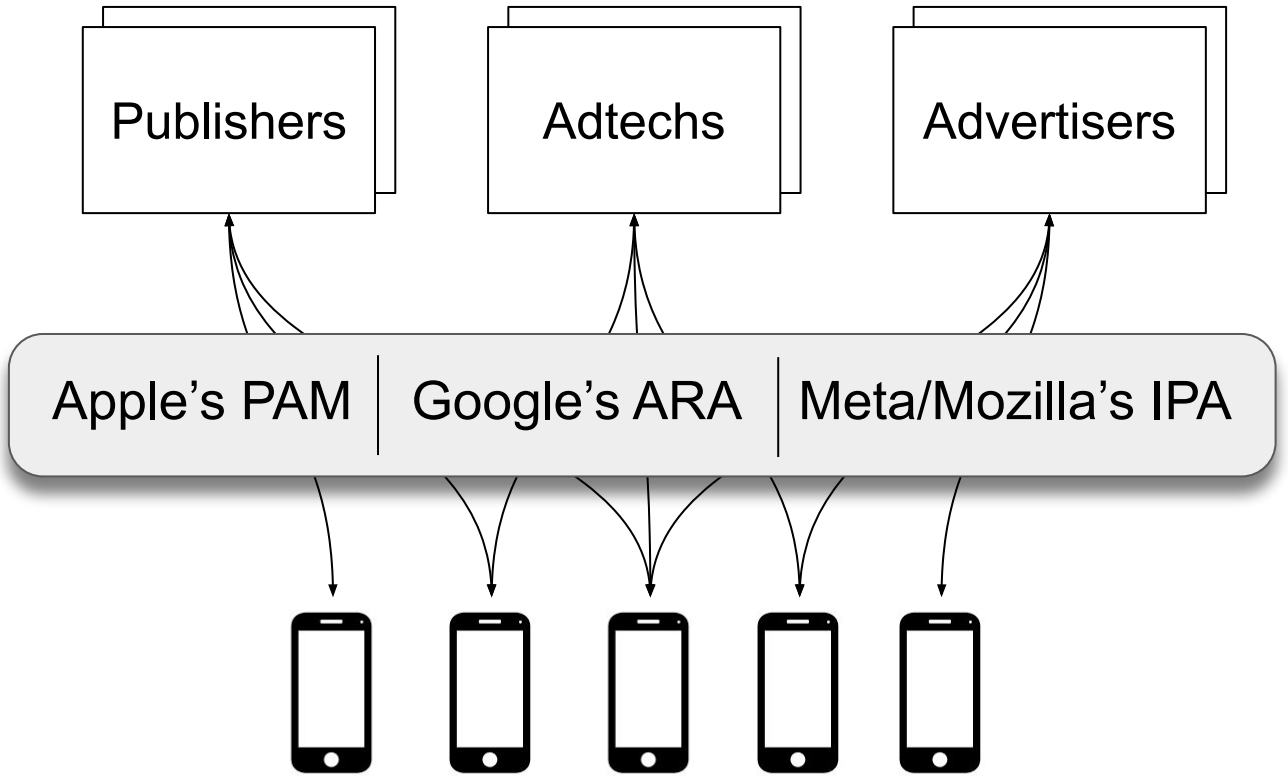
- Motivate and briefly describe our design of the DP budgeting component for a certain class of ad-measurement APIs, specifically those that do “on-device budgeting”
- This can serve as an example for the kind of work you too might be able to do in this space, with us or on your own
- My main message to you: with Private Systems knowledge, you have the background to engage an help out in this important change, so we can make it
  - Let me know if you’re interested in this, there’s LOTS of stuff to do here, including over the summer

Technical Take-Aways

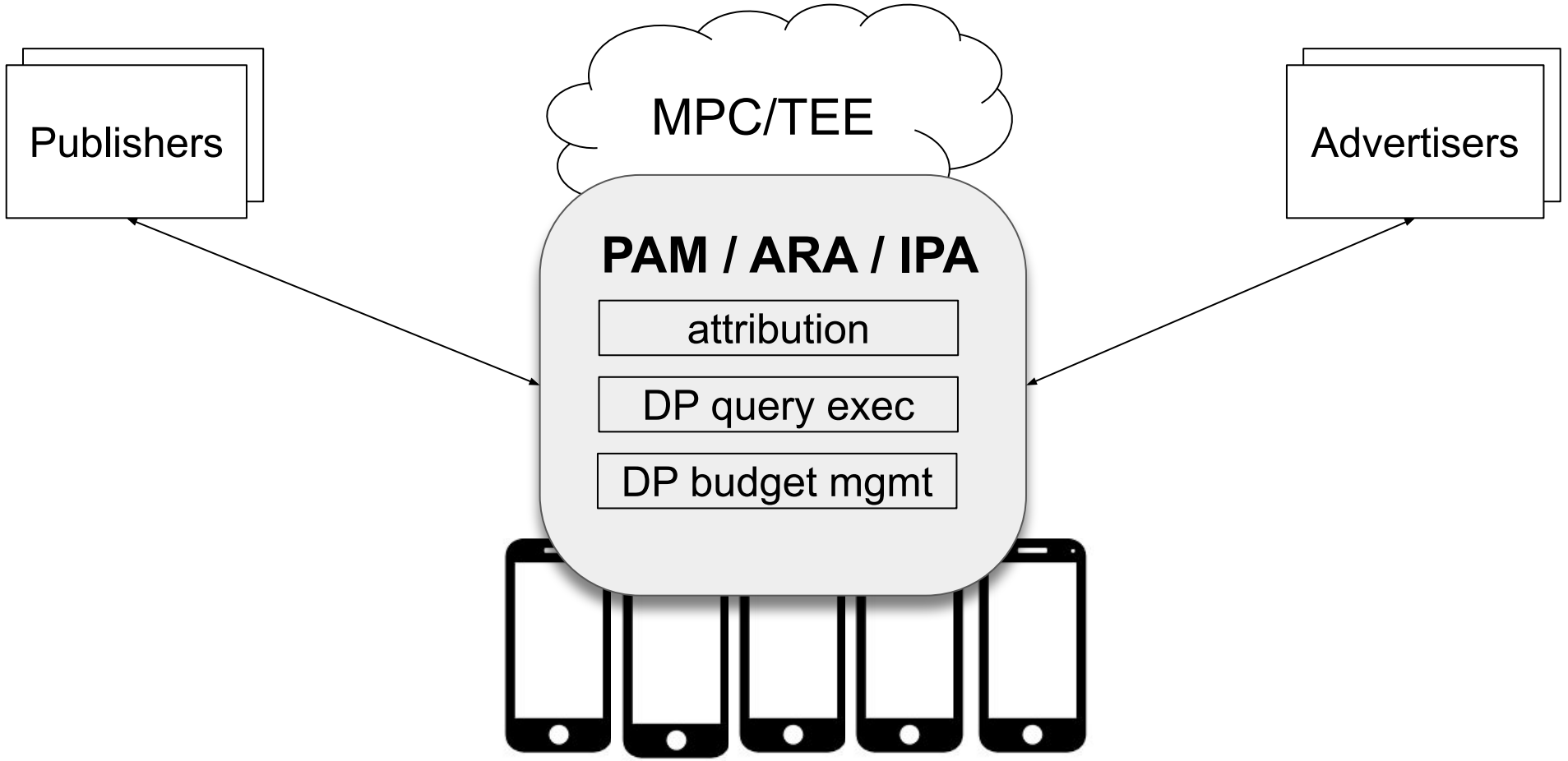
1. Traditional DP is not a good fit for on-device budgeting systems.
2. Individual DP (IDP) is better suited and can enable optimizations for more efficient budget management in these systems.
3. But IDP also brings negative consequences, such as the need to keep the privacy budgets private.

Private Ad-Measurement Systems

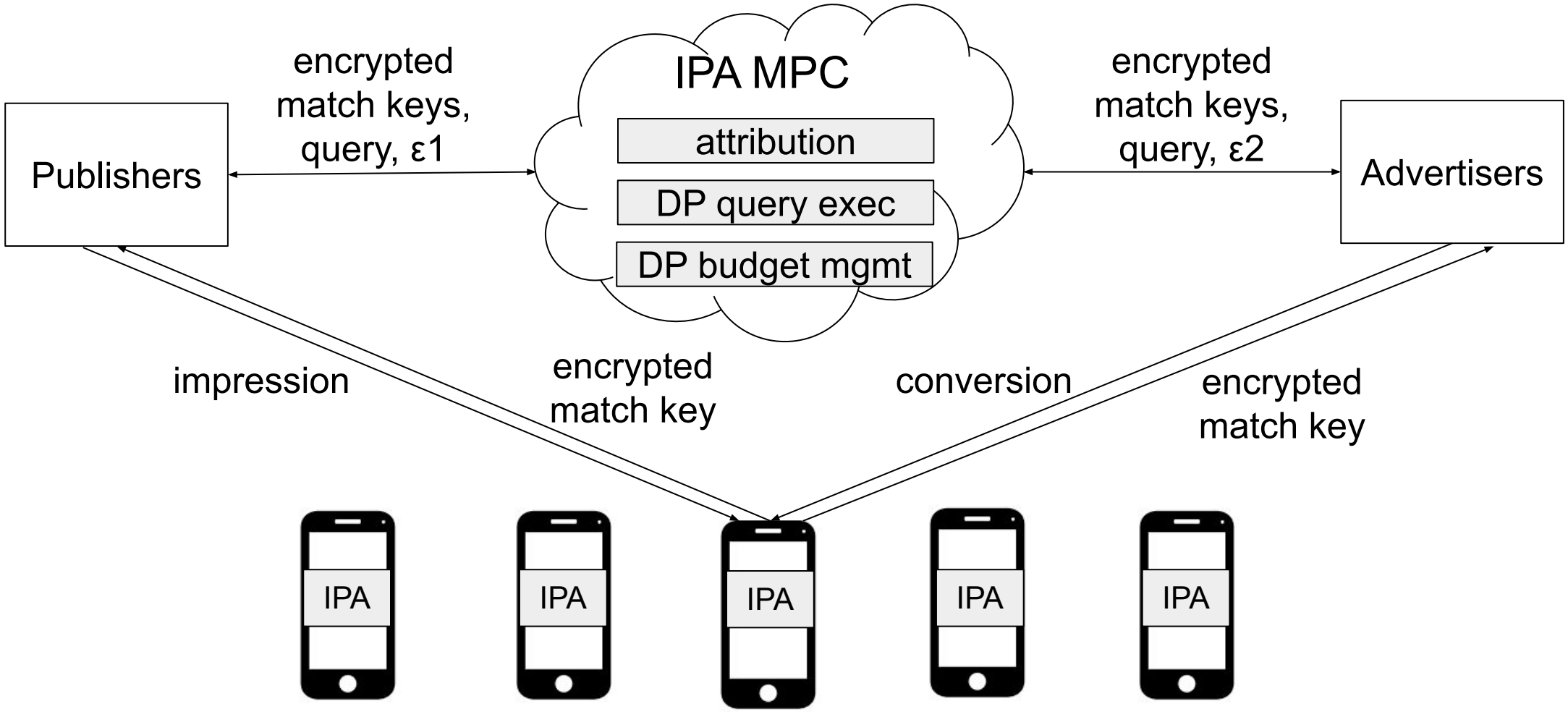
- Multiple designs, with differences and commonalities.
- Often designs incorporate DP mechanisms, but they lack clear formulations of DP desiderata.
- These desiderata are not obvious, especially for systems with **on-device budgeting**, such as PAM.



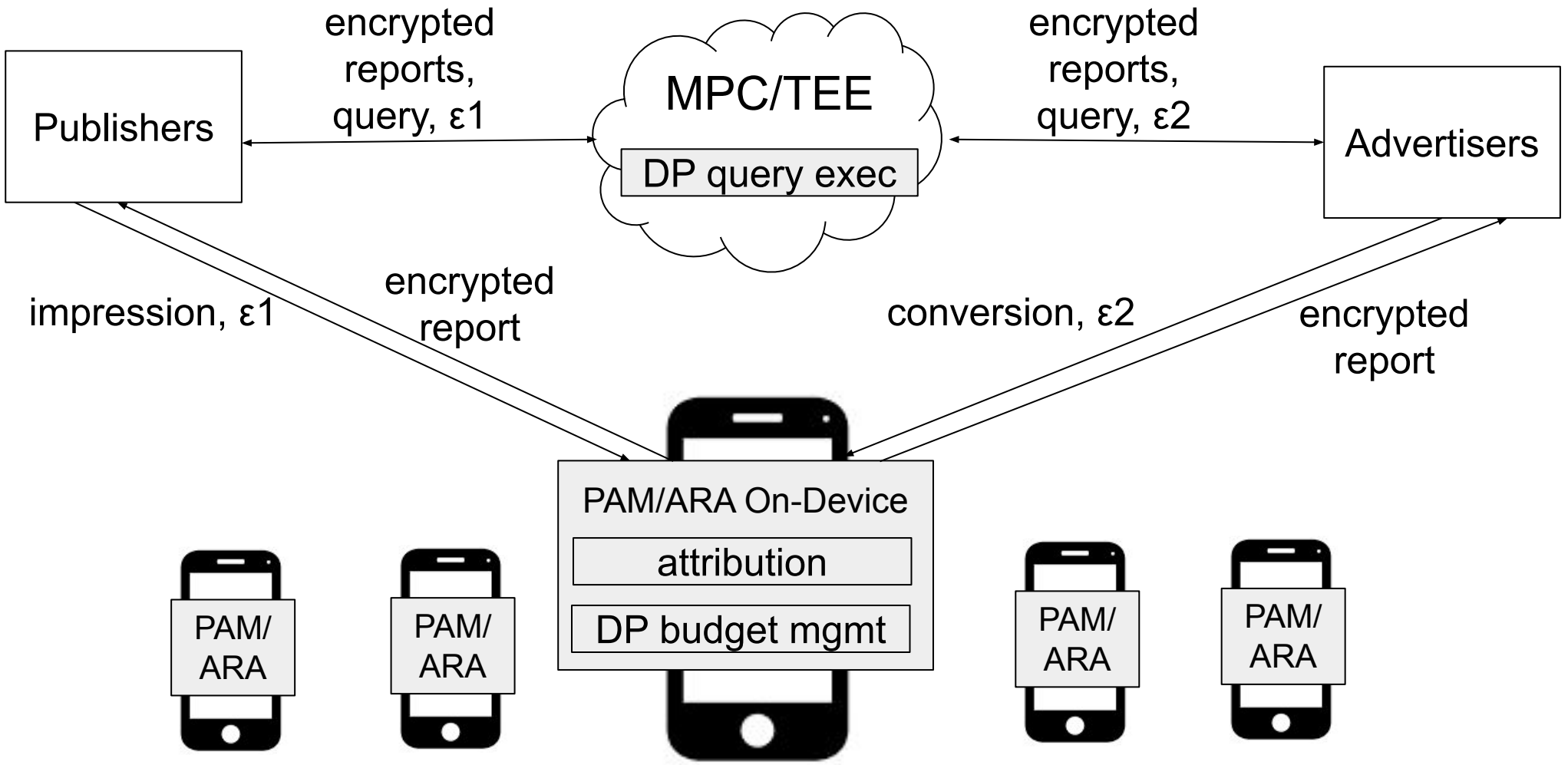
Common Architecture



Off-Device Designs (e.g., IPA)



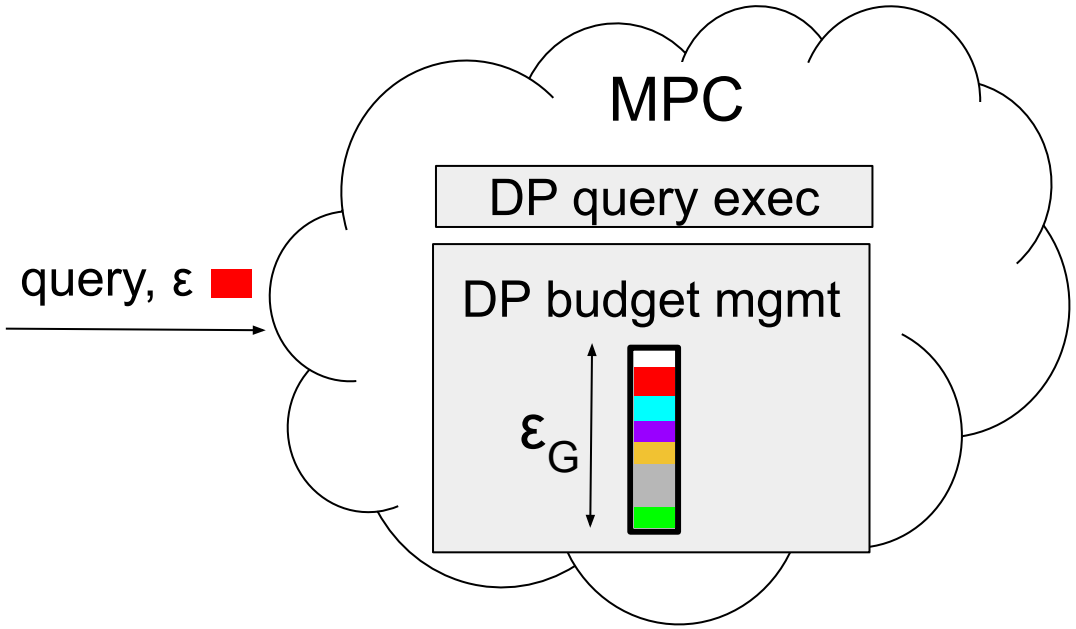
On-Device Designs (e.g., PAM, ARA)



DP Budget Mgmt: Off-Device vs. On-Device

(here, we discuss abstract models of operation, not specific systems)

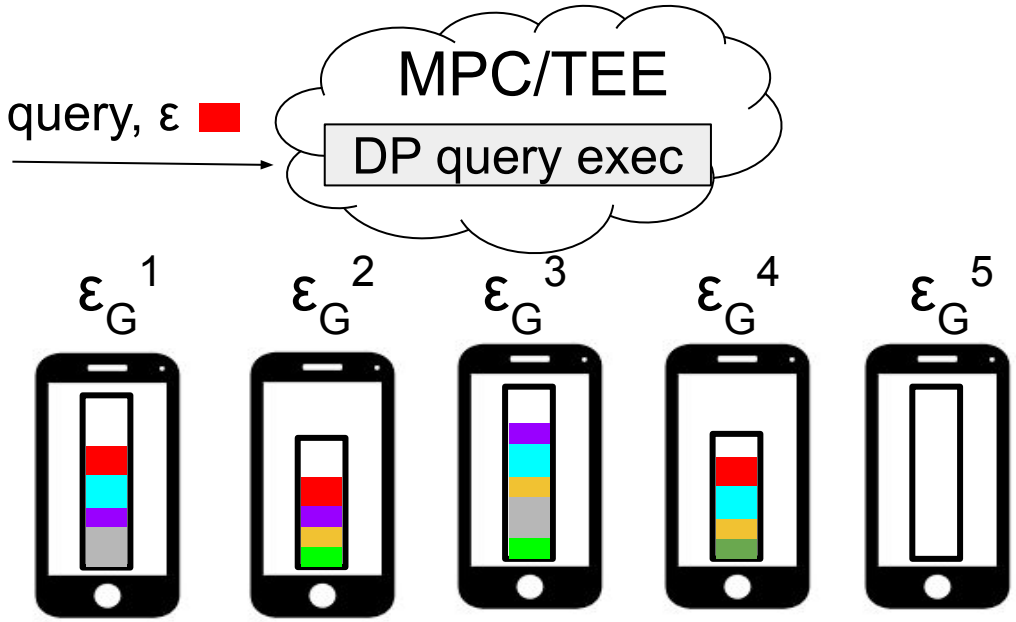
Off-device Budgeting



DP desideratum: Should satisfy ε_G-DP

(above: informal and under-specified)

On-device Budgeting

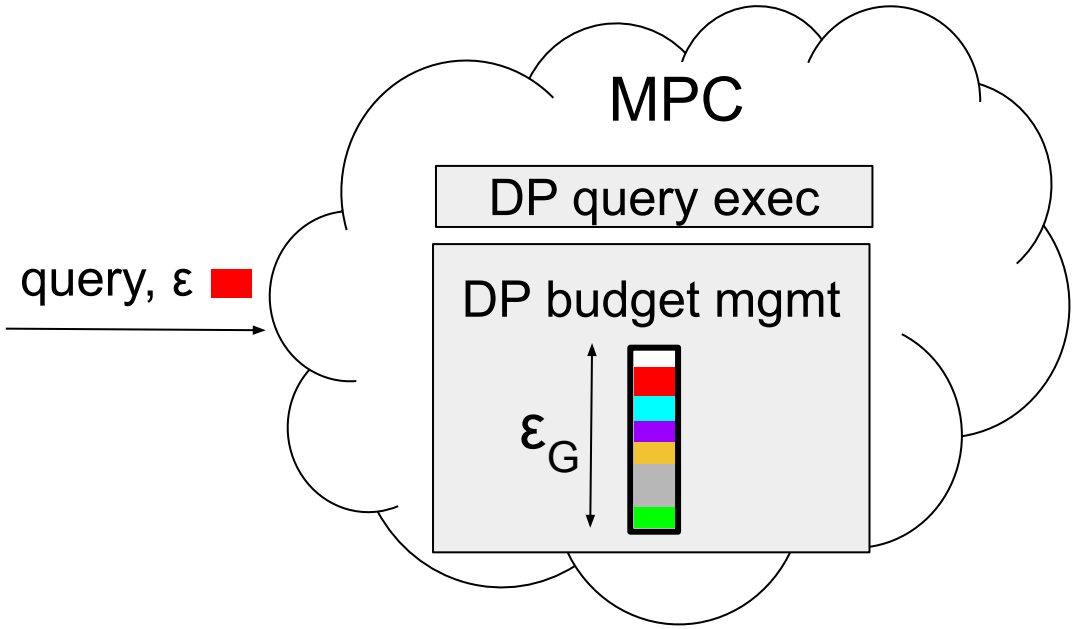


DP desideratum: ???

DP Budget Mgmt: Off-Device vs. On-Device

(here, we discuss abstract models of operation, not specific systems)

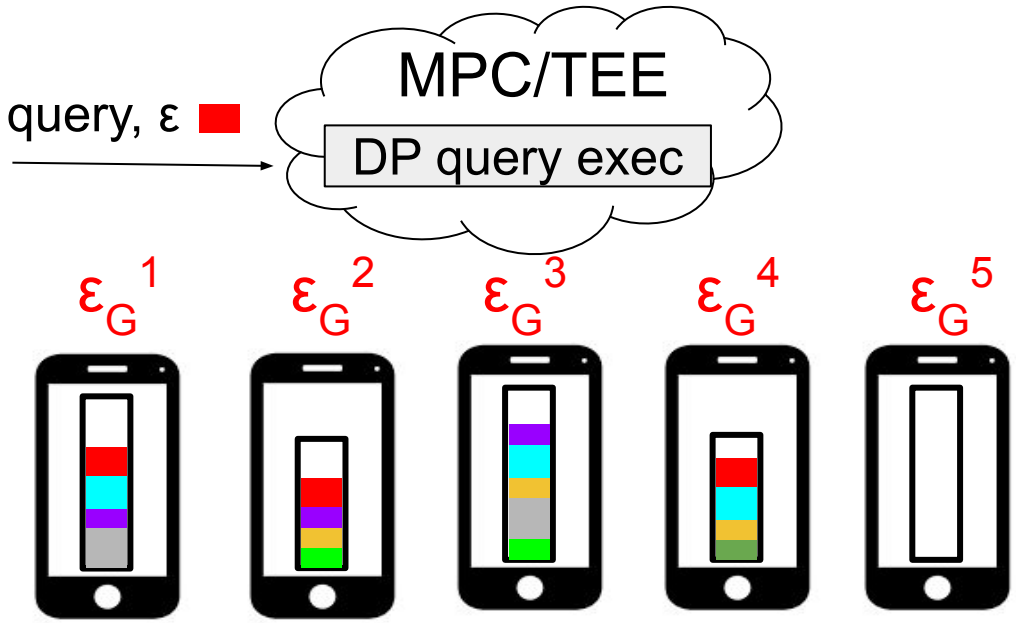
Off-device Budgeting



DP desideratum: Should satisfy ε_G-DP

(above: informal and under-specified)

On-device Budgeting



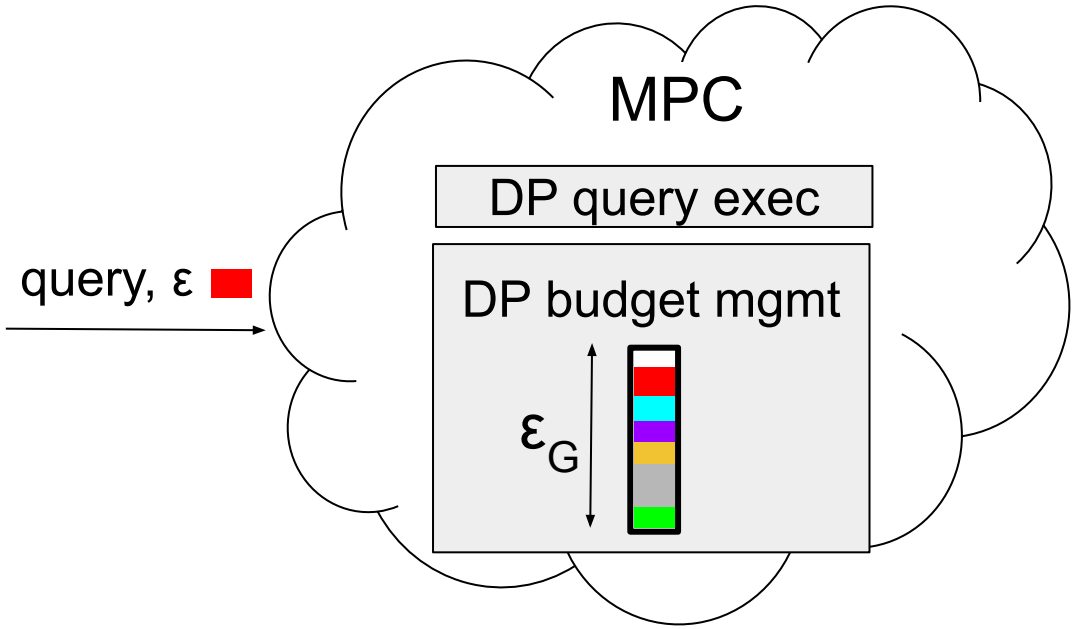
DP desideratum: ???

a) different ε_G settings

DP Budget Mgmt: Off-Device vs. On-Device

(here, we discuss abstract models of operation, not specific systems)

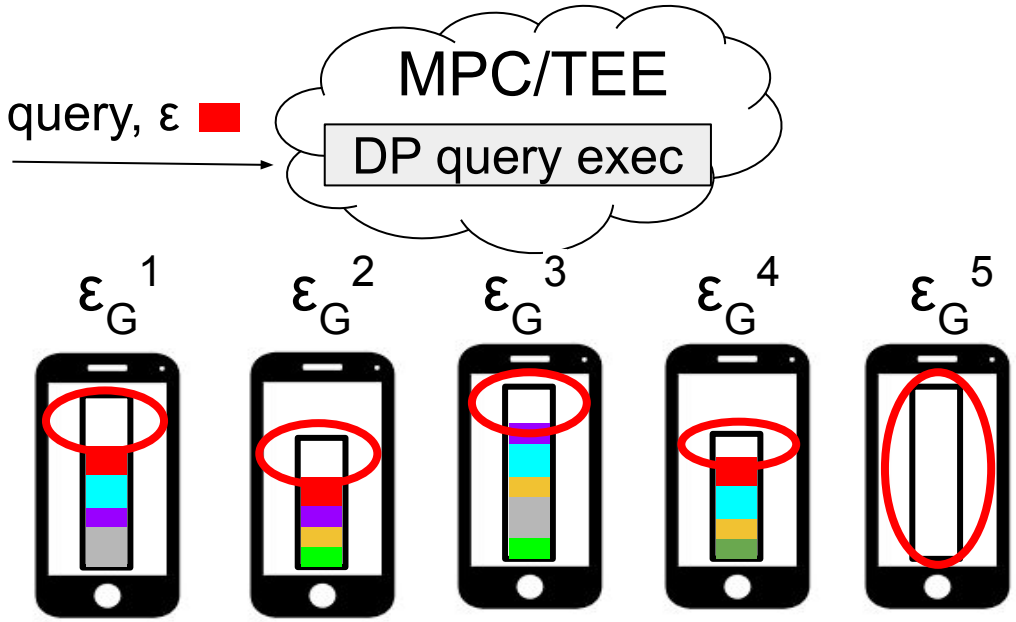
Off-device Budgeting



DP desideratum: Should satisfy ε_G-DP

(above: informal and under-specified)

On-device Budgeting



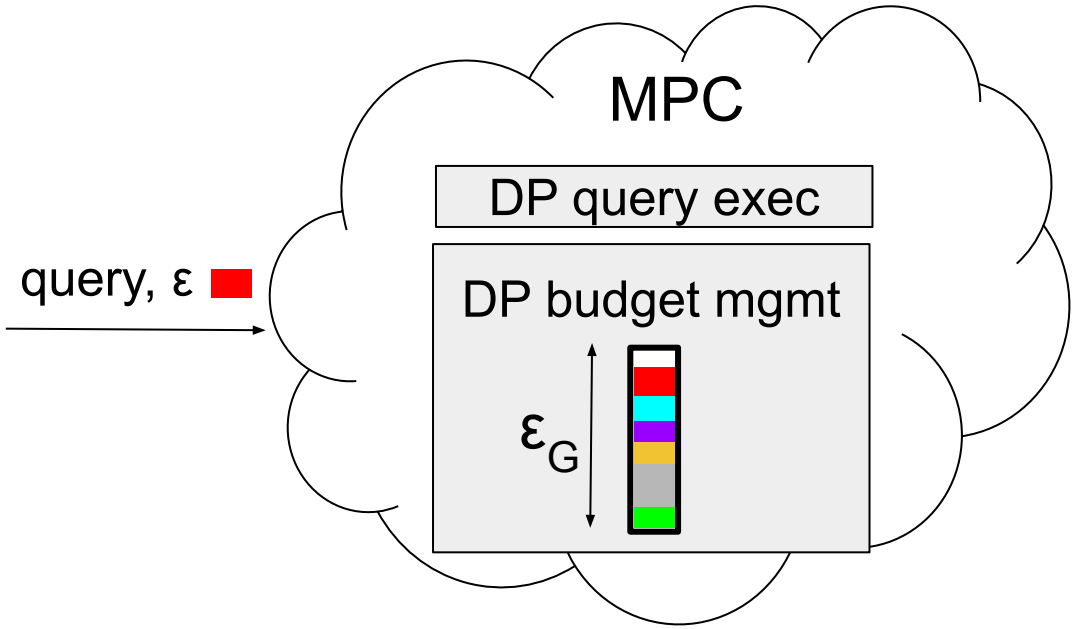
DP desideratum: ???

b) different remaining budgets

DP Budget Mgmt: Off-Device vs. On-Device

(here, we discuss abstract models of operation, not specific systems)

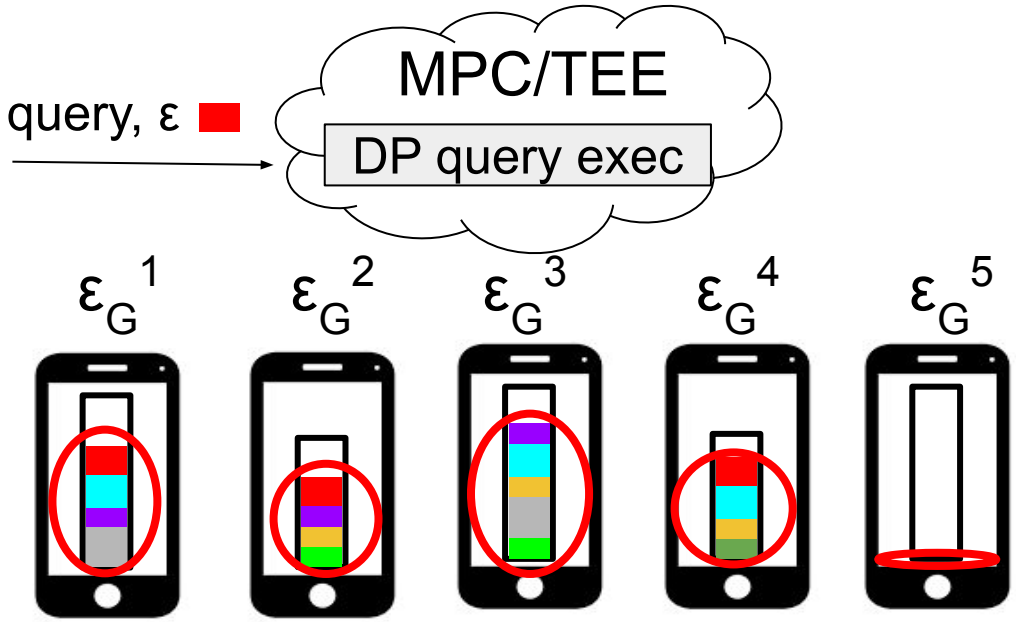
Off-device Budgeting



DP desideratum: Should satisfy ε_G-DP

(above: informal and under-specified)

On-device Budgeting



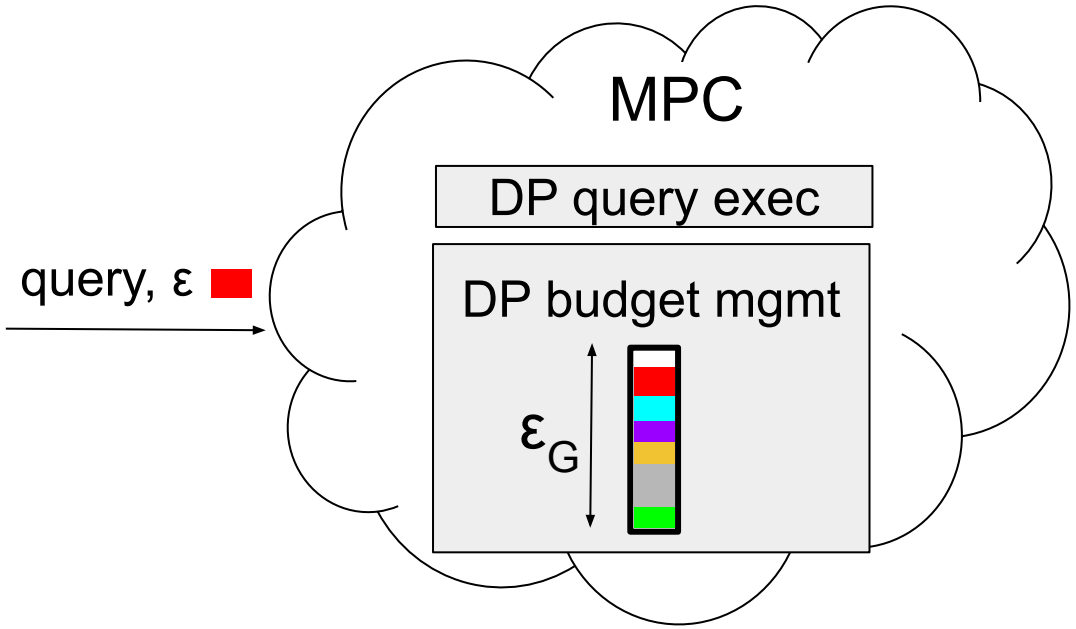
DP desideratum: ???

c) different consumed budgets

DP Budget Mgmt: Off-Device vs. On-Device

(here, we discuss abstract models of operation, not specific systems)

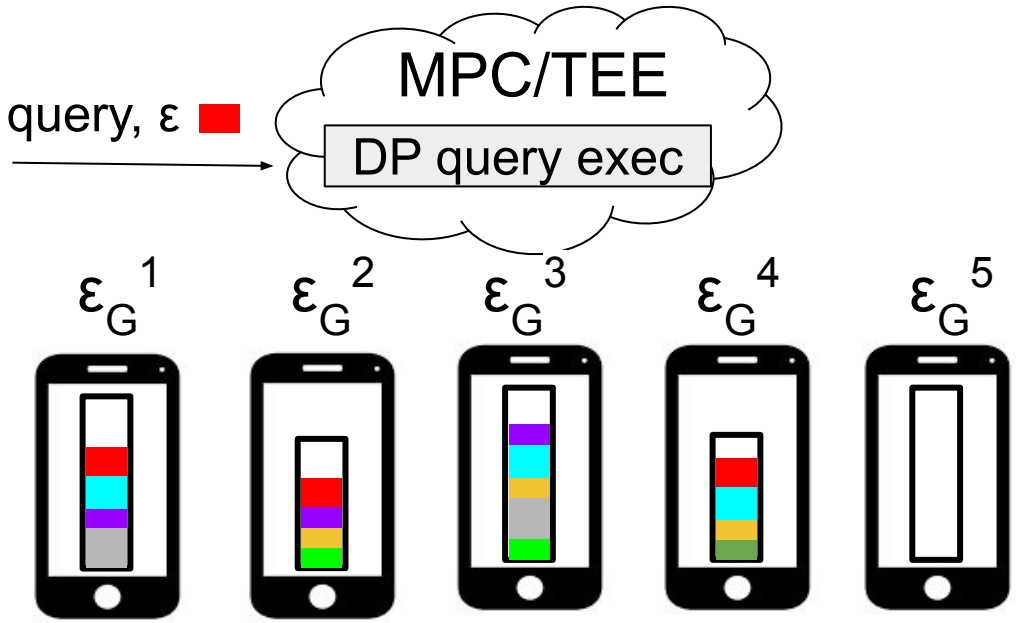
Off-device Budgeting



DP desideratum: Should satisfy ε_G-DP

(above: informal and under-specified)

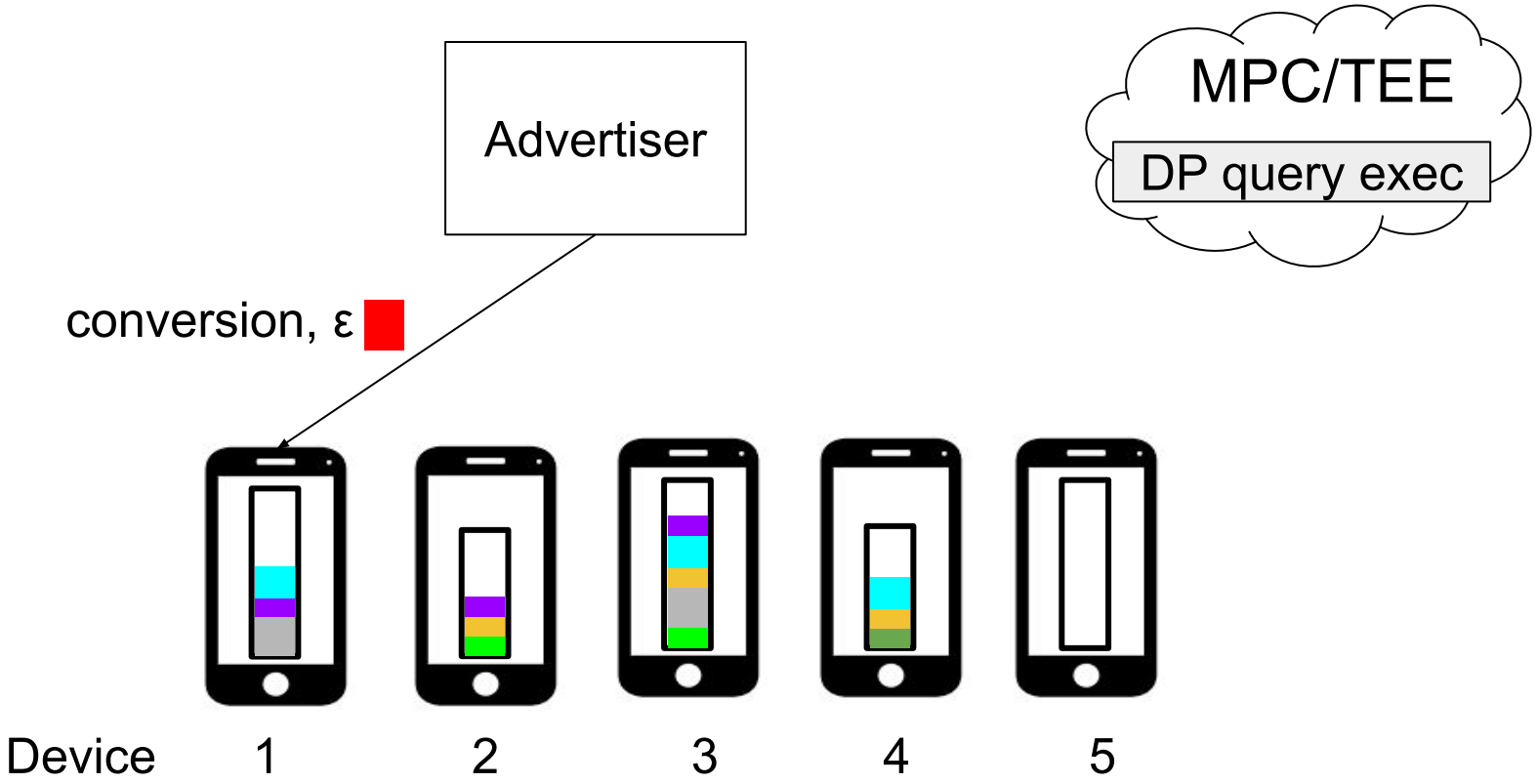
On-device Budgeting



DP desideratum: doesn't fit traditional DP well…

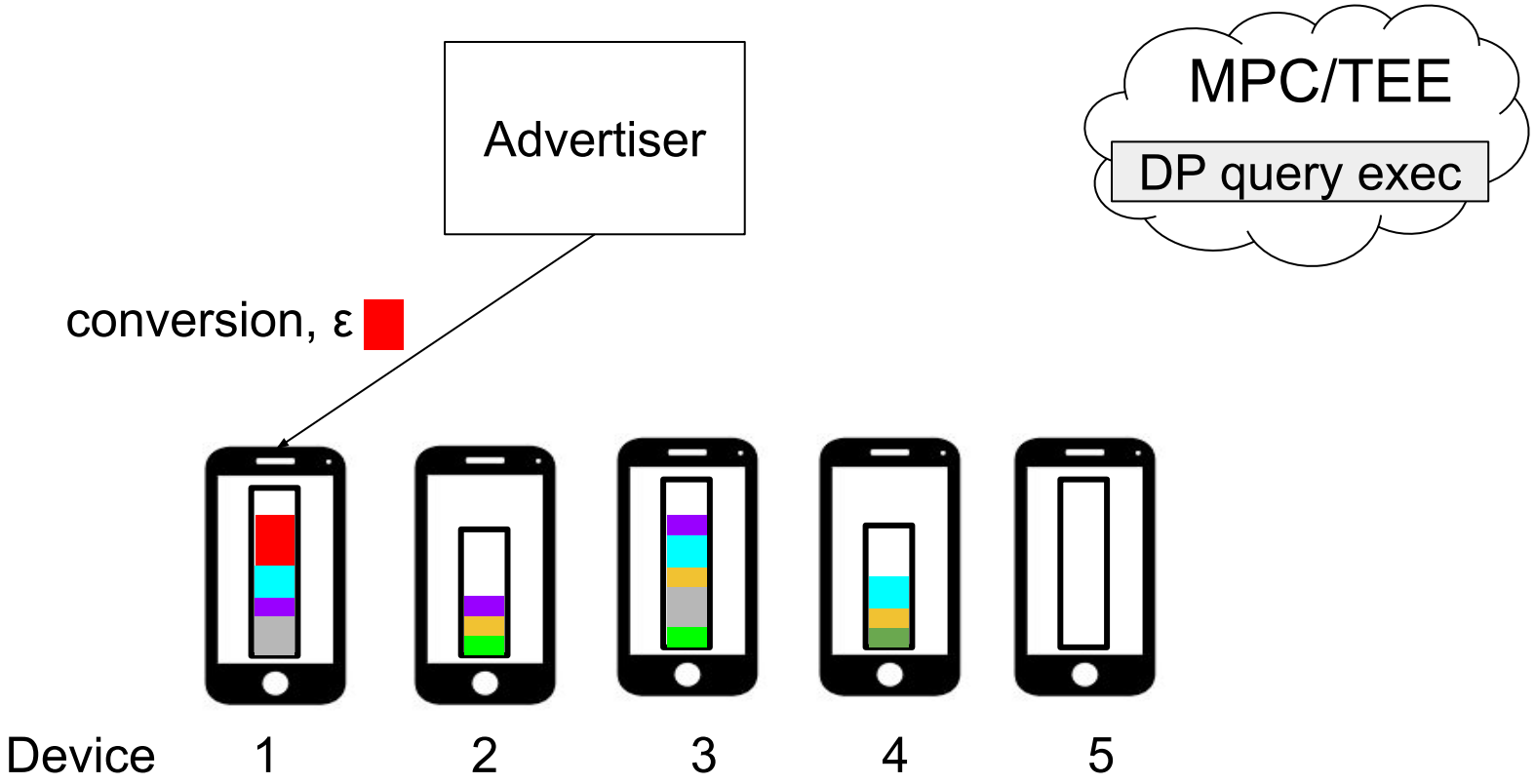
d) A Particularly Peculiar Behavior

On-device designs may exhibit a behavior that would be particularly challenging to capture with traditional DP.



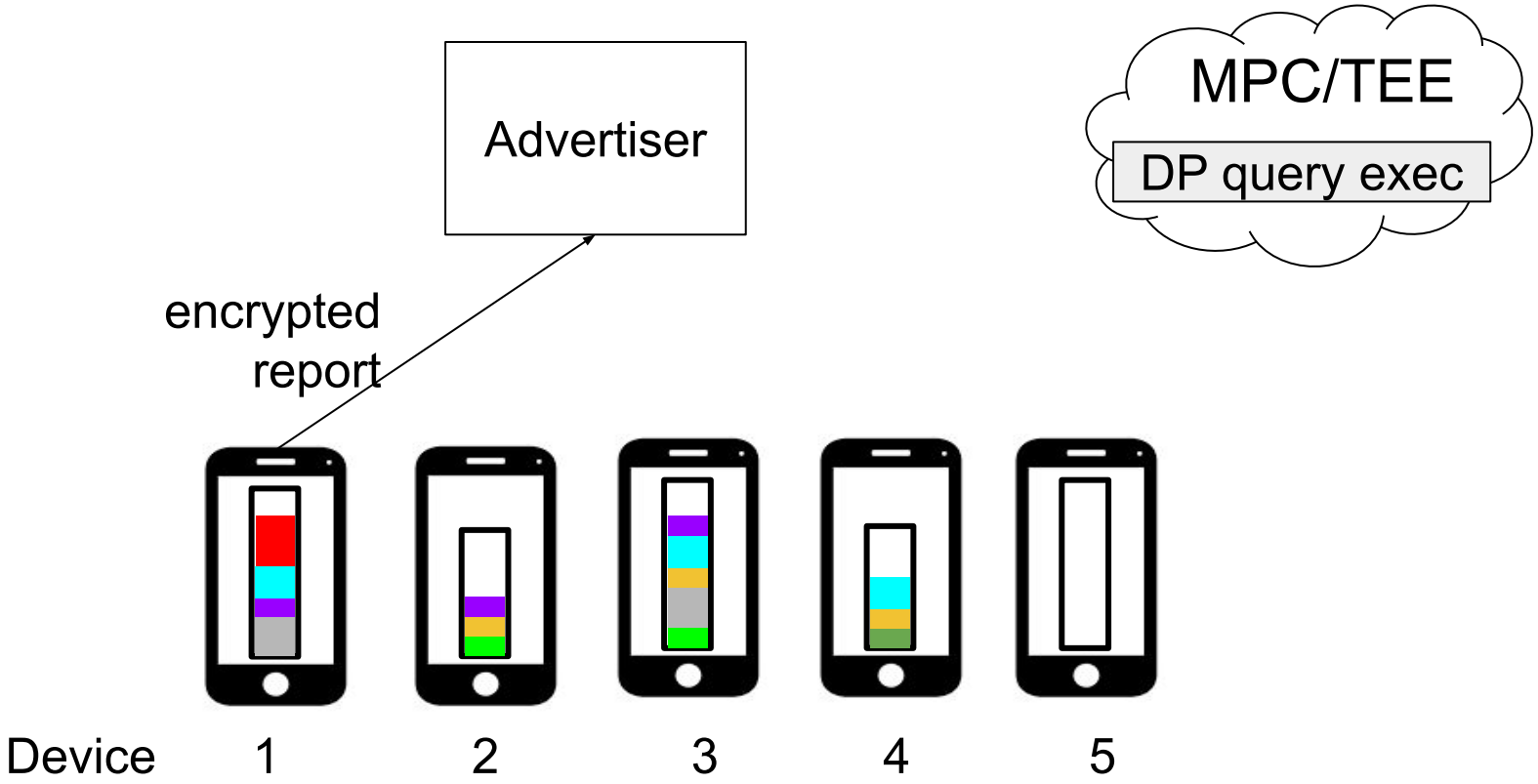
d) A Particularly Peculiar Behavior

On-device designs may exhibit a behavior that would be particularly challenging to capture with traditional DP.



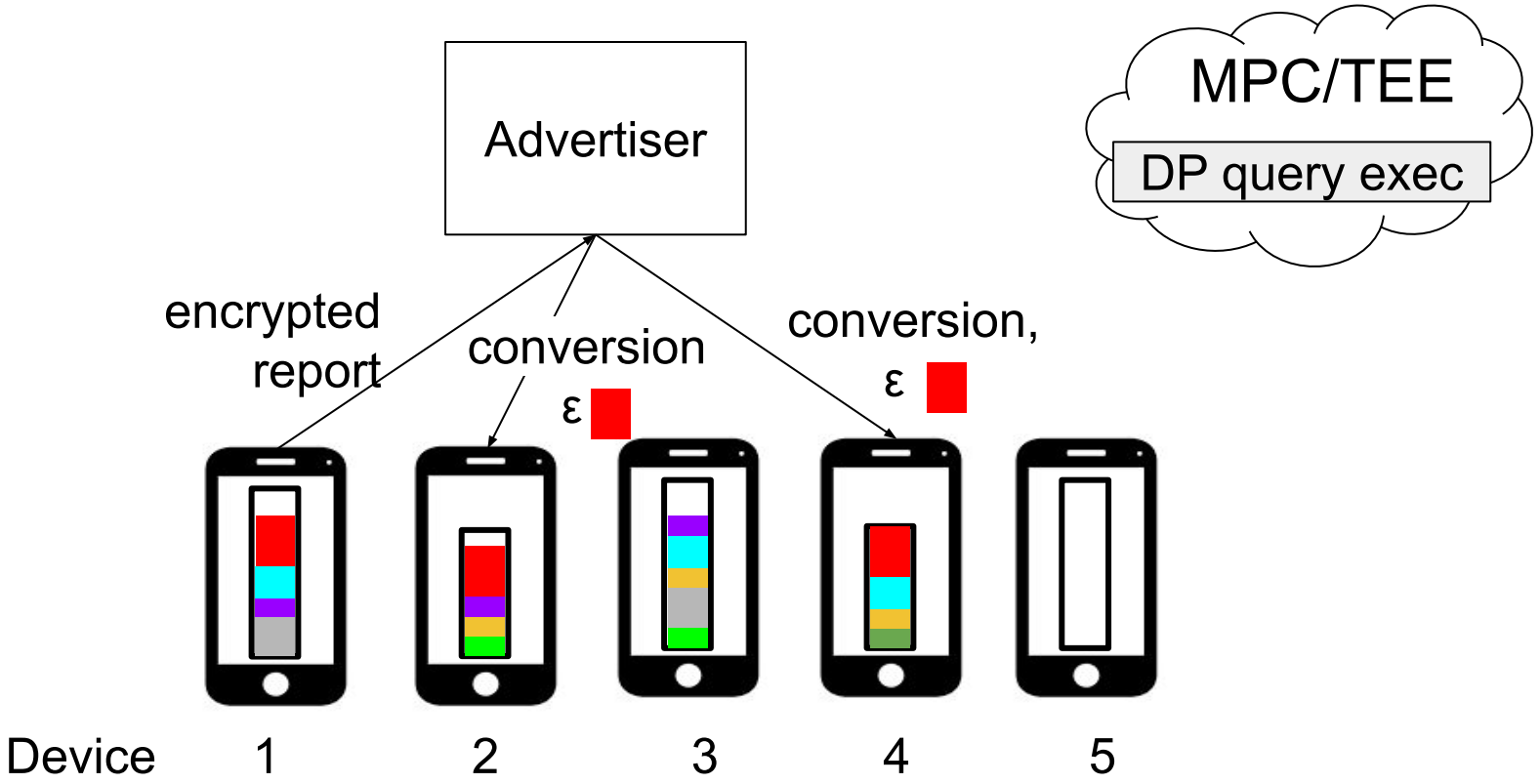
d) A Particularly Peculiar Behavior

On-device designs may exhibit a behavior that would be particularly challenging to capture with traditional DP.



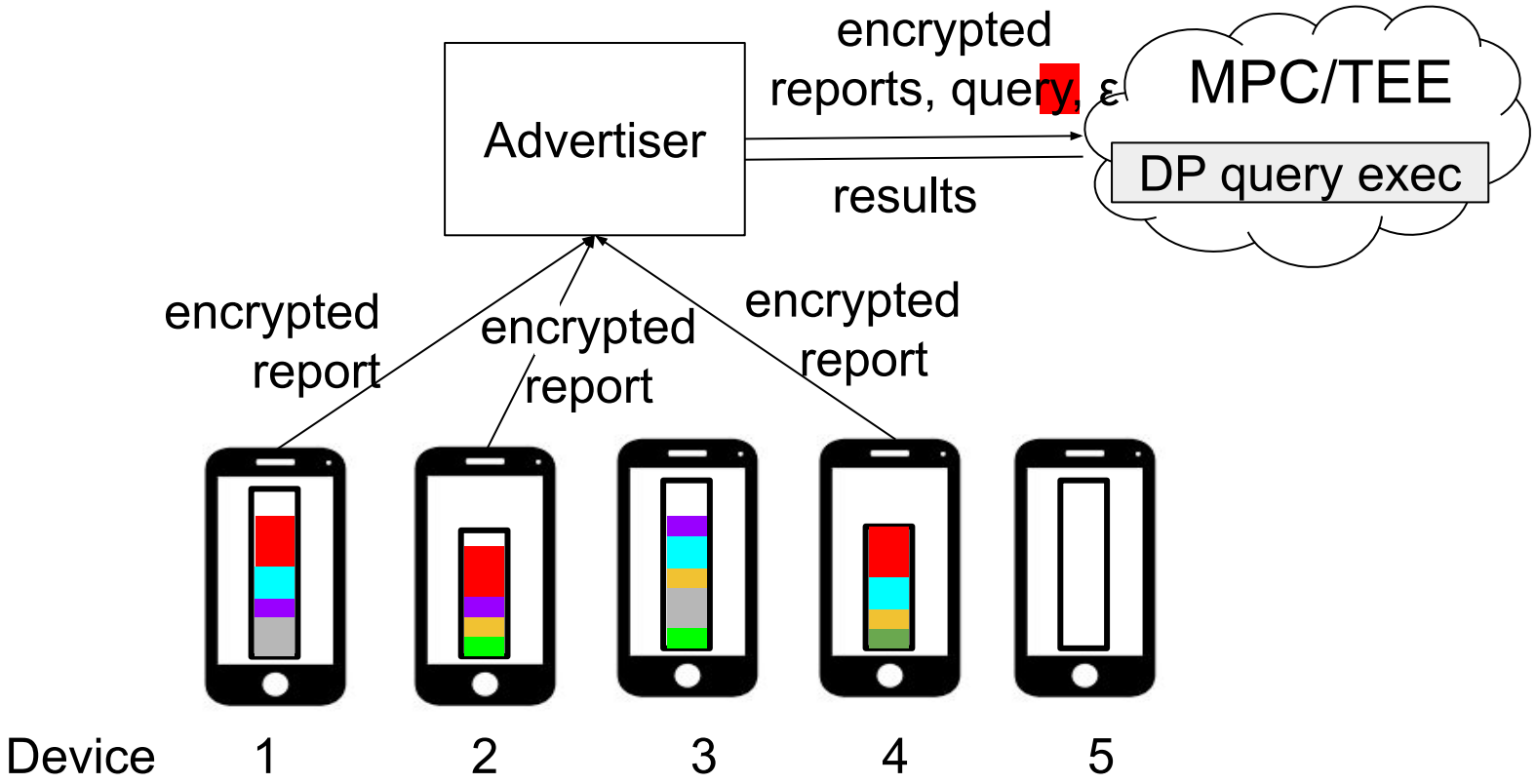
d) A Particularly Peculiar Behavior

On-device designs may exhibit a behavior that would be particularly challenging to capture with traditional DP.



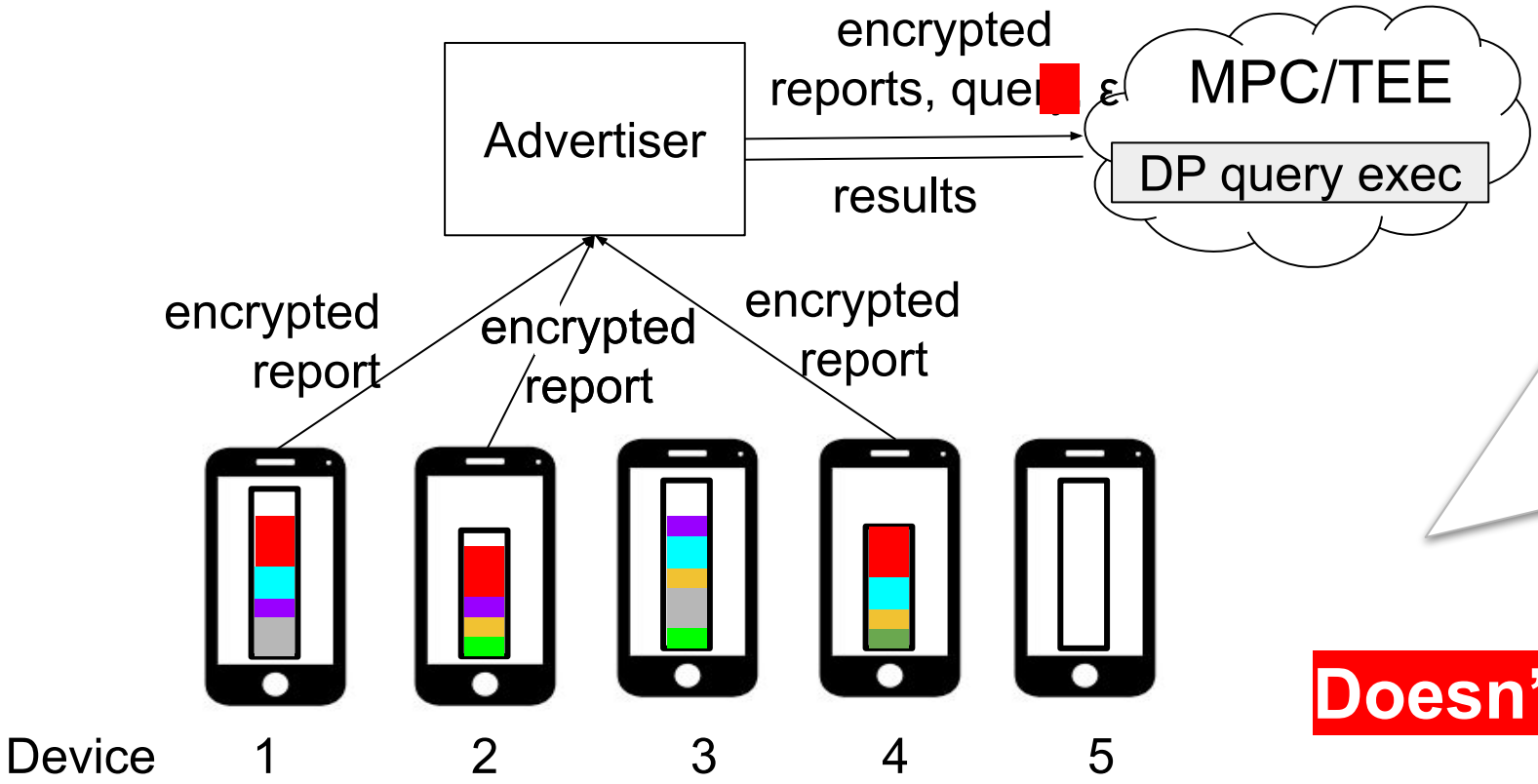
d) A Particularly Peculiar Behavior

On-device designs may exhibit a behavior that would be particularly challenging to capture with traditional DP.



d) A Particularly Peculiar Behavior

On-device designs may exhibit a behavior that would be particularly challenging to capture with traditional DP.



What just happened?

- Devices 1,2,4 consumed budget because their users had the conversion.
- Devices 3,5 did not consume budget b/c their users did not have the conversion.
- Budget is consumed (or not) based on **USER DATA!?**

Doesn't fit traditional DP, either…

(unless we model full conversion data as public info for advertiser)

Individual DP (IDP)
(a.k.a., Personalized DP)

IDP (*)

- Permits different users participating in centralized-DP computations to maintain separate DP guarantees.

(*) The definition we use here usually appears under the name of personalized DP (including in POPL'15), but more recently has appeared under the name of individual DP (NeurIPS'21). We adopt the individual-DP terminology b/c we consider it more intuitive and in line with other notions closely related to this definition (namely, individual sensitivity).

IDP (*)

- Permits different users/devices participating in centralized-DP queries to maintain **separate ε_i-DP guarantees**, and to account for privacy loss on the basis of **their data**.

Definition (from POPL'15): We say that data sets A and B differ in record r, written **$A \sim^r B$**, if A can be obtained from B by adding record r, or vice-versa.
Let ξ be a function from records to non-negative real numbers.
A randomized query Q provides ξ-individual DP if for all records r:
for all $A \sim^r B$ and any set of outputs $S \subseteq range(Q)$, we have:

$$Pr[Q(A) \in S] \leq Pr[Q(B) \in S] \cdot e^{\xi(r)}$$

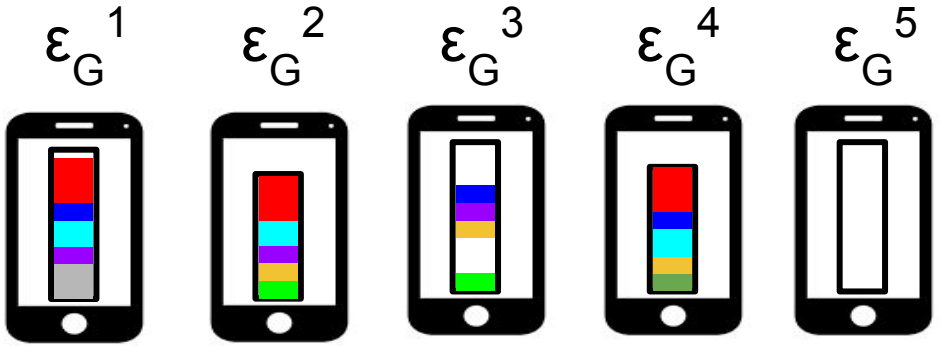
(*) We use the definition from POPL'15 (enclosed in slide), not to be confused with a DP relaxation introduced under the "individual DP" name. Ours is a generalization of DP, not a relaxation of it.

IDP Properties

- From the perspective of an individual user/device, individual DP is **as protective as** DP.
- All important DP properties hold for IDP (composition, post-processing, group privacy, adaptivity in privacy budgets).

If Q is ξ-IDP and sup(range(ξ)) = ε, then Q is also ε-DP.
If Q is ε-DP, then Q is ξ-IDP for ξ(r)=ε for all r.

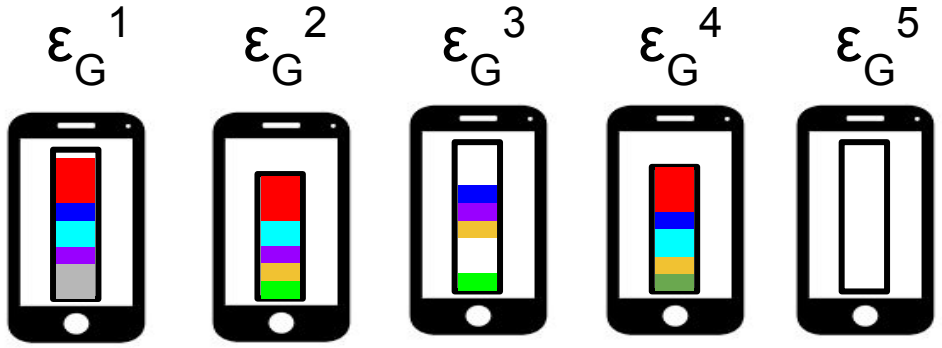
On-Device Budgeting Desideratum



DP Desideratum: Should satisfy ξ-IDP for $\xi(r) = \varepsilon_G^r$ for each device r.

(above: informal and under-specified – we're working on more complete spec, including user-epoch)

On-Device Budgeting Desideratum

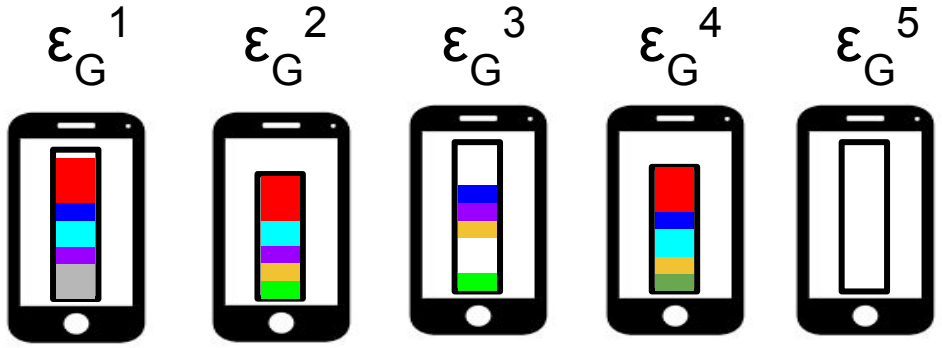


DP Desideratum: Should satisfy ξ-IDP for $\xi(r) = \varepsilon_G^r$ for each device r.

(above: informal and under-specified – we're working on more complete spec, including user-epoch)

- IDP lets us capture these systems' behaviors cleanly:
  a) different ε_G^r settings for each device r
  b) different remaining budgets
  c) different consumed budgets
  d) peculiar behavior of deducting zero for devices without a conversion

On-Device Budgeting Desideratum



DP Desideratum: Should satisfy ξ-IDP for $\xi(r) = \varepsilon_G^r$ for each device r.

(above: informal and under-specified – we're working on more complete spec, including user-epoch)

- IDP lets us capture these systems' behaviors cleanly:
  - a) different ε_G^r settings for each device r
  - b) different remaining budgets
  - c) different consumed budgets

  } easy to justify

  - **d) peculiar behavior of deducting zero for devices without a conversion**

  } **need individual sensitivity to justify**

Individual Sensitivity

Traditional DP

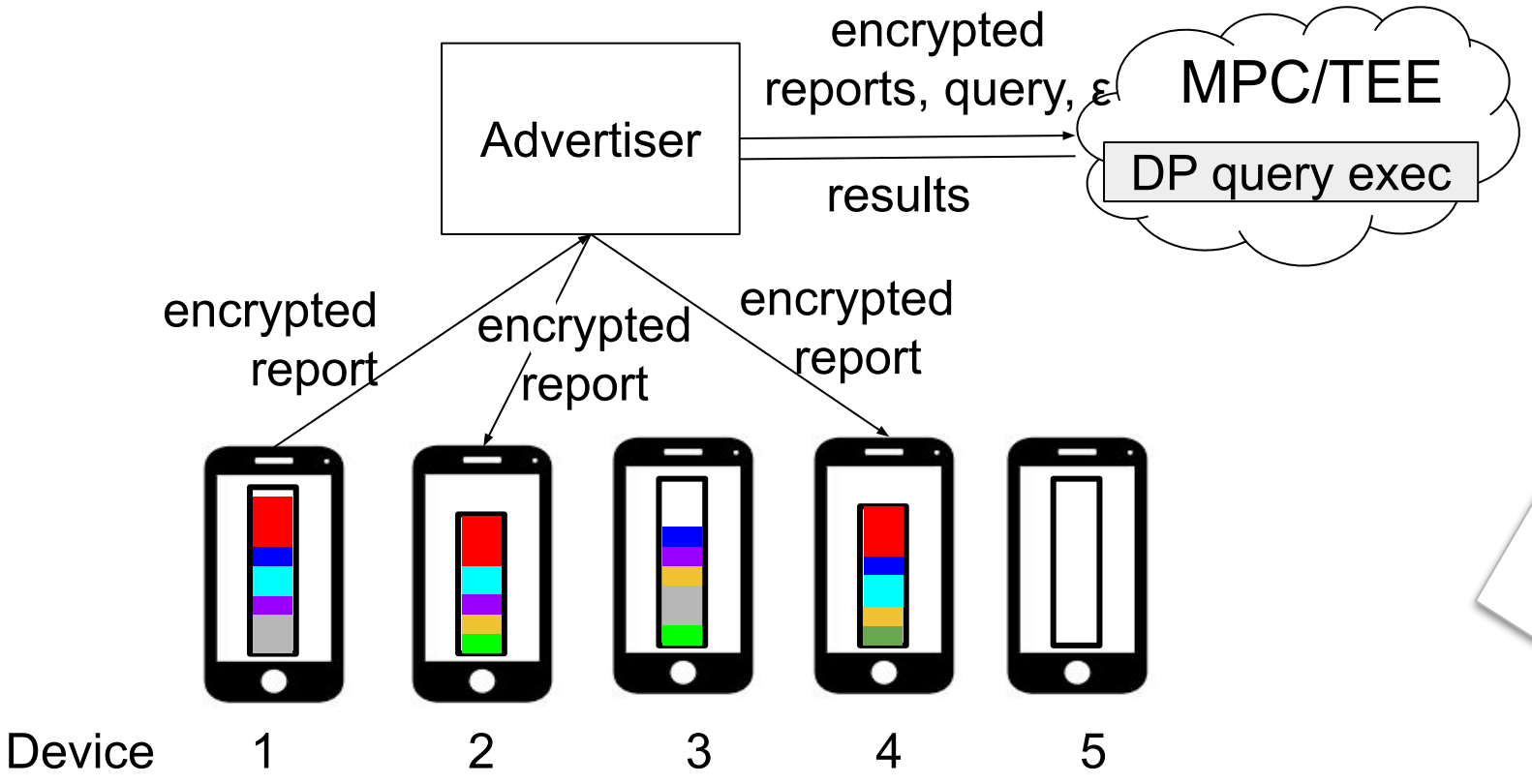
- **Global sensitivity, Δ**: maximum change in output that's possible by adding **any record r** to any database.
- Given noise scale σ, we deduct privacy loss proportional to **Δ/σ** from the system-wide budget.

IDP

- **Individual sensitivity, Δ(r)**: maximum change in output that's possible by adding **a given record r** to any database.
- Given noise scale σ, for each device r, we deduct privacy loss proportional to **Δ(r)/σ** from the device's own budget.

How IDP Justifies Peculiar Behavior

Individual sensitivity justifies deducting zero for devices without a conversion.



- Devices 1,2,4 consume budget because their users had the conversion.
- Devices 3,5 did not consume budget b/c their users did not have the conversion.
- **In IDP, devices 3,5 have Δ(r)=0, so makes sense to deduct Δ(r)/σ = 0 from their budgets!**

Implications of IDP

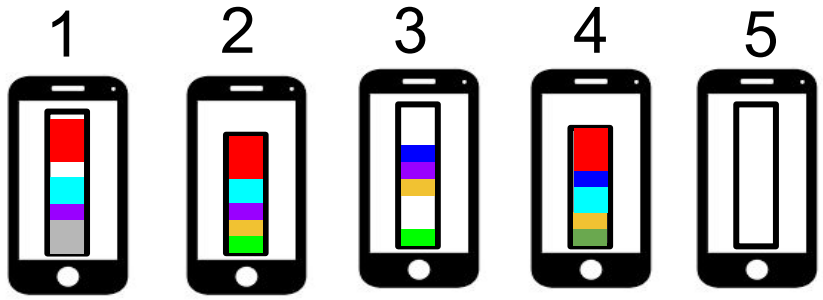
- Positive
- Negative

Implications of IDP

- **Positive**
- Negative

More Efficient Budgeting

- Per-device is more fine-grained, so more efficient, than global budgeting.
- Implicit use of individual sensitivity for devices without a conversion is already an IDP optimization.
- But IDP enables broader optimizations, since individual sensitivity is often lower than global sensitivity.



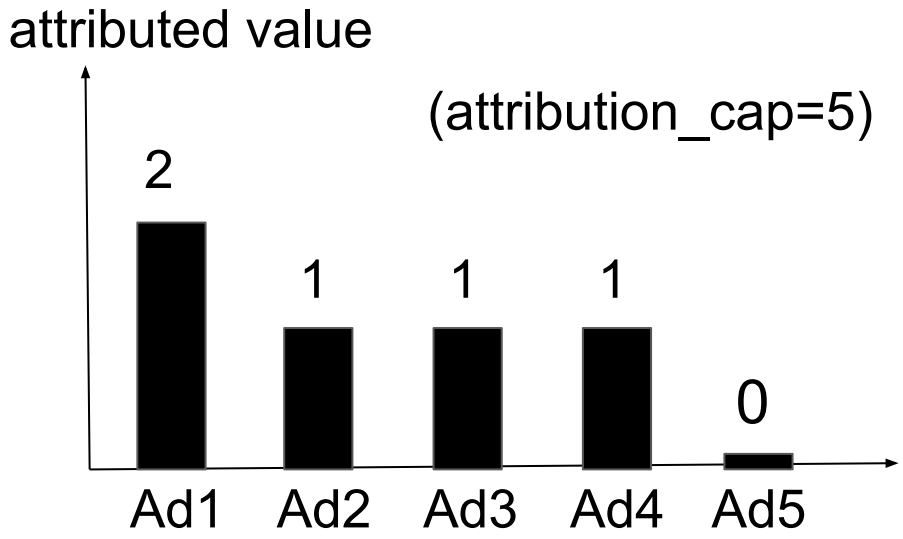
- Devices 3,5 use individual sensitivity.
- But devices 1,2,4 use global sensitivity.
  Why not use individual if smaller?

Examples of Individual < Global Sensitivity

Setting

- Conversion occurs, advertiser requests attribution histogram:
  - ads={Ad1,..,Ad5}
  - attribution_logic
  - attribution_cap=5
  - ε
- Assume advertiser's query will be a summation of attribution histograms from users with that conversion.
- Assume MPC/TEE uses Laplace(σ).

Example attribution histogram:



Examples of Individual < Global Sensitivity

Setting

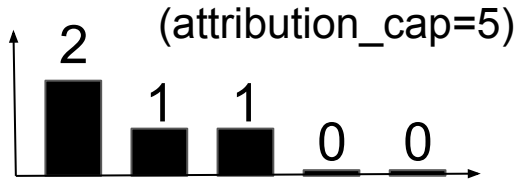
- Conversion occurs, advertiser requests attribution histogram:
  - ads={Ad1,..,Ad5}
  - attribution_logic
  - attribution_cap=5
  - ε
- Assume advertiser's query will be a summation of attribution histograms from users with that conversion.
- Assume MPC/TEE uses Laplace(σ).

Case 1: No impressions found

- Δ(r)=0
- Deduct zero privacy loss

Case 2: L1 of attribution histogram < attribution_cap

- Δ(r)=~3
- Deduct loss based on 3/σ, which is lower than 5/σ



Case 3: For user-epoch IDP, when performing attribution across epochs, deduct zero from epochs with no relevant ads.

Case 4: When performing attribution across epochs, ignore out-of-budget epochs and run over epochs w/ budget.

We're investigating other cases for specific attribution logics.

Implications of IDP

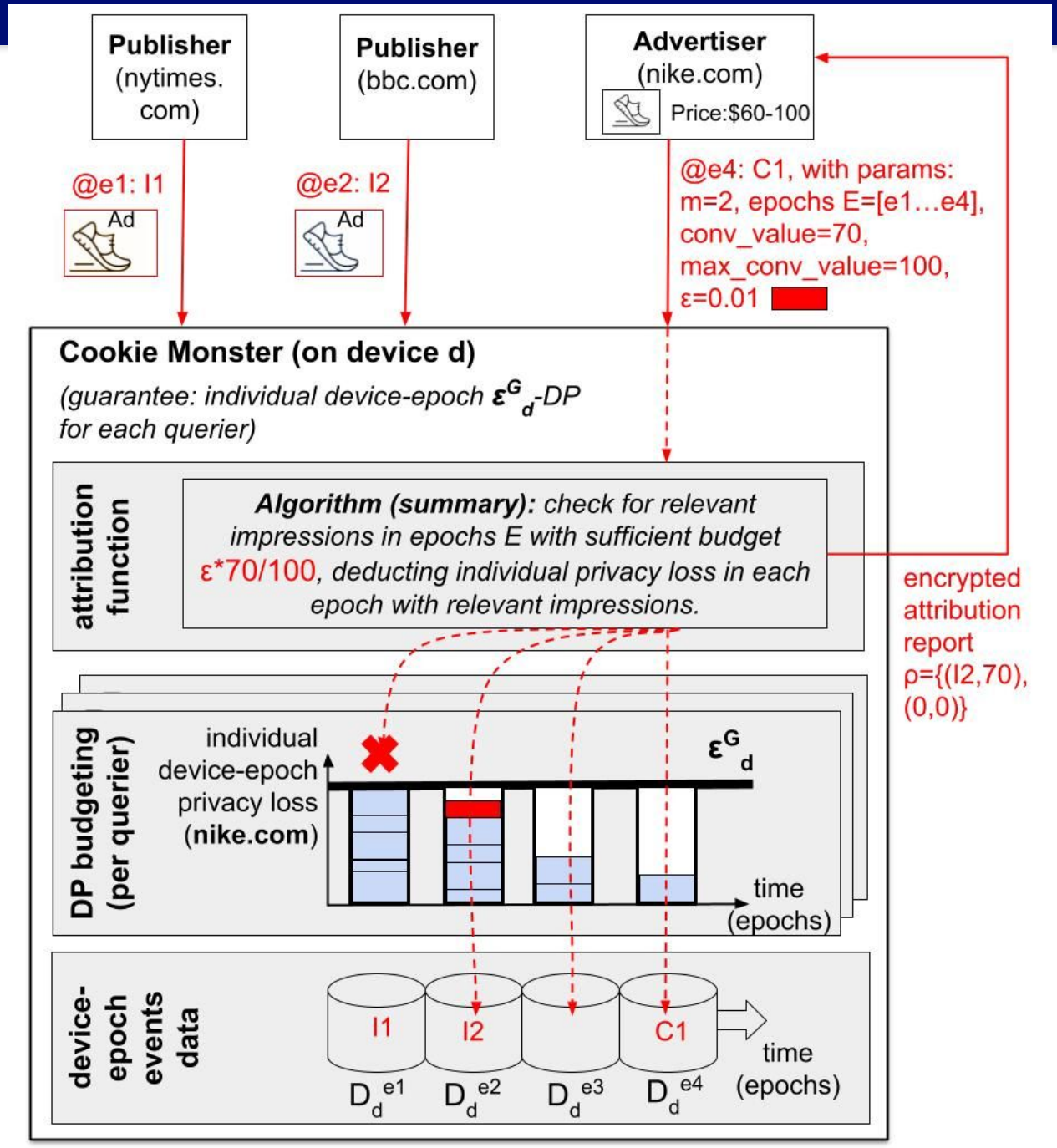
- Positive
- **Negative**

Must Keep Privacy Budgets Private!

- Individual sensitivity depends on data, so privacy loss depends on data, which means it must be **treated as data** and protected as such.
- This implies two things:
  1) Devices must **always answer** to report requests with a default value of individual sensitivity zero for the expected class of queries.
  2) We need to incorporate mechanisms into ad-measurement infra to help sites run reliable measurements while blind w.r.t. which report is real vs. default.
- Multiple mechanisms exist in the literature for 2):
  - Having the query executor drop reports w/o budget.
  - Regard available budgets as *data* and let sites query them in aggregate with DP, to understand what fraction of their users has the necessary budgets.
  - Designing mechanisms that work well for our setting is on our TODOs.

Cookie Monster: system design and prototype in ARA that we are developing at Columbia for DP budgeting component for on-device attribution systems

Design, development and evaluation are very much ongoing, let me know if you're interested in participating



Technical Take-Aways

1. Traditional DP is not a good fit for on-device budgeting systems.
2. Individual DP (IDP) is better suited and can enable optimizations for more efficient budget management in these systems.
3. But IDP also brings negative consequences, such as the need to keep the privacy budgets private.

The End