Navigating Clean Room and Privacy Enhancing Solutions

for AdTech and MarTech

Contents

An informal framework for navigating and evaluating emerging clean room and privacy-enhancing solutions

Ad-/MarTech Workflow
Data Enrichment

- Use two variants of common enrichment workflow to illustrate concepts
- Same principles will apply to many other workflows: onboarding, identity resolution, segmentation, lookalikes, attribution, real-time ad delivery, etc.

1st Party
Publisher or Intermediary

Email	Address
ann@example.com	123 A St. 1234
bob@example.com	456 B St. 12345
⋮	

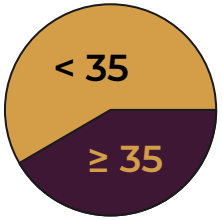
- Prefers to disclose as little data as possible to 2nd party
- Expects to receive agreed-upon output as part of transaction

2nd Party
Identity Graph Vendor

Email	Age	Income
ann@example.com	48	> $50K
bob@example.com	31	≤ $50K
⋮		

- Prefers exposure to as little data as possible from 1st party
- Wants to disclose only data that has business value to 1st party

Aggregate Output (for 1st Party)



Row-Level Output (for 1st Party)

Email	Age
ann@example.com	48
bob@example.com	31
⋮	

Clean Room Solutions
What Kind of Implementation?

- Preferences/requirements presented on previous slide **can be fully satisfied** in multiple ways
- Solutions grouped by **extent of disclosure to a 3rd party**

Trusted 3rd Party	Untrusted 3rd Party	No 3rd Party
SaaS platform and multi-sided data marketplace		Full integration with 2nd party infrastructure
Scalable, proprietary workflows enabled by platform		Support proprietary 2nd party workflows/capabilities
Secure data storage (fully accessible by 3rd party) • Governance & access controls (primarily for protecting data from 1st/2nd party, not 3rd party)	Secure data storage (limited or no access by 3rd party) • Governance & access controls • Non-cryptographic techniques • Cryptographic techniques	Secure joint computation (limited or no access by other party) • ~~Governance & access controls~~ • Non-cryptographic techniques • Cryptographic techniques

Extent of Disclosure to 3rd Party
What Informs Your Decision?

- One purpose of data protection technologies is to limit disclosure of data to the (newly introduced) 3rd party
- Your organization's goals w.r.t. this informs solution choice

- Is it acceptable to disclose data to a 3rd party?
  - If yes, trusted 3rd party may be sufficient
  - If not, costs of other two variants may be justified
- What disclosure is acceptable when considering…
  - Competitive advantage
  - Cybersecurity considerations
  - Regulatory compliance (current and future)
  - Internal policies
  - Legal liability
- How does this interact with non-privacy solution/vendor features?
  - Multi-sided marketplace/network participation
  - Scalability and cost
  - Standardization and stability
  - Proprietary capabilities offered by 3rd party

Clean Room Solutions
How/Where is Data Disclosed?

- Two disclosure points: to 3rd party and to intended recipient(s)
- Any individual technology might be used to limit disclosure to 3rd party, recipient(s), or both: technology ≠ purpose in use case

Not Disclosing Input and Intermediate Data (to 3rd or any party)

Trusted 3rd Party

- Encrypted data **(with key)** stored by 3rd party
- Governance, access controls, & auditability

Untrusted 3rd Party

- Encrypted data **(no key)** stored by 3rd party
- Minimization & redaction
- De-identification
- Pseudonymization
- Noise introduction

"Non-cryptographic"

No 3rd Party

- Encrypted data **(no key)** jointly processed by 1st and 2nd party
- Traditional hashing & tokenization
- Trusted Execution Environments (TEEs)
- Secure Multi-Party Computation (MPC)
- Homomorphic Encryption (HE)

"Cryptographic"

Disclosing as Little as Possible in Output Data (to 1st party)

- Minimization & redaction
- De-identification
- Pseudonymization
- Noise introduction
- K-anonymity
- Cohorts and clusters
- Data models
- Differential Privacy (DP)

Privacy-Enhancing Technologies
Maturity, Characteristics, & Costs

- Brief comparison of a traditional technique and PETs
- Some dimensions to consider: scalability/cost, maturity, longevity ("future-proof"-ness), accuracy/precision

- Traditional cryptographic hashing and tokenization
  - Susceptible to guess-and-check, persistent, impossible to federate/convert
- Non-cryptographic techniques (de-identification, anonymization, noise)
  - Highly scalable: no significant costs compared to in-the-clear
  - Not necessarily future-proof: difficult to quantify privacy benefits
- Techniques intended for aggregate results (cohorts, models, DP)
  - Highly scalable: no significant costs compared to in-the-clear
  - Not necessarily future-proof: difficult to quantify privacy benefits
  - Reduced Precision/accuracy of results
  - Sometimes (mis)used as less costly substitute to cryptographic techniques...
- Cryptographic techniques
  - TEEs: Least costly, most scalable (appropriate for AI/ML); relies on hardware
  - MPC: Scalability depends on workflow; future-proof
  - HE: Least scalable; secure/general like MPC but can run offline; future-proof

Example Scenario Variants
Data Enrichment

- Depending on what kind of data disclosure is acceptable, different PET solutions can be employed
- Other features enabled/prevented by choices may be a factor, as well

1. Publisher and ID graph vendor send data to 3rd party; 3rd party delivers results to publisher

- Both sides trust 3rd party
- Publisher and ID graph owner pay for 3rd party platform use by volume of data
- Workflow runs very quickly

2. ID graph vendor offers MPC-based portal that lets publisher run enrichment workflow

- No 3rd party involved, so no 3rd party fees
- Publisher's data never disclosed, and only intended results are disclosed to publisher
- Workflow takes ~60 minutes for 300M rows

3. Publisher and ID graph vendor build noisy data via 3rd party; both receive the results

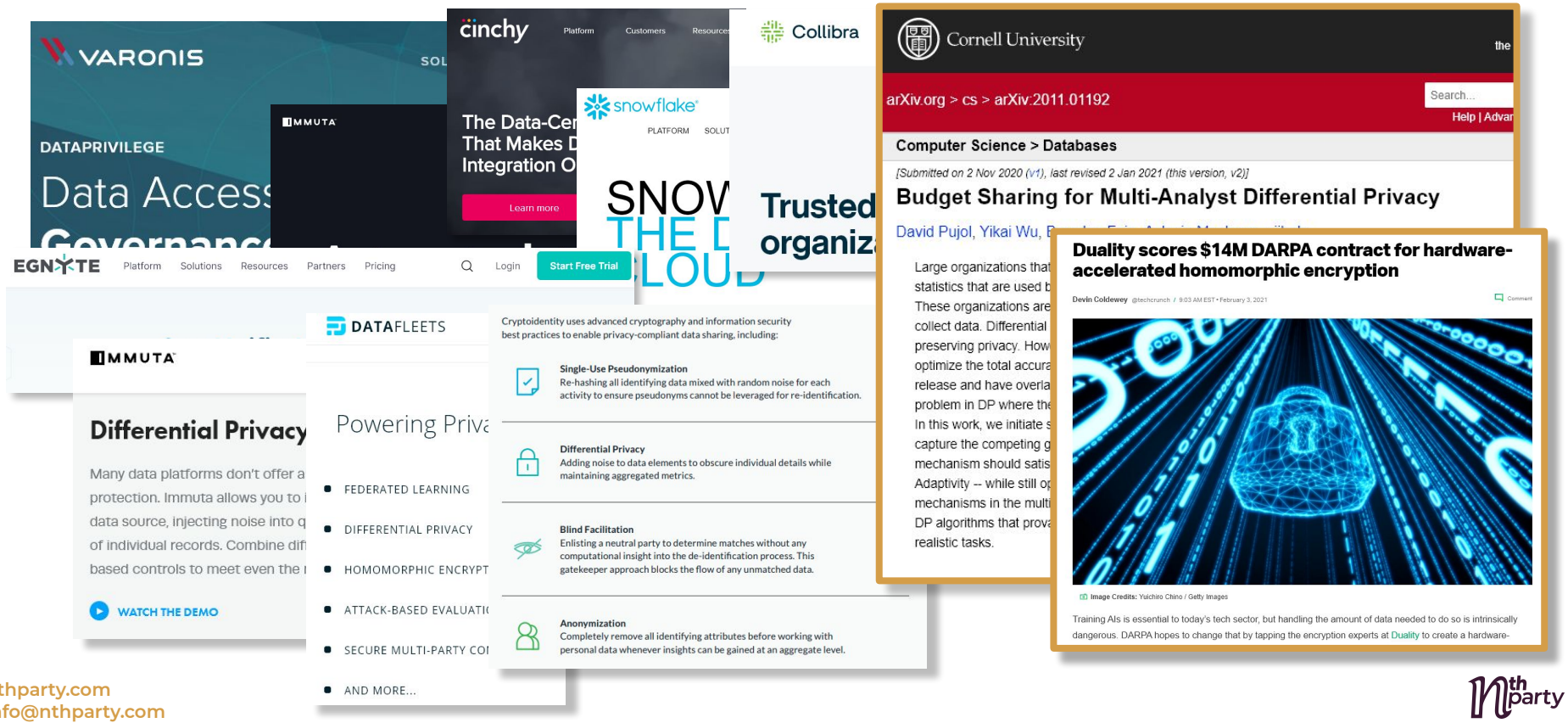
- 3rd party does not store raw data, but briefly sees it to build noisy model
- Results have an error margin of 2-3%
- Workflow takes ~1 minute for 300M rows

4. Publisher and ID graph vendor run workflow on 3rd party's TEE cloud infrastructure

- 3rd party sees no data, charges by volume
- Publisher's data never disclosed, and only intended results are disclosed to publisher
- Workflow takes ~5 minutes for 300M rows

Evaluating & Selecting
Other Considerations

- Noisy landscape: vendors list every PET, but many cannot offer them all
- Important to ascertain for what purpose each feature is present: how does it help you avoid disclosure and to which parties?



Privacy-Enhancing Technologies
Further Reading

- Resources below mostly focus on current/future state of PETs
- Vendors cannot decide on their own which PET solves your data non-disclosure needs; progress will involve collaboration

- "What Are Privacy-Enhancing Technologies (PETs) and How Will They Apply to Ads?" Facebook. https://about.fb.com/news/2021/08/privacy-enhancing-technologies-and-ads/
- "A guide to Privacy Enhancing Technologies (PETs) & how to adopt them!" Gilbert + Tobin via Lexology. https://www.lexology.com/library/detail.aspx?g=79a577a0-a4b7-409c-aca6-cef77154ffa9
- "Privacy Enhancing Technologies Adoption Guide". Centre for Data Ethics and Innovation. https://gov.uk/cdei
- "The Age of Collaborative Computing". Lunar Ventures via Medium. https://medium.com/lunar-ventures/the-age-of-collaborative-computing-e73374b7aedc