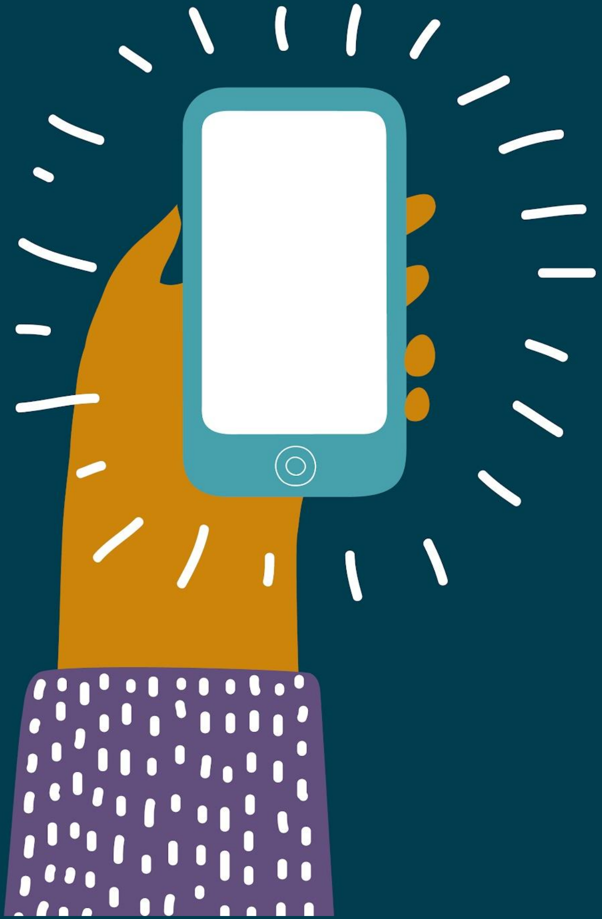
Practical Data Privacy

An Introduction to Privacy Technology futures

/thoughtworks

Privacy Enhancing Technologies: From Labs to Reality



thoughtworks

What we’ll cover today

Practical Applications of Privacy Technology

In this talk, we’ll walk through some plausible use cases and review how we solve privacy problems now and what the future might look like if developers and data scientists embraced privacy enhancing technologies.

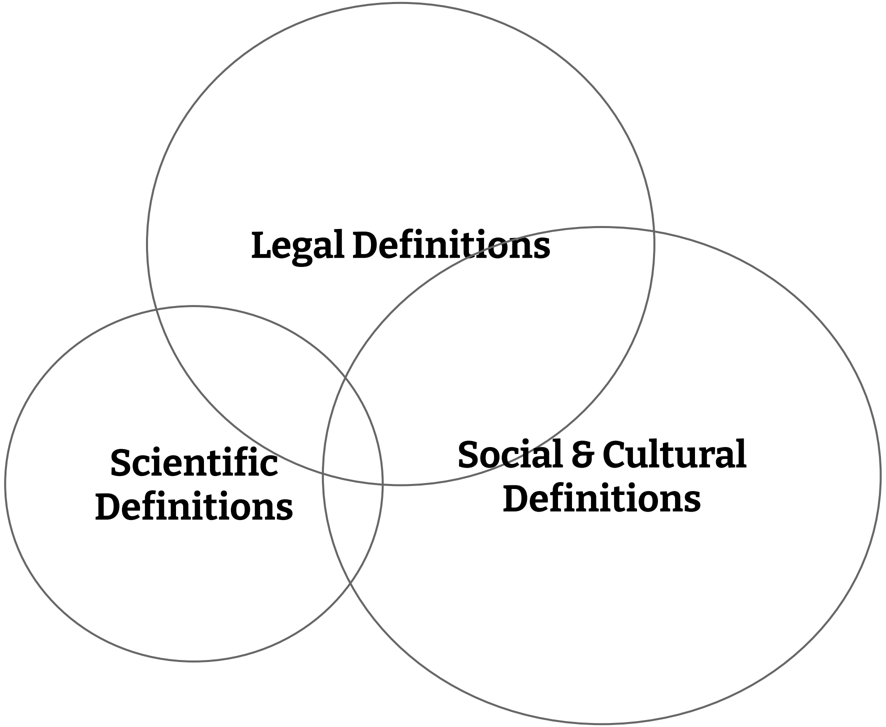
Real-world use cases for privacy technology

Why privacy engineering?

Applicable areas for privacy technology

Ways to learn more

What even is privacy?



Legal Definitions

Scientific Definitions

Social & Cultural Definitions

Why now?

Privacy Engineering: A growing field!

In the ever-changing regulatory landscape, data privacy is taking a more central role than ever. Finding new and improved ways to manage data can mean that we keep working on important problems while also reducing risk!

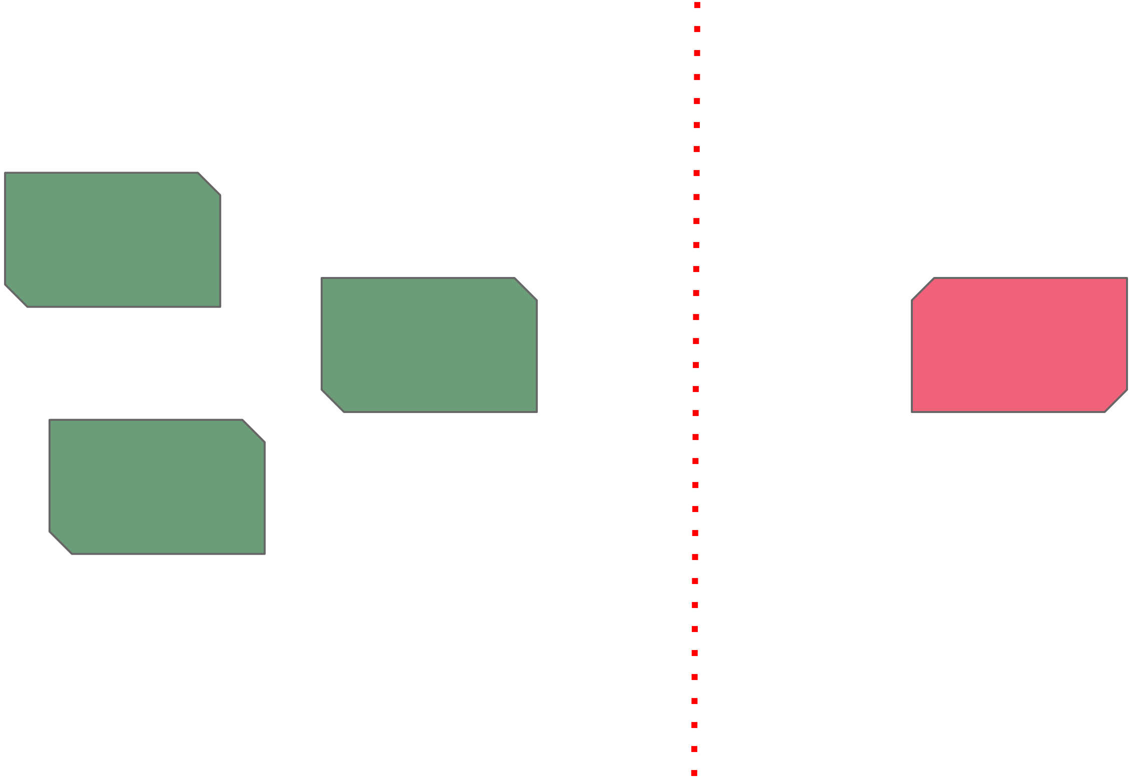
The field of privacy engineering is growing quickly – and could be a potential shift should this talk interest you!

- **Increased data → Increased risk**
- **New & changing data regulations**
- **Consumer demand**
- **Technological advances**
- **Social justice**
- **Data benefits**

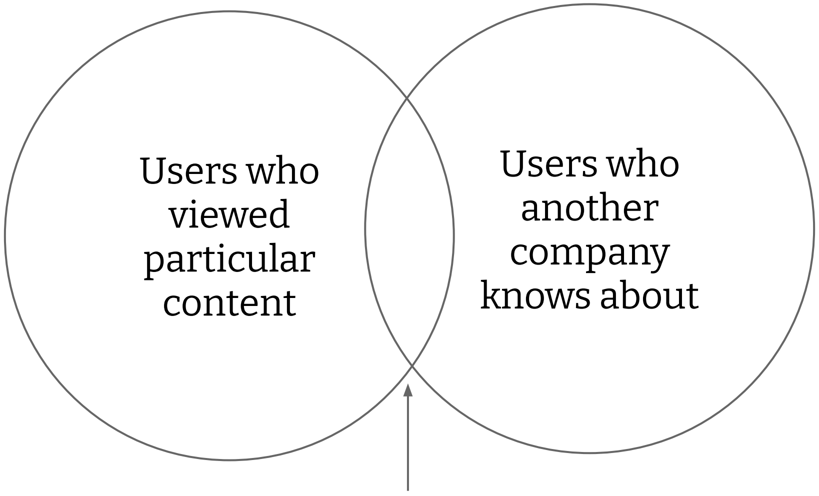
Federated Data Mesh: How can we find shared customers without sharing data?

thoughtworks

Problem Statement



Desired Outcome



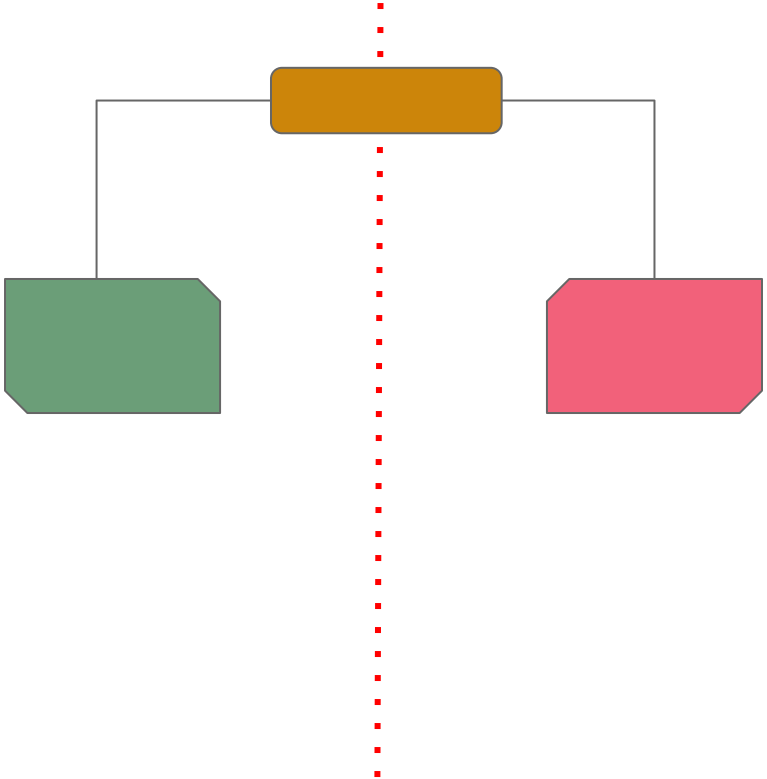
Users who viewed particular content

Users who another company knows about

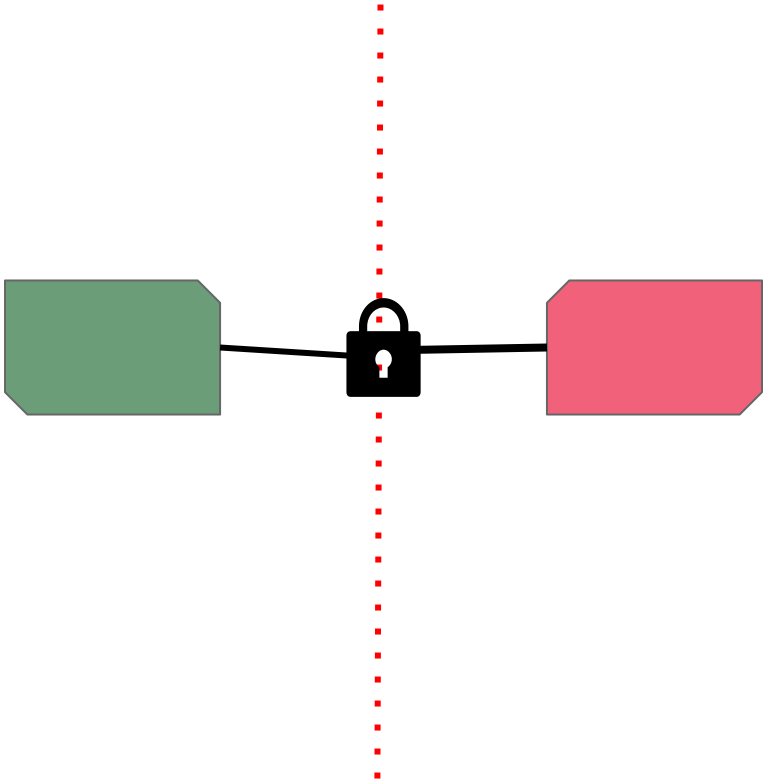
Users we want to find

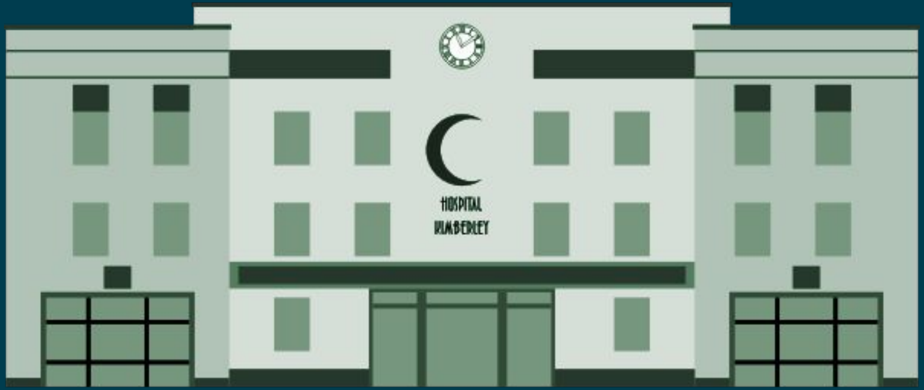
© 2022 Thoughtworks

Current Solution



Future Solution: Private Set Intersection





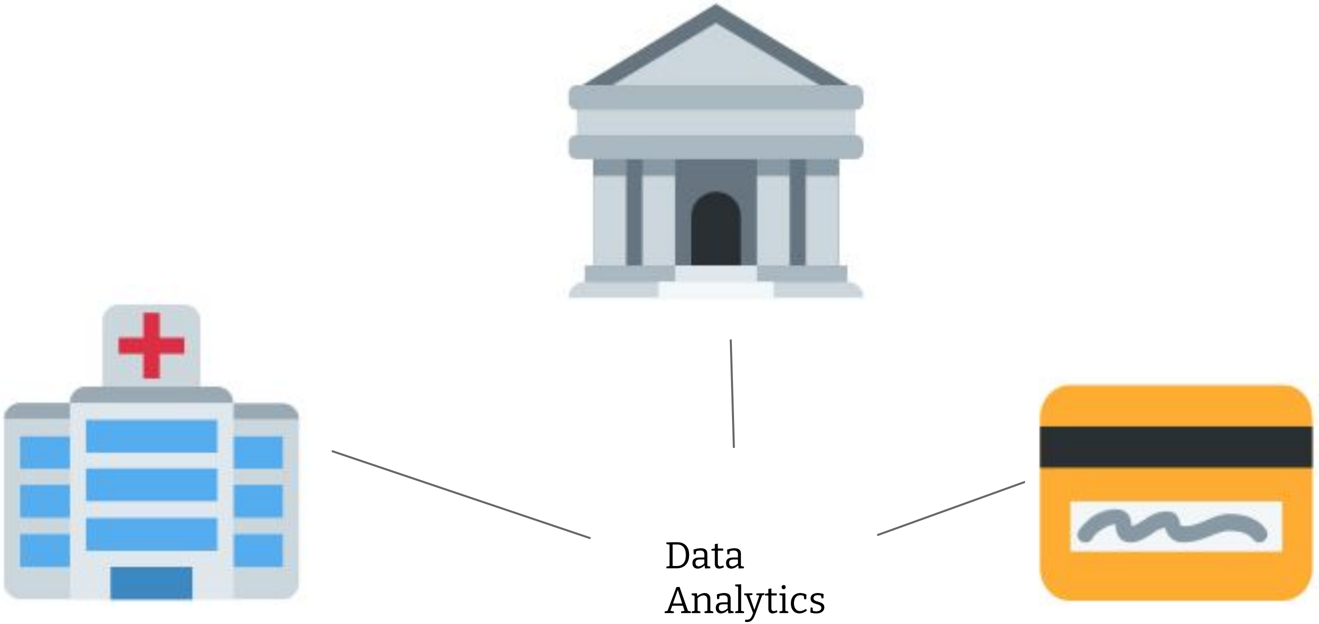
Shared Sensitive Data Computations: How do we estimate costs of public and private health services?

/thoughtworks

Problem Statement

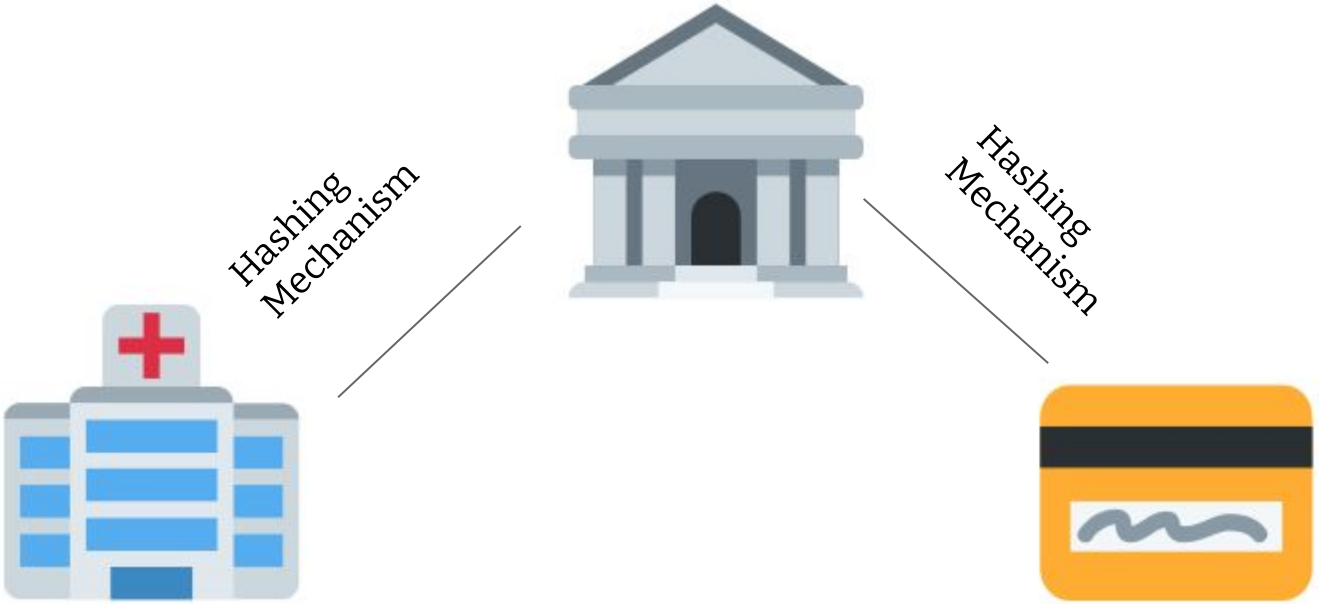
© 2022 Thoughtworks

Desired Outcome

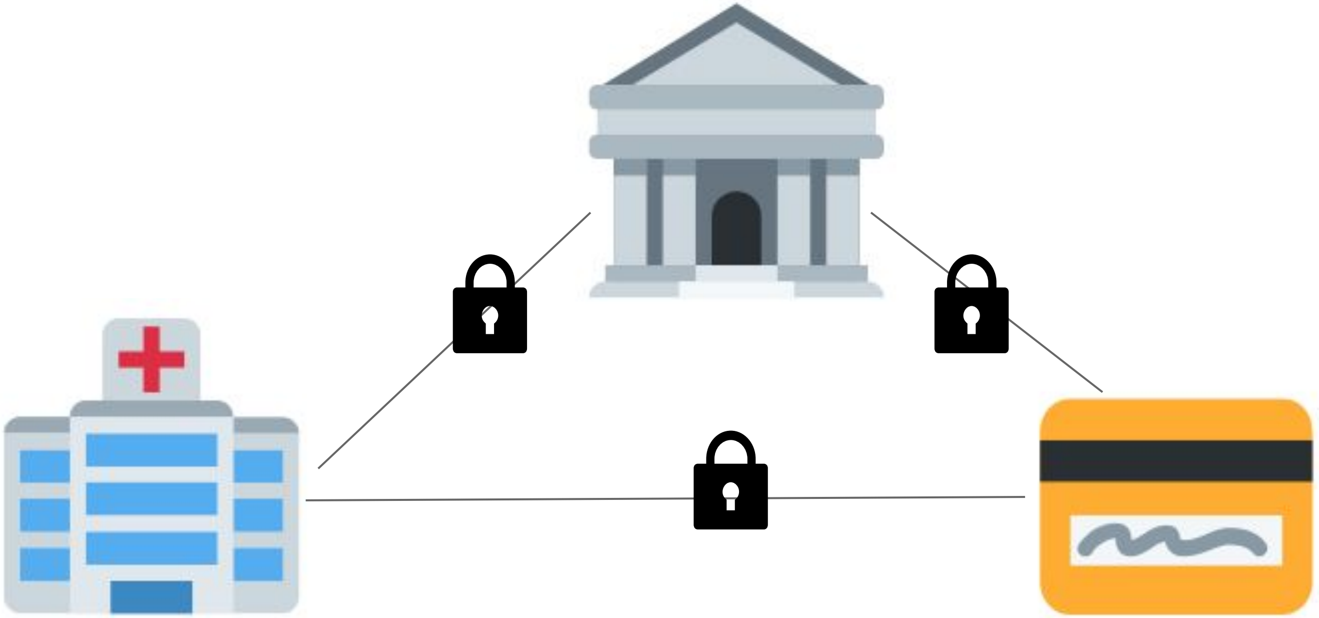


Data Analytics

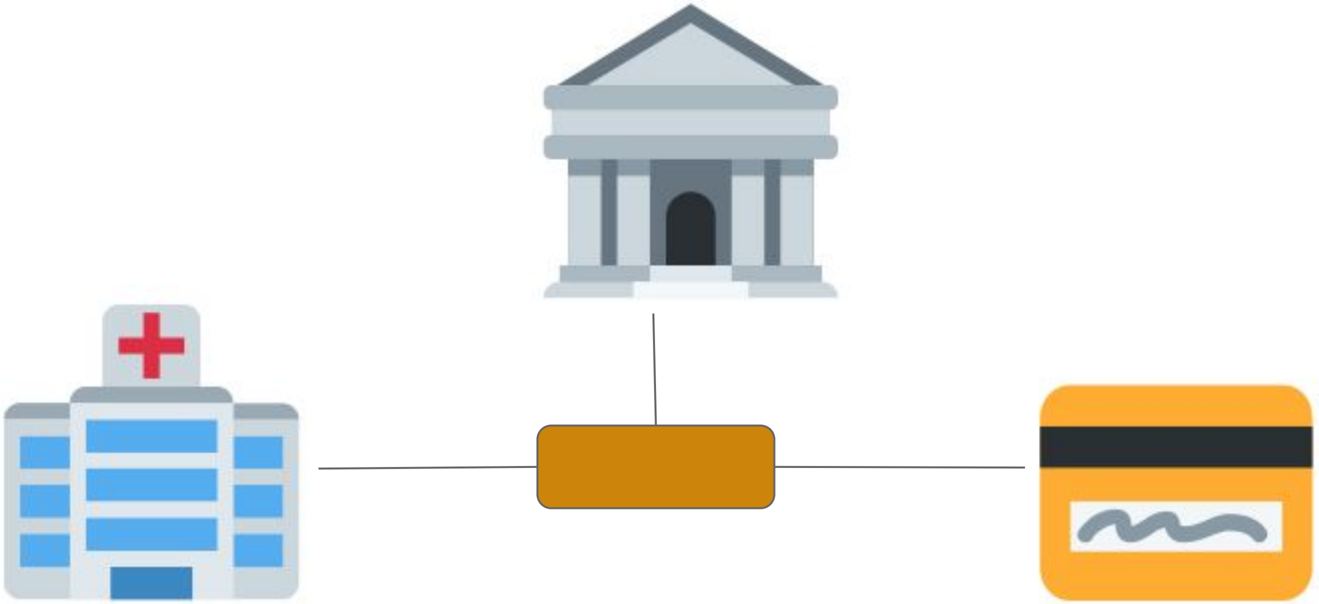
Current Solution



Future Solution: Multi-Party Computation



Future Solution: Federated Analytics



Anonymized Machine Learning: GDPR Compliance in large-scale machine learning

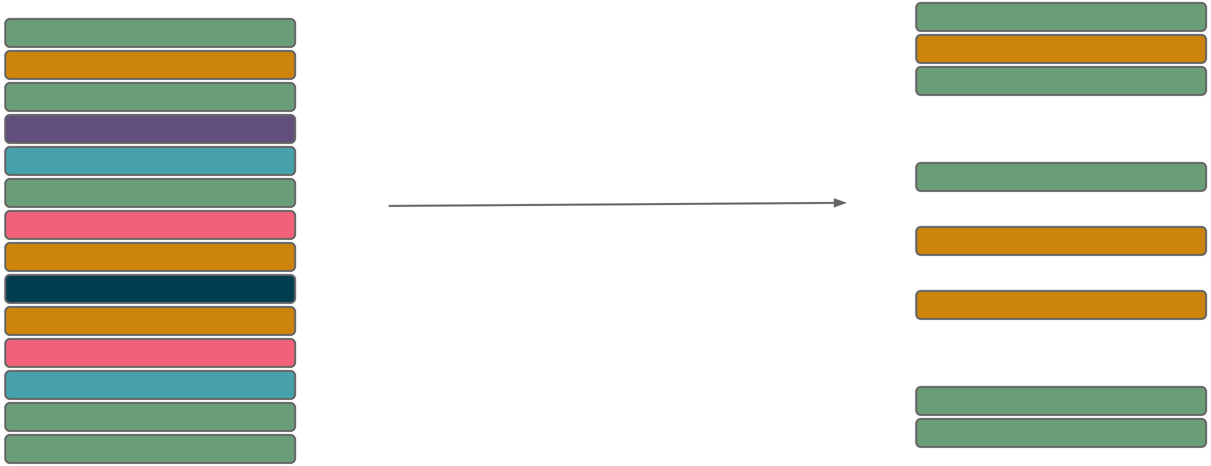
/thoughtworks

Problem Statement

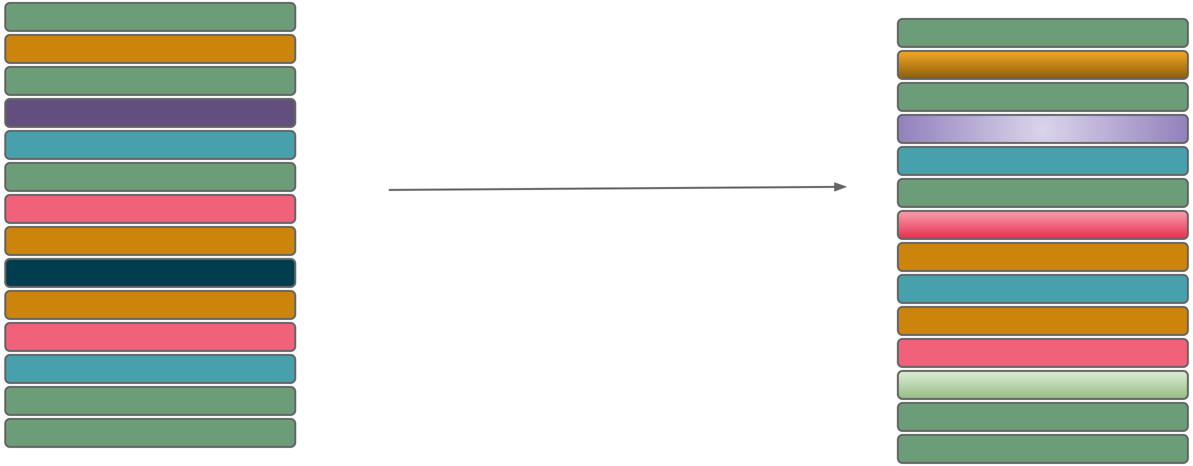
Desired Outcome

💳 – 💥

Current Solution: K-Anonymity

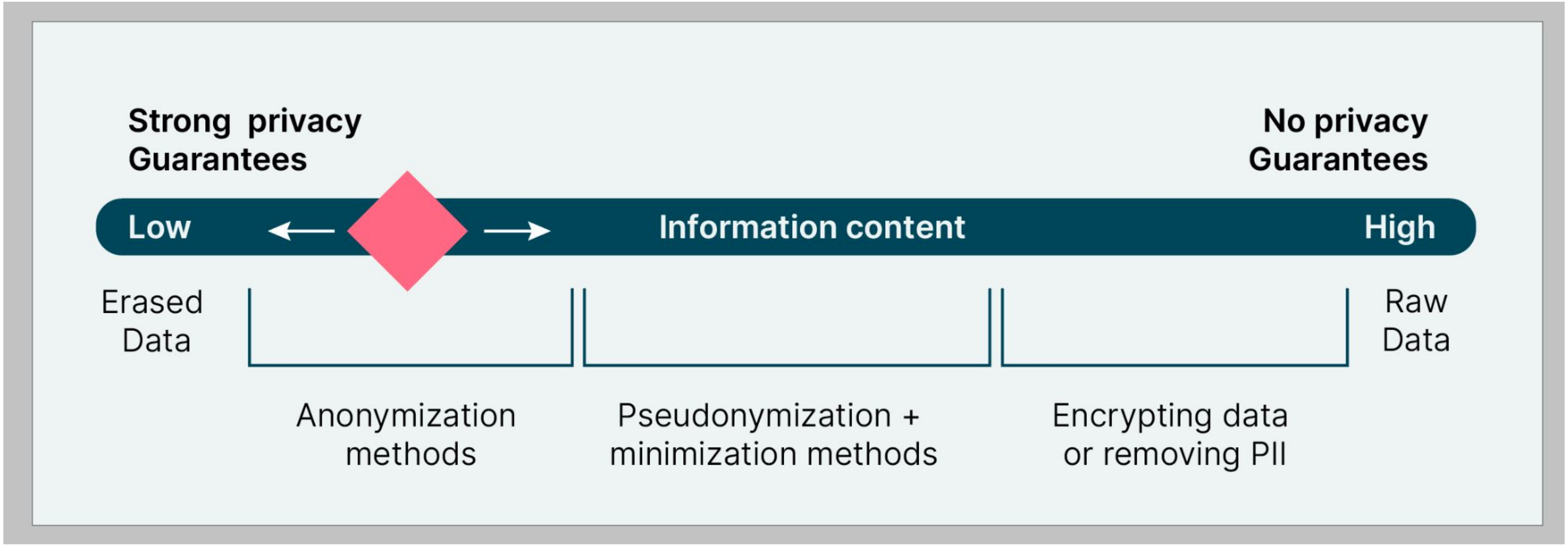


Future Solution: Differential Privacy



Privacy vs. Information Continuum

Thinking through the privacy vs. utility “tradeoff”


Strong privacy Guarantees
No privacy Guarantees
Low
Information content
High
Erased Data
Raw Data
Anonymization methods
Pseudonymization + minimization methods
Encrypting data or removing PII

Where can I use PETs?

Enabling safer and privacy-aware data usage

When working in government, healthcare or financial services, PETs are becoming not only more prevalent, but seen as a requirement to enter. By leveraging PETs in your work, you are introducing state-of-the-art privacy protection and often reducing the attack space for information security risk.

- Highly regulated industries
- Data sharing or collaboration
- Sensitive or proprietary data
- Anytime you handle person-related data!

How do I learn more?

Become a privacy engineer!

There is ***so much more*** that we didn’t have a chance to cover today. If your interest is sparked, please update your Summit goals and start learning now! The world needs many more privacy engineers.

1. Check out the references slide!

2. Read my new book!

3. Pick a technique to apply to your next project!

Questions? Thoughts? Please reach out!

Katharine Jarmul
Principal Data Scientist
katharine.jarmul@thoughtworks.com

@kjam on Twitter

/thoughtworks

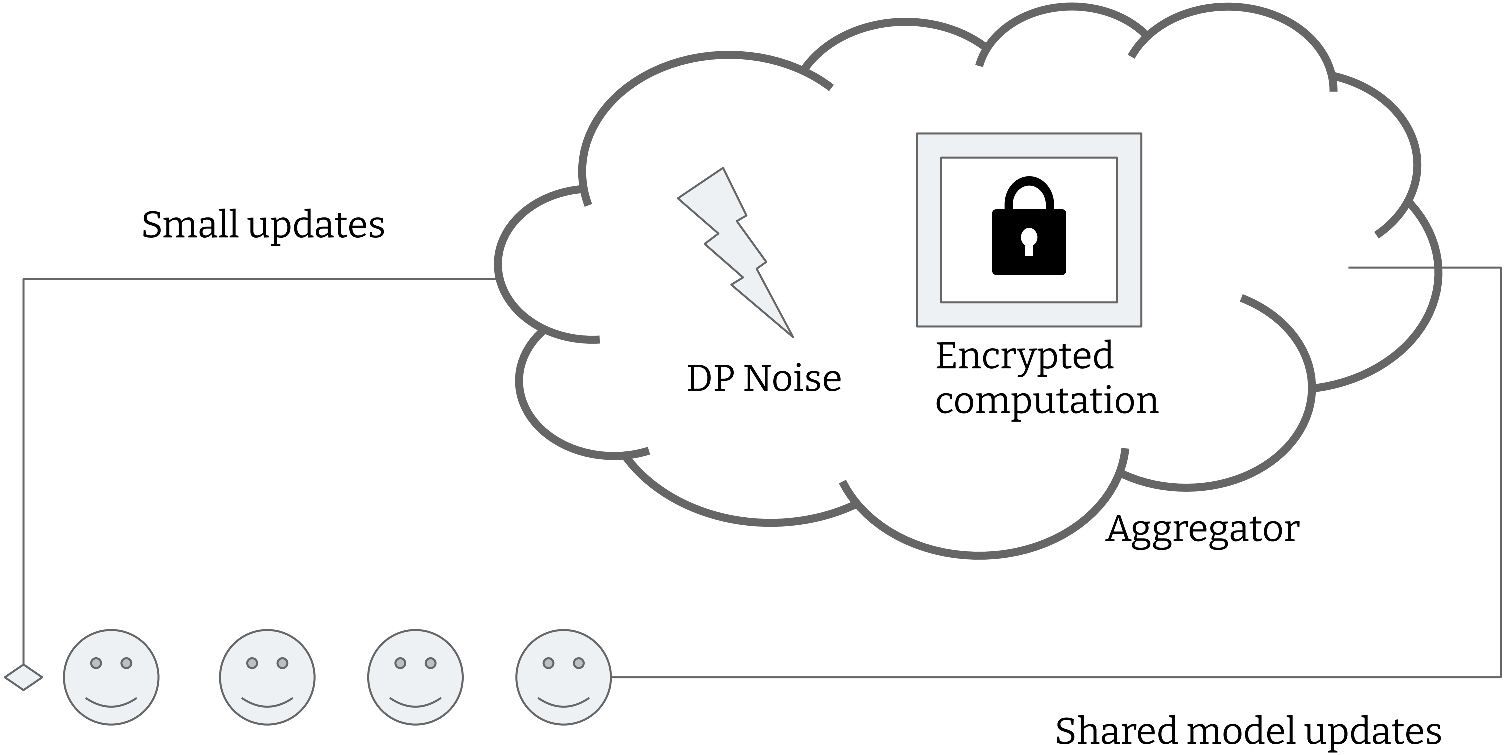
References

- Damien Desfontaines Differential Privacy blog series: https://desfontain.es/privacy/friendly-intro-to-differential-privacy.html
- Similar examples: Google Private Join & Compute: https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/private_join_and_compute.pdf
- Similar examples: NVIDIA's Clara: https://developer.nvidia.com/blog/federated-learning-clara/

Additional Learning

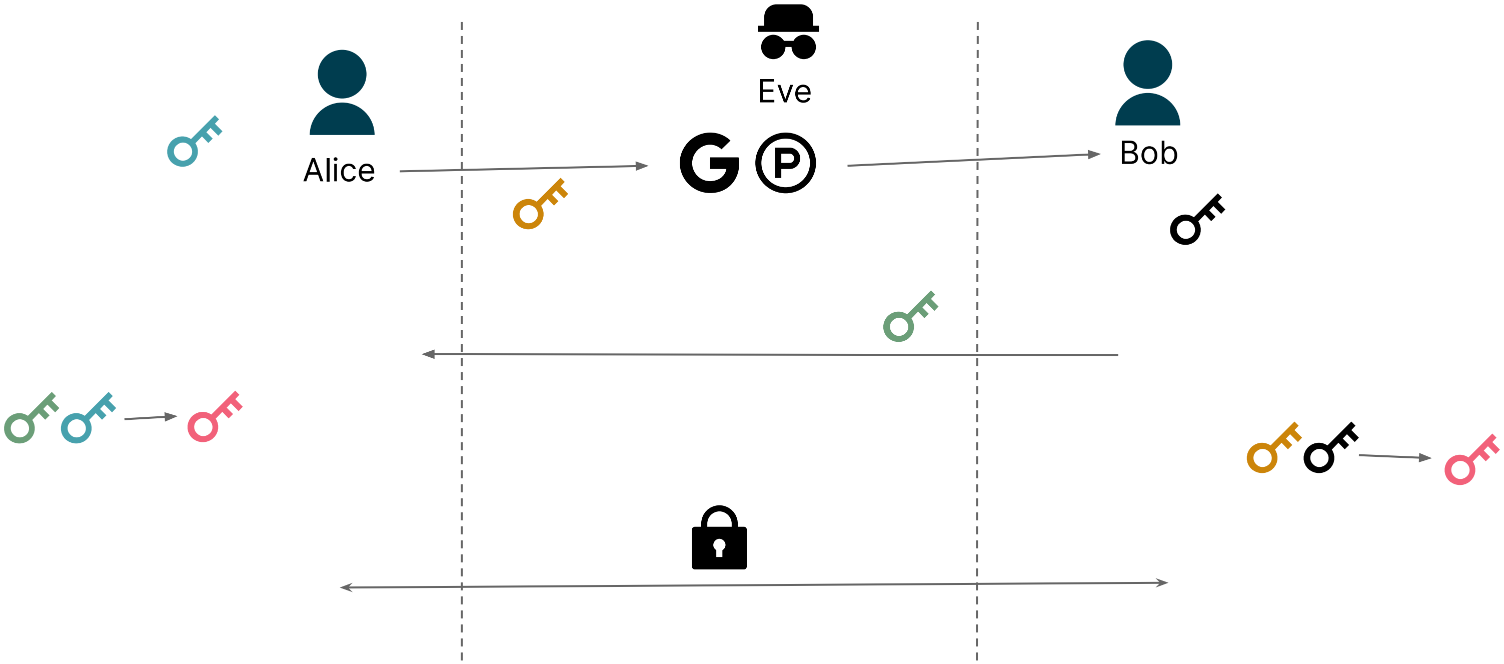
- Practical Data Privacy - Early Release: https://www.oreilly.com/library/view/practical-data-privacy/9781098129453/
- Foundations of Private Computation: https://courses.openmined.org/courses/foundations-of-private-computation
- Federated ML at the Edge Talk: https://www.infoq.com/news/2021/12/jarmul-ml-edge/
- Learn MPC: https://www.mpcalliance.org/learn

Appendix: Private & Secure Federated Learning



Appendix: Diffie-Hellman Key Exchange

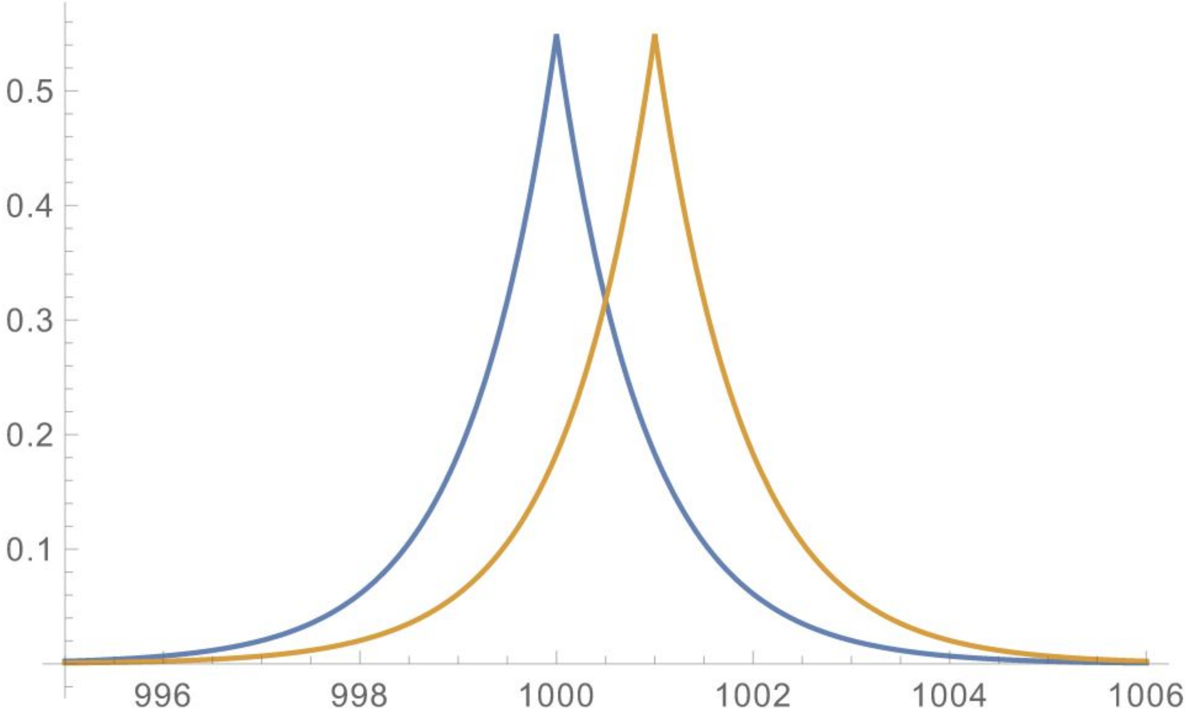
Finding Private Joins with Shared Keys



© 2022 Thoughtworks

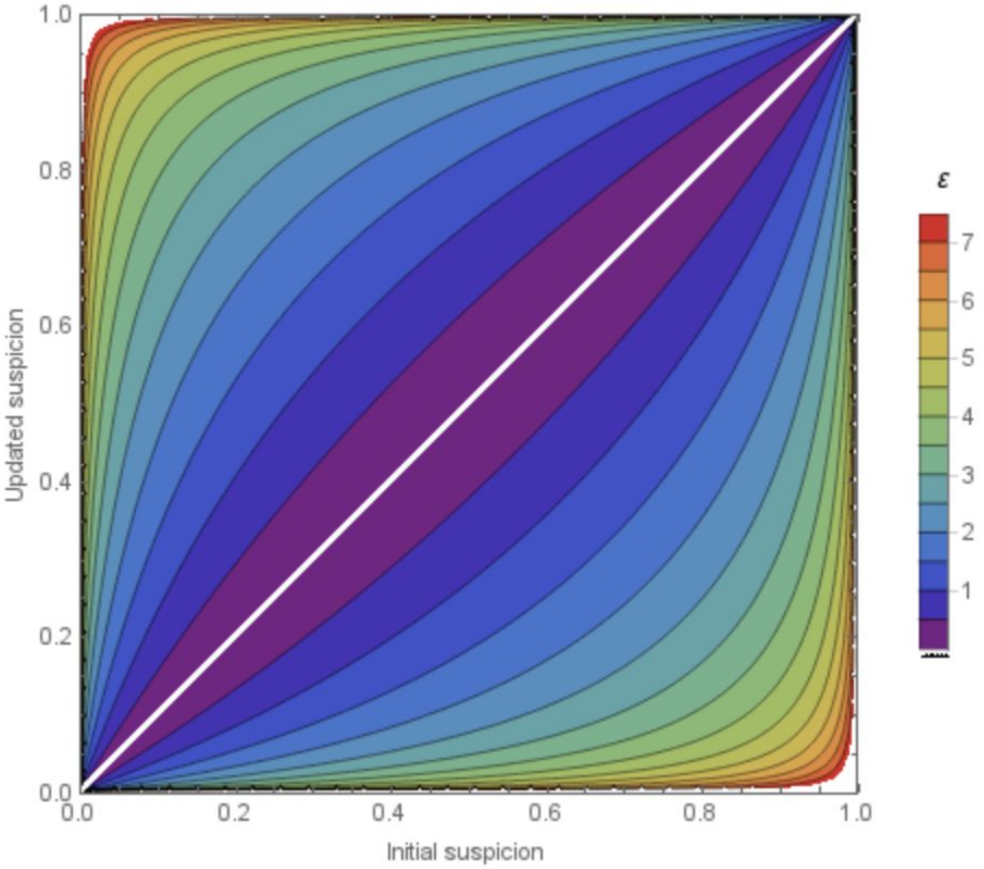
Appendix: What is Differential Privacy?

Building Intuition: Returning a Count: Is the real value 1000 or 1001? 🤔



Appendix: Bounding the Attacker's Information Gain

Differential privacy parameters allow us to bound the potential information gain based on a probability-driven attacker (here: Bayesian reasoning).


Updated suspicion
Initial suspicion
ε