Privacy and Security in Distributed Data Markets

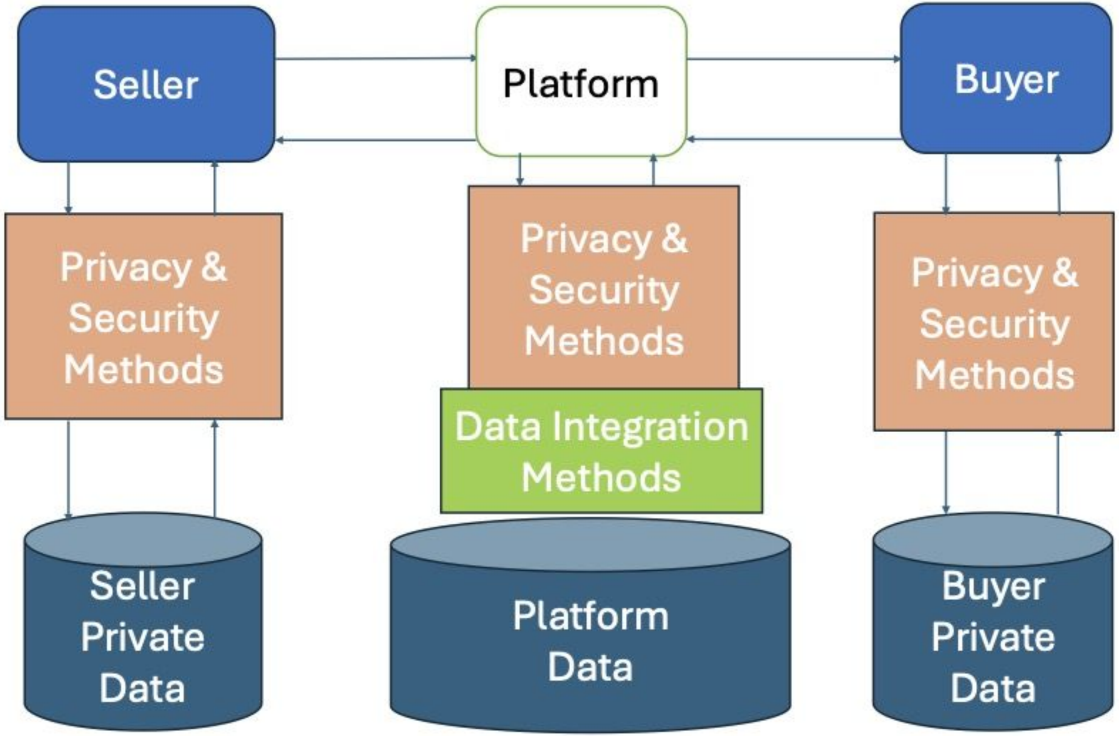
Daniel Alabi, Sainyam Galhotra, Shagufta Mehnaz, Zeyu Song, Eugene Wu

SIGMOD 2025 Tutorial

Part 2: Privacy and Security Risks

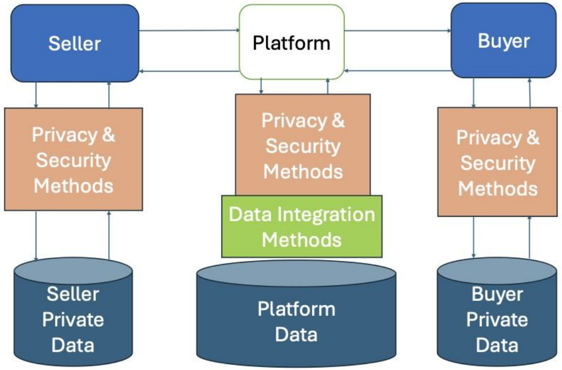
https://dacesresearch.org/tutorials/sigmod2025/

Protect Information in Data Markets



Protect Information in Data Markets

1. Protect buyers from *malicious* sellers
2. Protect sellers from *malicious* buyers
3. Prevent *unauthorized* users from accessing:
   a. Seller private data
   b. Buyer private data
   c. Platform private data
4. Prevent manipulation of data acquisition mechanisms:
   a. Data discovery
   b. Data valuation
   c. Data negotiation
   d. Data delivery



Privacy and Security Attacks

- Naively allowing query access to data markets is risky for users/orgs
  - Linkage attacks
  - Reconstruction attacks
  - Inference attacks
  - Plaintext/ciphertext attacks
- Naive designs of data markets is risky for valuation
  - Manipulation of pricing and negotiation mechanisms
  - Less trust in data markets

Motivates the need for robust *privacy and security protections*

Privacy and Security Attacks

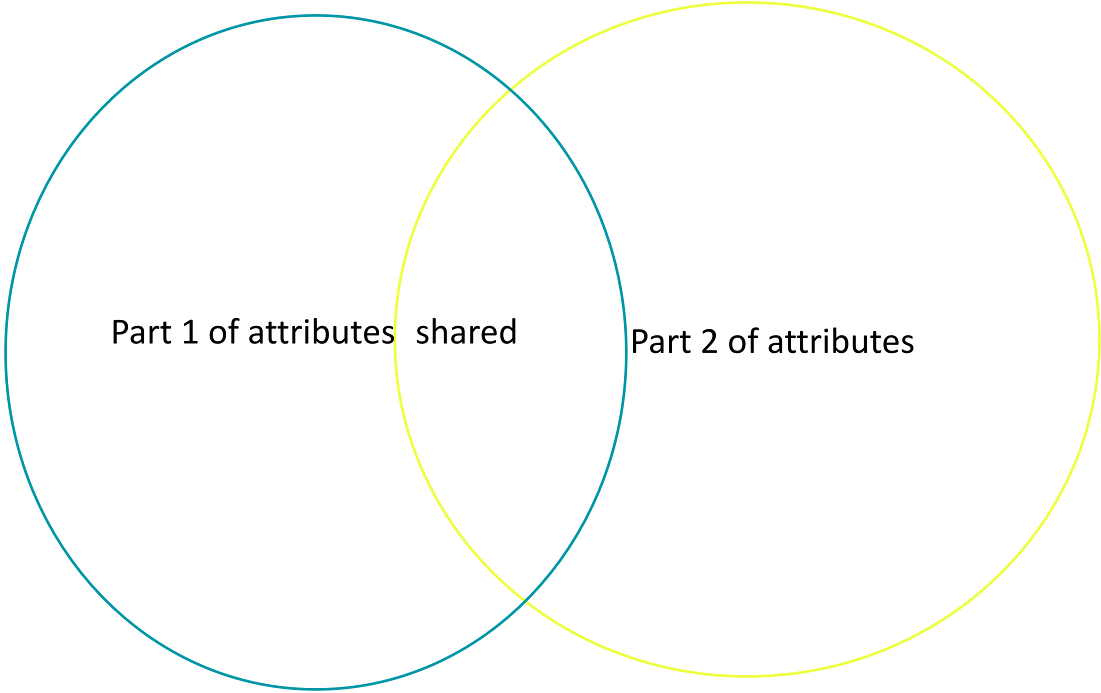
- Naively allowing query access to data markets is risky for users/orgs
  - **Linkage attacks**
  - Reconstruction attacks
  - Inference attacks
  - Plaintext/ciphertext attacks
- Naive designs of data markets is risky for valuation
  - Manipulation of pricing and negotiation mechanisms
  - Less trust in data markets

Motivates the need for robust *privacy and security protections*

Linkage Attacks

Perform join on one or more datasets

Can uniquely identify individuals



De-identification attempt

"Anonymize the Data": Are we happy with this solution? Why or why not?

Name	Sex	Blood	…	HIV?
James	M	B	…	N
Peter	M	O	…	Y
…	…	…	…	…
Paul	M	A	…	N
Eve	F	B	…	Y

Name	Sex	Blood	…	HIV?
XXXXX	M	B	…	N
XXXXX	M	O	…	Y
…	…	…	…	…
XXXXX	M	A	…	N
XXXXX	F	B	…	Y

De-identification attempt

“Anonymize the Data”: Not sufficient because of linkage attacks!

87% of US population (used to) have unique date of birth, gender, and postal code!

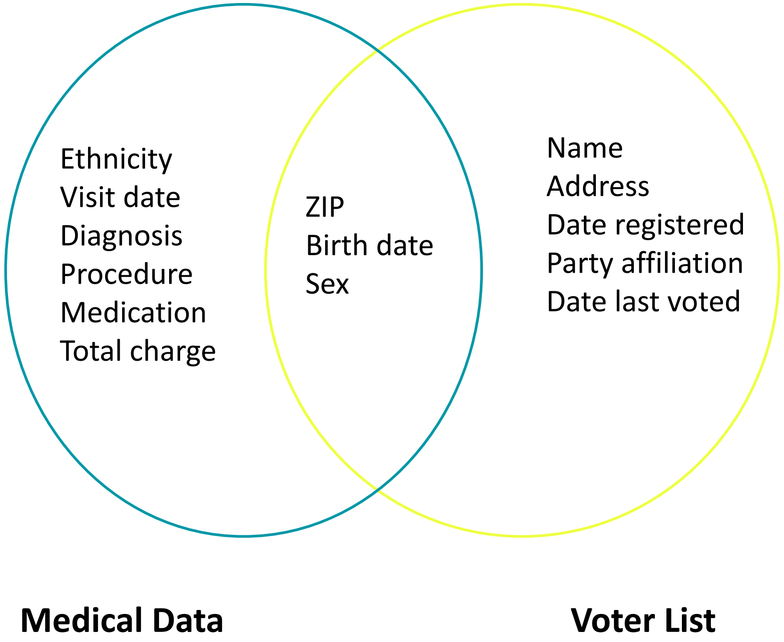
[Golle and Partridge ‘09]

De-identification attempt

“Anonymize the Data”: Reidentification via Linkage

Can uniquely identify > 60% of the U.S. population [Sweeney ’00, Golle ‘06, Sweeney ‘97]

Name	Sex	Blood	…	HIV?
XXXXX	M	B	…	N
XXXXX	M	O	…	Y
…	…	…	…	…
XXXXX	M	A	…	N
XXXXX	F	B	…	Y



Privacy and Security Attacks

- Naively allowing query access to data markets is risky for users/orgs
  - **Linkage attacks**
  - **Reconstruction attacks**
  - Inference attacks
  - Plaintext/ciphertext attacks
- Naive designs of data markets is risky for valuation
  - Manipulation of pricing and negotiation mechanisms
  - Less trust in data markets

Motivates the need for robust *privacy and security protections*

Reconstruction Attack

<u>Reconstruction attack</u>: If we have dataset $x \in \{0,1\}^n$ and person i has sensitive bit x_i and attacker/adversary gets $q_S(x) = \sum_{i \in S} x_i$ for $O(n)$ random $S \subseteq [n]$.

Reconstruction Attack

<u>Reconstruction attack</u>: If we have dataset $x \in \{0,1\}^n$ and person i has sensitive bit x_i and attacker/adversary gets $q_S(x) = \sum_{i \in S} x_i$ for $O(n)$ random $S \subseteq [n]$.

[Dinur-Nissim '03]: With high probability, adversary can reconstruct 0.99 fraction of the dataset $x \in \{0,1\}^n$ if noise added to each query is less than $o(\sqrt{n})$.

Privacy and Security Attacks

- Naively allowing query access to data markets is risky for users/orgs
  - **Linkage attacks**
  - **Reconstruction attacks**
  - **Inference attacks**
  - Plaintext/ciphertext attacks
- Naive designs of data markets is risky for valuation
  - Manipulation of pricing and negotiation mechanisms
  - Less trust in data markets

Motivates the need for robust *privacy and security protections*

Inference Attacks

Inference attack: Attacker gets $\tilde{O}(n^2)$ count queries with noise $o(n)$ and needs to know if someone is in the dataset or not.

Inference Attacks

Public Access to Genome-Wide Data: Five Views on Balancing Research with Privacy and Protection

P3G Consortium, George Church, Catherine Heeney, Naomi Hawkins, Jantina de Vries, Paula Boddington, Jane Kaye, Martin Bobrow, Bruce Weir

Just over twelve months ago, *PLoS Genetics* published a paper [1] demonstrating that, given genome-wide genotype data from an individual, it is, in principle, possible to ascertain whether that individual is a member of a larger group defined solely by aggregate genotype frequencies, such as a forensic sample or a cohort of participants in a genome-wide association study (GWAS). As a consequence, the National Institutes of Health (NIH) and Wellcome Trust agreed to shut down public access not just to individual genotype data but even to aggregate genotype frequency data from each study published using their funding. Reactions to this decision span the full breadth of opinion, from “too little, too late—the public trust has been breached” to “a heavy-handed bureaucratic response to a practically minimal risk that will unnecessarily inhibit scientific research.” Scientific concerns have also been raised over the conditions under which individual identity can truly be accurately determined from GWAS data. These concerns are addressed in two papers published in this month's issue of *PLoS Genetics* [2],[3]. We received several submissions on this topic and decided to assemble these viewpoints as a contribution to the debate and ask readers to contribute their thoughts through the PLoS online commentary features.

Privacy and Security Attacks

- Naively allowing query access to data markets is risky for users/orgs
  - **Linkage attacks**
  - **Reconstruction attacks**
  - **Inference attacks**
  - **Plaintext/ciphertext attacks**
- Naive designs of data markets is risky for valuation
  - Manipulation of pricing and negotiation mechanisms
  - Less trust in data markets

Motivates the need for robust *privacy and security protections*

Plaintext/Ciphertext Attacks

A datamarket could encrypt the interaction between buyers/sellers/platforms

Plaintext/Ciphertext Attacks

A datamarket could encrypt the interaction between buyers/sellers/platforms. The encryption scheme should be secure against one or more threat models:

Ciphertext-only attack

Known-plaintext attack

Chosen-plaintext attack

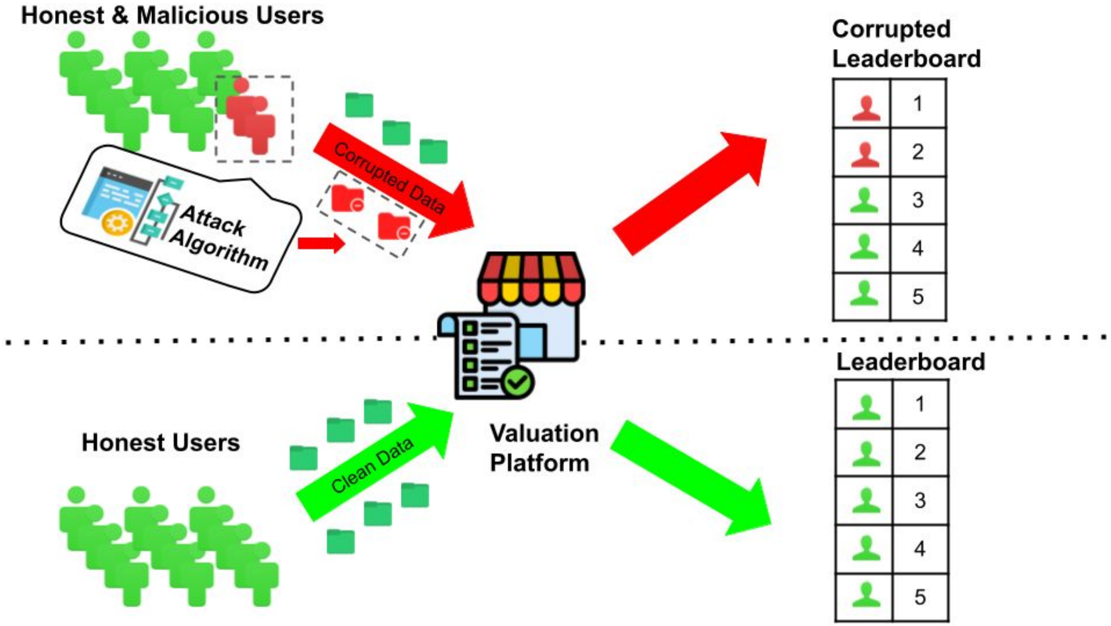
Chosen-ciphertext attack

Privacy and Security Attacks

- Naively allowing query access to data markets is risky for users/orgs
  - Linkage attacks
  - Reconstruction attacks
  - Inference attacks
  - Plaintext/ciphertext attacks
- Naive designs of data markets is risky for valuation
  - **Manipulation of pricing and negotiation mechanisms**
  - Less trust in data markets

Motivates the need for robust *privacy and security protections*

Valuation Attacks



Valuation Attacks

MemAttack: Efficiently Attacking Memorization Scores

by Do, Chandrasekaran, Alabi (2025)

Influence estimation tools—such as memorization scores—are widely used to understand model behavior, attribute training data, and inform dataset curation. However, recent applications in data valuation and responsible machine learning raise the question:

Can these scores themselves be adversarially manipulated?

In this work, we present a systematic study of the feasibility of attacking memorization-based influence estimators. We propose efficient mechanisms that allow an adversary to perturb specific training samples or small subsets of data to inflate or suppress their corresponding influence scores, all while maintaining high utility on natural downstream tasks. Our attacks are practical, requiring only black-box access to model outputs and incur moderate computational overhead. We empirically validate our methods on MNIST, SVHN, and CIFAR-10, showing that even state-of-the-art estimators are vulnerable to targeted score manipulations. In addition, we provide a theoretical analysis of the stability of memorization scores under adversarial perturbations, revealing conditions under which influence estimates are inherently fragile. Our findings highlight critical vulnerabilities in influence-based attribution and suggest the need for robust defenses.

Valuation Attacks

In large datasets, a small subset of highly influential (memorized) training examples disproportionately affects the model's predictions and generalization capabilities, while the majority of examples have little to no impact. Influence scores quantify how much each datapoint affects the model's predictions.

Influence scores can be used to price data.

- Tom Yan and Ariel D Procaccia. **If you like shapley then you'll love the core.** *AAAI 2021*
- Tianshu Song, Yongxin Tong, and Shuyue Wei. **Profit allocation for federated learning**. *In 2019 IEEE International Conference on Big Data (Big Data), pages 2577–2586. IEEE, 2019.*
- Jiachen T Wang and Ruoxi Jia. **Data banzhaf: A robust data valuation framework for machine learning**. *AISTATS 2023*.
- Tianhao Wang, Johannes Rausch, Ce Zhang, Ruoxi Jia, and Dawn Song. **A principled approach to data valuation for federated learning**. *Federated Learning: Privacy and Incentive, pages 153–167, 2020.*

Valuation Attacks

Memorization Score

$$\mathtt{mem}(\mathcal{A}, \mathbf{z}, q(\mathbf{z})) := \Pr_{(x,y)\leftarrow q(\mathbf{z}),h\leftarrow\mathcal{A}(\mathbf{z}\cup q(\mathbf{z}))}[h(x) = y] - \Pr_{(x,y)\leftarrow q(\mathbf{z}),h\leftarrow\mathcal{A}(\mathbf{z})}[h(x) = y]$$

Quantifies how much a new example would change the performance of a classifier.

Valuation Attacks via Memorization Scores

1) **Out-of-Distribution (OOD) Replacement Attack.**
2) **Pseudoinverse Attack (PINV)**
3) **EMD Attack**: Maximize Wasserstein distance between original and perturbed data points
4) **DeepFool (DF) Perturbation Attack**: Sample points along decision boundary

Valuation Attacks: Experimental Results

{Loss Curvature, Confidence Event, and Privacy Score} are proxies for the memorization scores.

We evaluate on MNIST, SVHN, CIFAR-10 datasets.

Higher scores correspond to more memorization from the attack data points.

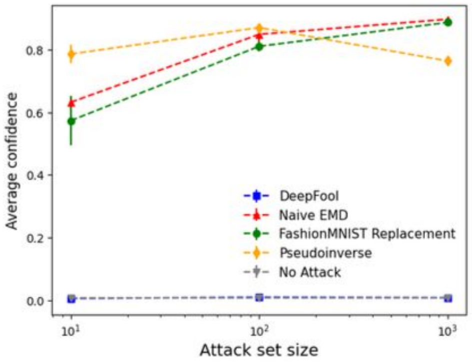
Attack	Loss Curvature			Confidence Event			Privacy Score		
	MNIST	SVHN	CIFAR-10	MNIST	SVHN	CIFAR-10	MNIST	SVHN	CIFAR-10
None	0.00±0.00	0.01±0.00	0.09±0.00	0.01±0.00	0.06±0.00	0.23±0.00	0.47±0.00	0.49±0.00	0.19±0.00
OOD	0.13±0.00	0.02±0.00	**0.14±0.00**	0.62±0.00	0.52±0.00	0.61±0.00	0.08±0.00	-0.10±0.00	0.09±0.00
PINV	**0.14±.00**	**0.14±0.00**	0.08±0.00	**0.85±0.00**	**0.81±0.00**	**0.66±0.00**	**0.29±0.01**	**0.51±0.00**	**0.79±0.00**
EMD	0.06±0.00	0.00±0.00	-0.05±0.00	0.51±0.00	0.68±0.00	0.54±0.00	-0.03±0.00	-0.03±0.00	0.01±0.00
DF	0.00±0.00	0.00±0.00	0.01±0.00	0.00±0.00	0.00±0.00	0.00±0.00	-0.02±0.00	-0.01±0.00	-0.02±0.00

Valuation Attacks: Experimental Results

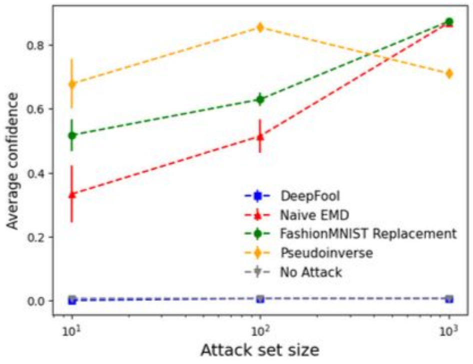
{Loss Curvature, Confidence Event, and Privacy Score} are proxies for the memorization scores.

We evaluate on (standard) deep neural network architectures: VGG, ResNet, MobileNet.

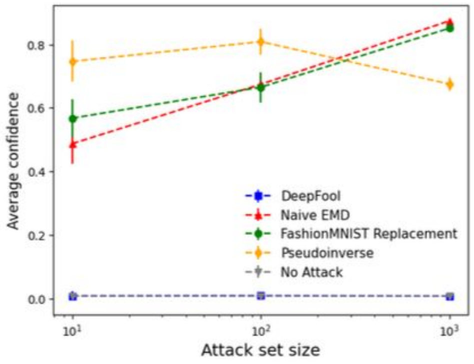
Higher scores correspond to more memorization from the attack data points.



(a) VGG-11



(b) ResNet-18



(c) MobileNet-v2

Conclusion: Privacy and Security Attacks

- Naively allowing query access to data markets is risky for users/orgs
  - Reconstruction attacks
  - Inference attacks
  - Plaintext/ciphertext attacks
- Naive designs of data markets leads is risky for valuation
  - Manipulation of pricing and negotiation mechanisms
  - Less trust in data markets

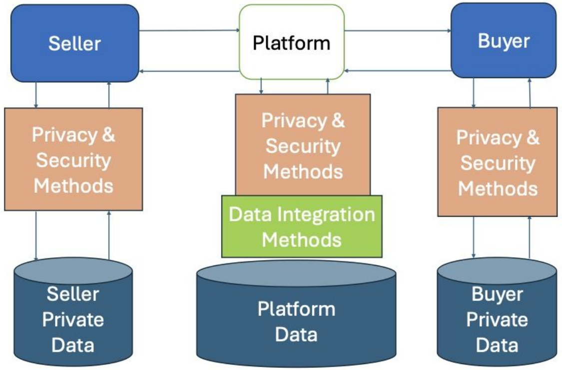
Need to provide robust *privacy and security protections*

Conclusion: Privacy and Security Attacks

Need to provide robust *privacy and security protections* via security definitions:

1. **Security guarantee**: what is the scheme/protocol in the data market intended to prevent the attacker from doing?
2. **Threat model**: what is the power of the adversary in the data market? What actions can the attacker perform?

Protect Information in Data Markets



1. Protect buyers from *malicious* sellers
2. Protect sellers from *malicious* buyers
3. Prevent *unauthorized* users from accessing:
   a. Seller private data
   b. Buyer private data
   c. Platform private data
4. Prevent manipulation of data acquisition mechanisms:
   a. Data discovery
   b. Data valuation
   c. Data negotiation
   d. Data delivery

Next: How do we *protect* the information?