Introduction to Differential Privacy

Tudor Cebere

February 6, 2024

1. Motivation

2. *What* is DP

3. *How* is DP achieved

4. *Where* to find DP

Notes M2 S1 Informatique Fondamentale 2021-2022		DEPARTEMENT D'INFORMATIQUE																	
		CR01	CR02	CR03	CR04	CR05	CR06	CR07	CR08	CR09	CR10	CR11	CR12	CR13	CR14	CR15	CR16	CR17	CR18
INE	**N° Etudiant**	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
0110014058Y	3700392		15,5		19,25		16,75			14,5			15	18,13					
0110035443T	3700252		14	16	14	15						#10,5		15,25	14,65				
0111016015G	3700640	14	15		15,25			#14	17,71	16	14,67								
0113031566Y	3700351				13,75	11						13			14,63			13,5	13,5
0209024246T	3700354		12	16	10							12,5			13,11		15,5		
0409008788V	3700202		14					13,05						15,63				18,5	13,4
0809035111R	3600547		13				12,58			11		15,5	16	13,13					
0810013837Y	3700428				8,75			12,8	16,14			12						10,5	
0810047127H	3700419	#14	#14		17,5		14,83	14		15	15,17	15		#12,63					
0815906608J	3700281	14,5	9				14,08			14,5	#7,5					14,6			
0909063132A	3600314		#10	17	16,25	18						13,5			15,5		15		
0914300277Z	3500166				16,75					17,5			14,5	11,13			17		
0EVL7V000Z9	3800655	#12						14,05			13,83			#12,5		15,1		16,5	13,2
0EVL7V001F5	3800603							12,4	12,79						16,03		12,5	13	14,3
1006001573W	3900521			9,5	#9,25			12,8				11			14,01	14,9	#8,5		13,4
1198027050R	3800499	11,5				14		12,95								13,9		14,5	12,7
1209005144B	3900459		15		15,25		19			18,5		14		12,56					
1209032876V	3700098	#14				18		14,05							18,23			18,5	16,2
1409037518V	3700309	14,5						12,9	13,93									13,5	13,2
1509018315Z	3800626	14,5			11,25	15			16,57	#10,5	9	13,5							
1510022290B	3700442									13,5	15		10	12,13			11		14,2
1608004677W	3500471	17,5		18,5	19,75	17,5											17,5		
1710012202R	3700301		8,5		12,75		16,25			13,5			6				12		
1808028029H	3500286		17	16,5	17,75					17		15,5		#13,13			17		
1810025457A	3700300		#12	15	18,75		14,17			15	#12,5	#11,5	#11,5	14			15,5		
2009006907K	3900472							12,95	16,14						15,1		#12	15,5	12,6
2009013043E	3900528							#12,95							13,17	15,8	13	17	16,1
213043839HE	3900638			#8,5		15,5		12,8							11,6	15,1		10	13,8
213043915KF	3900627							12,4	16	14,5					13,07			10,5	15,3
2408016882E	3500550				18,75					18,5		15	17				19		
2409023405B	3800537	14						14,05	16	13,5								16,5	14,4

Notes M2 S1 Informatique Fondamentale 2021-2022		DEPARTEMENT D'INFORMATIQUE																	
		CR01	CR02	CR03	CR04	CR05	CR06	CR07	CR08	CR09	CR10	CR11	CR12	CR13	CR14	CR15	CR16	CR17	CR18
INE	N° Etudiant	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
0110014058Y	3700392		15,5		19,25		16,75			14,5			15	18,13					
0110035443T	3700252		14	16	14	15						#10,5		15,25	14,65				
0111016015G	3700640	14	15		15,25			#14	17,71	16	14,67								
0113031566Y	3700351				13,75	11						13			14,63			13,5	13,5
0209024246T	3700354		12	16	10							12,5			13,11		15,5		
0409008788V	3700202		14					13,05						15,63				18,5	13,4
0809035111R	3600547		13				12,58			11		15,5	16	13,13					
0810013837Y	3700428				8,75			12,8	16,14			12						10,5	
0810047127H	3700419	#14	#14		17,5		14,83	14		15	15,17	15		#12,63					
0815906608J	3700281	14,5	9				14,08			14,5	#7,5					14,6			
0909063132A	3600314		#10	17	16,25	18						13,5			15,5		15		
0914300277Z	3500166				16,75					17,5			14,5	11,13			17		
0EVL7V000Z9	3800655	#12						14,05			13,83			#12,5		15,1		16,5	13,2
0EVL7V001F5	3800603							12,4	12,79						16,03		12,5	13	14,3
1006001573W	3900521			9,5	#9,25			12,8				11			14,01	14,9	#8,5		13,4
1198027050R	3800499	11,5				14		12,95								13,9		14,5	12,7
1209005144B	3900459		15		15,25		19			18,5		14		12,56					
1209032876V	3700098	#14				18		14,05							18,23			18,5	16,2
1409037518V	3700309	14,5						12,9	13,93									13,5	13,2
1509018315Z	3800626	14,5			11,25	15			16,57	#10,5	9	13,5							
1510022290B	3700442									13,5	15		10	12,13			11		14,2
1608004677W	3500471	17,5		18,5	19,75	17,5											17,5		
1710012202R	3700301		8,5		12,75		16,25			13,5			6				12		
1808028029H	3500286		17	16,5	17,75					17		15,5		#13,13			17		
1810025457A	3700300		#12	15	18,75		14,17			15	#12,5	#11,5	#11,5	14			15,5		
2009006907K	3900472							12,95	16,14						15,1		#12	15,5	12,6
2009013043E	3900528							#12,95							13,17	15,8	13	17	16,1
213043839HE	3900638			#8,5		15,5		12,8							11,6	15,1		10	13,8
213043915KF	3900627							12,4	16	14,5					13,07			10,5	15,3
2408016882E	3500550				18,75					18,5		15	17				19		
2409023405B	3800537	14						14,05	16	13,5								16,5	14,4

Notes M2 S1 Informatique Fondamentale 2021-2022

		DEPARTEMENT D'INFORMATIQUE																			MATHS
		CR01	CR02	CR03	CR04	CR05	CR06	CR07	CR08	CR09	CR10	CR11	CR12	CR13	CR14	CR15	CR16	CR17	CR18	QCS (M1IF)	Probabilités avancées - MATH4104
INE	N° Etudiant	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6
0110014058Y	3700392		15,5		19,25		16,75			14,5			15	18,13							
0110035443T	3700252		14	16	14	15						#10,5		15,25	14,65						
0111016015G	3700640	14	15		15,25			#14	17,71	16	14,67										
0113031566Y	3700351				13,75	11						13			14,63			13,5	13,5		
0209024246T	3700354		12	16	10							12,5			13,11		15,5				
0409008788V	3700202		14					13,05						15,63				18,5	13,4		
0809035111R	3600547		13				12,58			11		15,5	16	13,13							
0810013837Y	3700428				8,75			12,8	16,14			12						10,5			
0810047127H	3700419	#14	#14		17,5		14,83	14		15	15,17	15		#12,63							
0815906608J	3700281	14,5	9				14,08			14,5	#7,5					14,6					
0909063132A	3600314		#10	17	16,25	18						13,5			15,5		15				
0914300277Z	3500166				16,75					17,5			14,5	11,13			17			15,89	
0EVL7V000Z9	3800655	#12						14,05			13,83			#12,5		15,1		16,5	13,2		
0EVL7V001F5	3800603							12,4	12,79						16,03		12,5	13	14,3		
1006001573W	3900521			9,5	#9,25			12,8				11			14,01	14,9	#8,5		13,4		
1198027050R	3800499	11,5				14		12,95								13,9		14,5	12,7		
1209005144B	3900459		15		15,25		19			18,5		14		12,56							
1209032876V	3700098	#14				18		14,05							18,23			18,5	16,2		
1409037518V	3700309	14,5						12,9	13,93									13,5	13,2		
1509018315Z	3800626	14,5			11,25	15			16,57	#10,5	9	13,5									
1510022290B	3700442									13,5	15		10	12,13			11		14,2		
1608004677W	3500471	17,5		18,5	19,75	17,5											17,5				17
1710012202R	3700301		8,5		12,75		16,25			13,5			6				12				
1808028029H	3500286		17	16,5	17,75					17		15,5		#13,13			17				
1810025457A	3700300		#12	15	18,75		14,17			15	#12,5	#11,5	#11,5	14			15,5				
2009006907K	3900472							12,95	16,14						15,1		#12	15,5	12,6		
2009013043E	3900528							#12,95							13,17	15,8	13	17	16,1		
213043839HE	3900638			#8,5		15,5		12,8							11,6	15,1		10	13,8		
213043915KF	3900627							12,4	16	14,5					13,07			10,5	15,3		
2408016882E	3500550				18,75					18,5		15	17				19			17,1	
2409023405B	3800537	14						14,05	16	13,5								16,5	14,4		

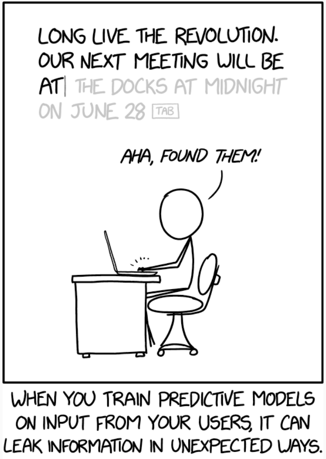
Attacks on anonymization

- Netflix dataset: Narayanan and Shmatikov 2008

- Twitter graph using Flickr: Narayanan and Shmatikov 2009

Aggregated statistics

- Reconstruction attacks on US Census Data: Dinur and Nissim 2003
- Membership inference attacks on Genomics Data: Homer et al. 2008



Evaluating and Testing Unintended Memorization in Neural Networks - N. Carlini

Attacks on ML models

- Reconstruction attacks on Neural Networks: Haim et al. 2022
- Membership inference attacks Neural Networks: Shokri et al. 2017

Outline

- We need a way to *measure* privacy.
- Auxiliary data breaks anonimization techniques
- Aggregated statistics or machine learning models **do not** protect privacy

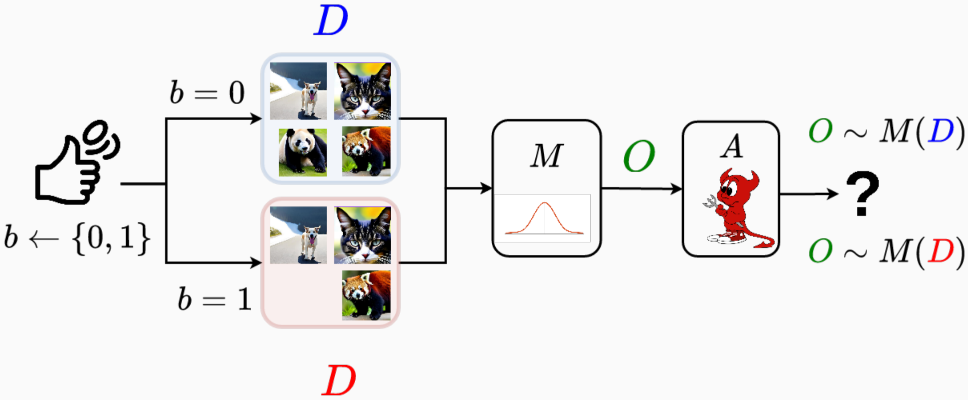
1. Motivation

2. ***What* is DP**

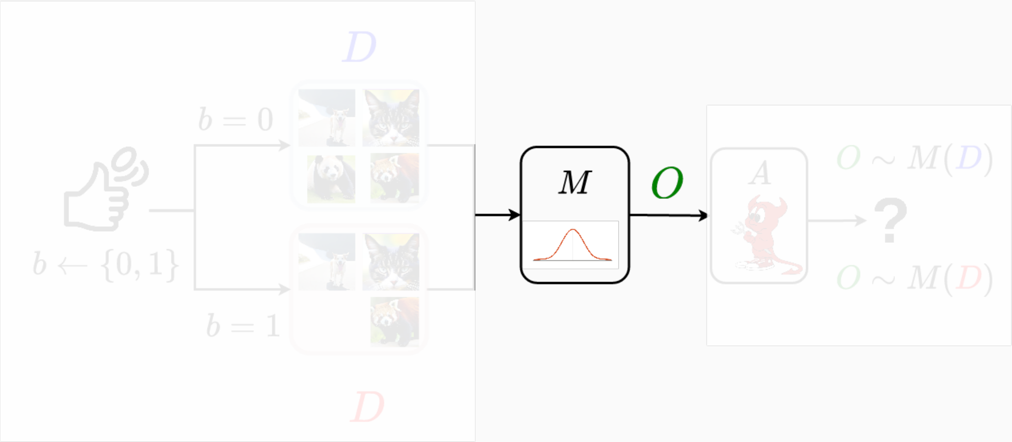
3. *How* is DP achieved

4. *Where* to find DP

Game of Privacy



Differential Privacy (Dwork et al. 2006)



(ϵ, δ) Differential Privacy in Hypothesis Testing

Differential Privacy as Hypothesis Testing

Given a output O of a (ϵ, δ)-DP mechanism M and two neighboring datasets D, D', consider the following hypothesis testing experiment:

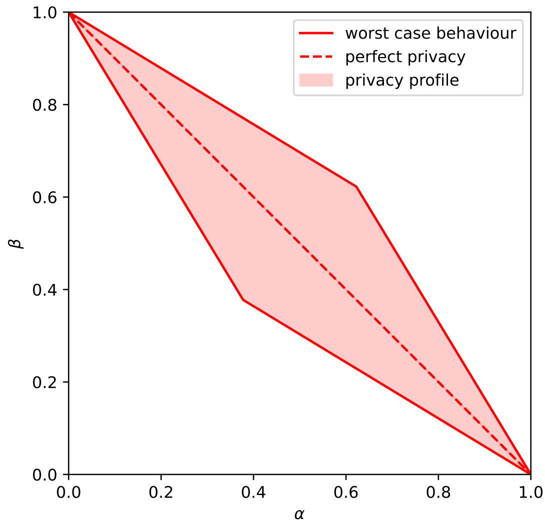
$$\begin{aligned} H_0 &: O \text{ was computed on } D \\ H_1 &: O \text{ was computed on } D' \end{aligned} \tag{1}$$

Any rejection rule and it's expectation of Type I (α) and II (β) errors, satisfies:

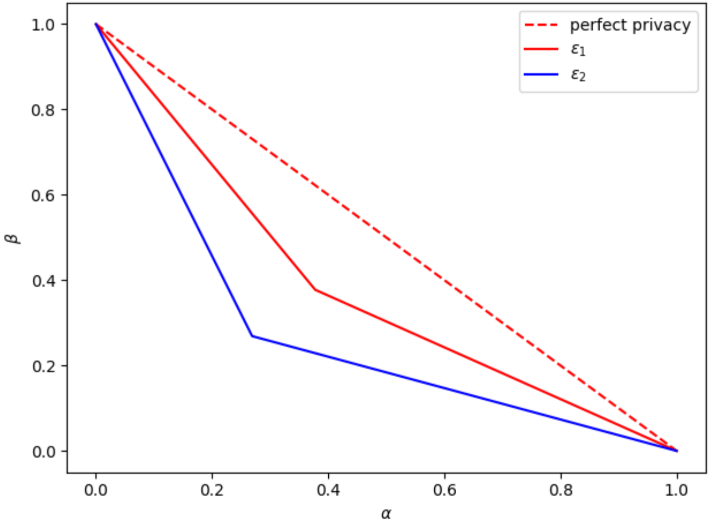
$$\begin{aligned} \alpha + e^{\epsilon}\beta &\geq 1 - \delta \\ \beta + e^{\epsilon}\alpha &\geq 1 - \delta \end{aligned} \tag{2}$$

A statistical framework for differential privacy - L. Wasserman et al.

Privacy Profiles



Quizz: What can we say about ϵ_1 and ϵ_2?



Differential Privacy (Dwork et al. 2006)

(ϵ, δ) Differential Privacy (DP)

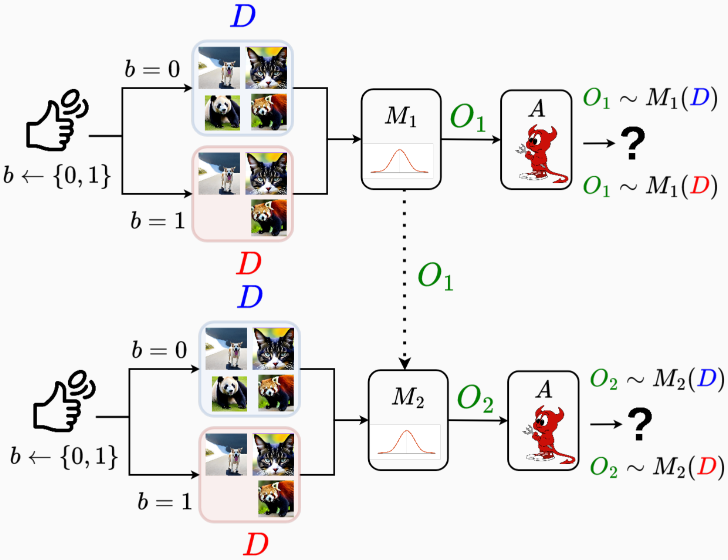
A mechanism $M : \mathcal{X}^* \to \mathcal{Y}$ is (ϵ, δ)-DP if for all neighboring datasets D and D' the following inequality holds for all $S \in \mathcal{Y}$:

$$P[M(D) \in S] \leq e^{\epsilon} P[M(D') \in S] + \delta \tag{3}$$

Fundamental properties of Differential Privacy

- Composition
- Post-Processing
- Group Privacy

Composition



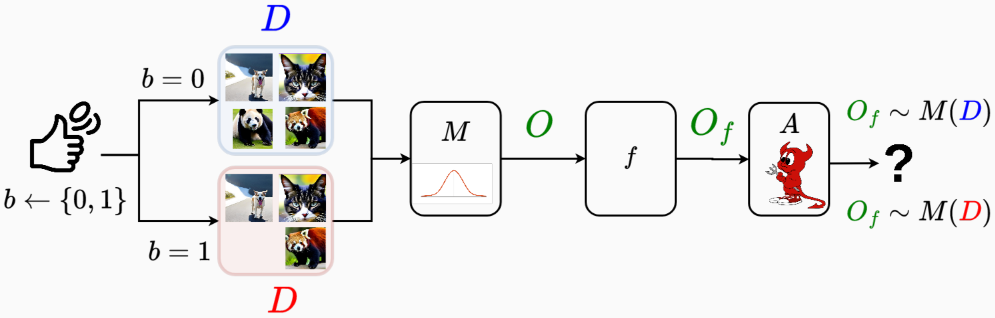
Composition Theorem of Differential Privacy

Composition

The individual privacy guarantees of multiple mechanisms $M_1 \ldots M_N$ can be composed into a single privacy guarantee.

- **Note:** The composition can be *sequential* or *adaptive*.
- **Implication:** DP does not restrain the number of released statistics.

Post-processing



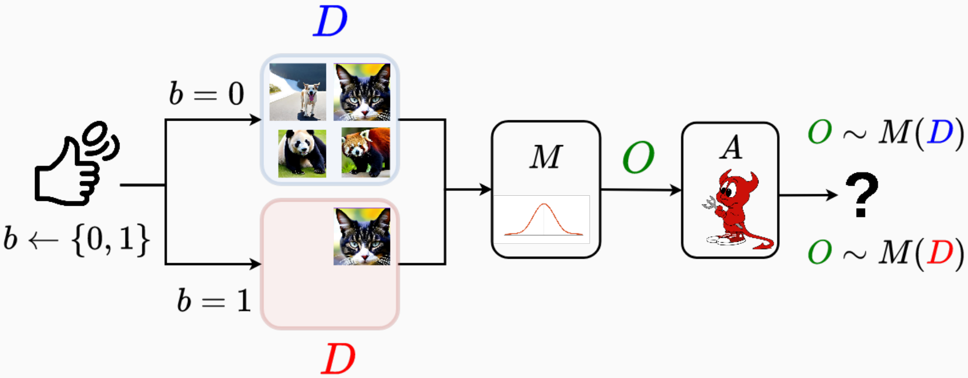
Post-processing Theorem of Differential Privacy

Post-processing

Differential privacy is immune to post-processing.

- **Implication:** it is safe to release DP-processed results and post-process as needed any DP-mechanism output, the privacy leakage of the statistic is *upperbounded* by the original mechanism.

Group Privacy



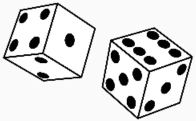
Group Privacy in Differential Privacy

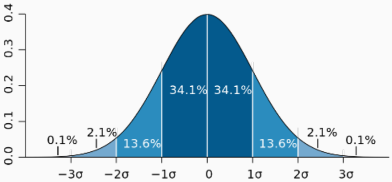
Group Privacy

The neighboring relationship of DP extends to multiple samples.

- **Linear Decay:** Privacy guarantees degrade linearly with the group size.
- **Implication:** Users can have multiple data entries (e.g. location data).

1. Motivation
2. *What* is DP
3. ***How* is DP achieved**
4. *Where* to find DP




0.4
0.3
0.2
0.1
0.0
0.1%
2.1%
13.6%
34.1%
34.1%
13.6%
2.1%
0.1%
−3σ
−2σ
−1σ
0
1σ
2σ
3σ

Randomized Response (1965, S.L. Warner) is a statistical method used to encourage truthful responses via *plausible deniability*.

Mechanism

Let X be a sensitive attribute with binary response: Yes (1) or No (0). The subject flips a *biased* coin in private and follows this protocol:

- If the coin comes up heads, the subject tells the truth.
- If the coin comes up tails, the subject lies.

Privacy Guarantees

Let $X_i \in \{0,1\}$ be a sensitivite attribute for each datasample, $D = \{X_1, \dots X_j \dots X_N\}$, $D' = \{X_1, \dots X_j' \dots X_N\}$ and $\xi \in [0, \frac{1}{2}]$, the randomized response mechanism is defined as:

$$M_{RR}(X_i) = \begin{cases} X_i, & \text{with probability } \frac{1}{2} + \xi \\ 1 - X_i, & \text{with probability } \frac{1}{2} - \xi \end{cases} \quad (4)$$

Theorem. M_{RR} is $(O(\xi), 0)$-Differentially Private for $\xi < \frac{1}{4}$.

Note. Via M_{RR} we can compute an unbiased estimator for $\frac{1}{N}\sum_{1}^{N} X_i$.

Randomized Response Mechanism

- Want to rerun the experiment? $\leftarrow$ **Composition**
- Want to run some processing on top of the results? $\leftarrow$ **Post-Processing**

Zoo of Mechanisms

- **Selection problems**: Exponential Mechanism
- **Continous support problems**: Laplace Mechanism, Gaussian Mechanism
- ...

1. Motivation

2. *What* is DP

3. *How* is DP achieved

4. ***Where* to find DP**

Applications

- Apple: New Word Discovery, Emoji Popularity, Safari Usage Reports
- Google: GBoard
- Microsoft: Global Victim-Perpetrator Synthetic Dataset, Windows Telemetry
- U.S. Census Bureau: US Census Data 2020

Software for Differential Privacy

- OpenDP
- Tumult Labs Analytics
- Google Privacy
- OpenMined - PyDP

Resources to learn Differential Privacy

- Algorithms for Private Data Analysis by Gautam Kamath
- The complexity of differential privacy by Salil Vadhan
- The Algorithmic Foundations of DP by Cynthia Dwork and Aaron Roth
- Algorithms for Private Data Analysis by Sasho Nikolov
- Privacy Preserving Machine Learning by Aurélien Bellet

Dinur, Irit and Kobbi Nissim (2003). "Revealing information while preserving privacy". In: *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. Association for Computing Machinery. ISBN: 1581136706 (cit. on p. 8).

Dwork, Cynthia et al. (2006). "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Theory of Cryptography* (cit. on pp. 14, 18).

Haim, Niv et al. (2022). "Reconstructing training data from trained neural networks". In: *Advances in Neural Information Processing Systems* 35, pp. 22911–22924 (cit. on p. 10).

Homer, Nils et al. (2008). "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays". In: *PLOS Genetics*. DOI: 10.1371/journal.pgen.1000167 (cit. on p. 8).

Narayanan, Arvind and Vitaly Shmatikov (2008). "Robust De-anonymization of Large Sparse Datasets". In: *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pp. 111–125. DOI: 10.1109/SP.2008.33 (cit. on p. 7).

Narayanan, Arvind and Vitaly Shmatikov (2009). “De-anonymizing Social Networks”. In: *2009 30th IEEE Symposium on Security and Privacy*, pp. 173–187. DOI: 10.1109/SP.2009.22 (cit. on p. 7).

Shokri, Reza et al. (2017). “Membership inference attacks against machine learning models”. In: *2017 IEEE symposium on security and privacy (SP)*. IEEE, pp. 3–18 (cit. on p. 10).