Towards Formalizing Speech Privacy with Differential Privacy

Yang Cao (Hokkaido University)

2023/10/2 @ SPSC Seminar

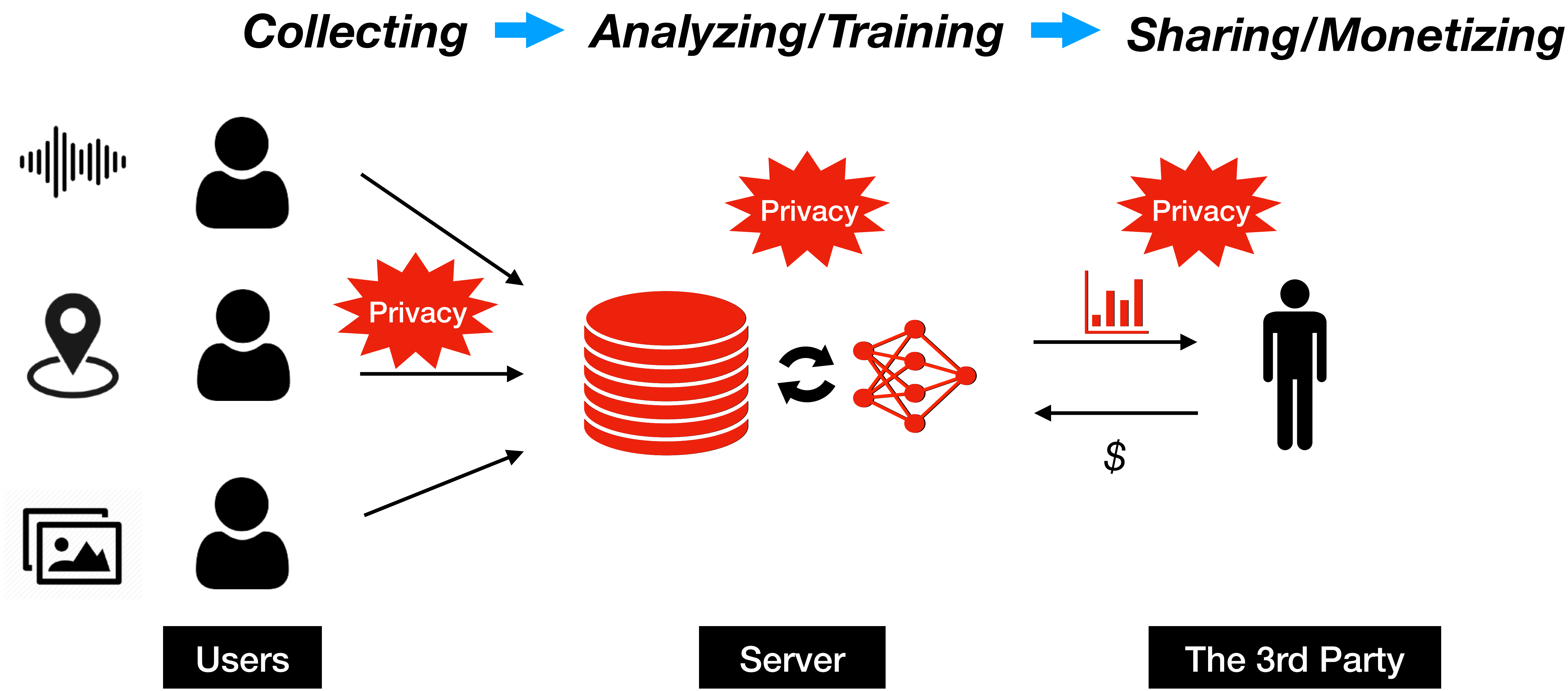
Outline

- Scenario and Motivation
    - why we need to formalize speech privacy?
- A brief history of privacy definitions
    - from k-Anonymity to Differential Privacy
- Our Studies for Formalizing Speech Privacy
    - [**ICME20**] Voice-Indistinguishability
    - [**ICASSP23**] General or Specific? Investigating Effective Speech Privacy Protection in Federated Learning for Speech Emotion Recognition
- Open Problems and Future Directions

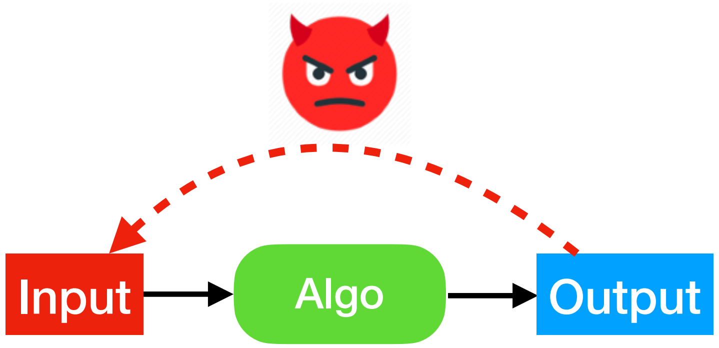
Outline

- Scenario and Motivation
  - why we need to formalize speech privacy?
- A brief history of privacy definitions
  - from k-Anonymity to Differential Privacy
- Our Studies for Formalizing Speech Privacy
  - [**ICME20**] Voice-Indistinguishability
  - [**ICASSP23**] General or Specific? Investigating Effective Speech Privacy Protection in Federated Learning for Speech Emotion Recognition
- Open Problems and Future Directions

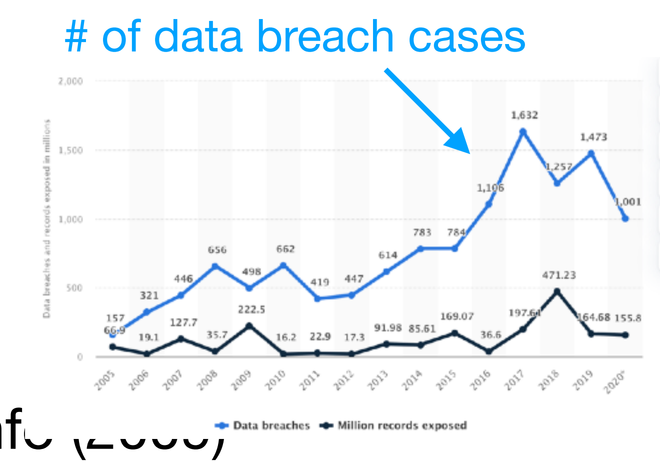
Scenario: Pipeline in Data Science



Privacy Concerns



- **Privacy Attacks**
  - *Data reconstruction attack* against statistical info [1] and ML models [2]
  - *Membership inference attack* against machine learning models [3]
- **Real-world Privacy Incidents**
  - De-identified AOL search log can be re-identified (2006)
  - NIH's DNA dataset discloses users' disease (2008)
  - Netflix anonymized watch history dataset reveals user's sensitive inf

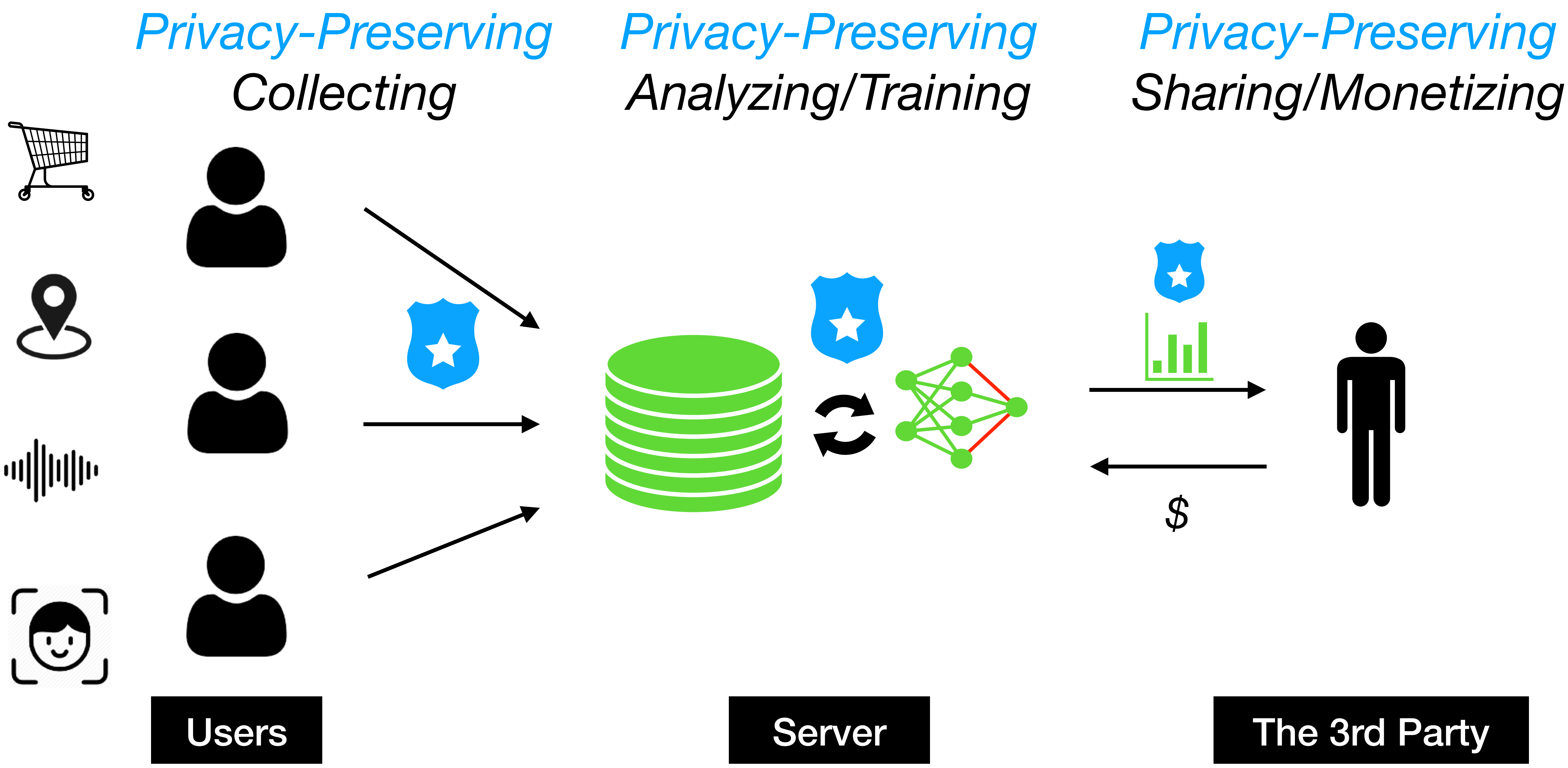


  - Facebook-Cambridge Analytica Data Scandal (2018)
  - Apple collects users' speech data for Siri quality evaluation process (2020)
- ⚠️ Privacy issues may hinder the development of data science
  - Individuals or organizations are not willing to share their data

[1] Dinur et al., "Revealing Information While Preserving Privacy." ACM PODS 2013.
[2] Papernot et al., "SoK: Security and Privacy in Machine Learning." IEEE Euro S&P 2018.
[3] Shokri et al., "Membership inference attacks against machine learning models." IEEE S&P 2017.

Privacy-Enhancing Technologies (PET)

is indispensable for Data-Driven Society



Why We Need to Formalize Privacy

- If privacy is the goal, we need to clarify **What Privacy Is**.
- Privacy is often an **ambiguous concept**, like
  - “the data is invisible to the adversary”
  - “my identify is invisible to the server”
  - “my identify is ε-differentially private to the server”
- We need to have a **mathematically quantifiable metrics** about the privacy risk
  - what is the scenario, what is the secret, who is the adversary, what kinds of attacks, etc..

Outline

- Scenario and Motivation
  - why we need to formalize speech privacy?
- A brief history of privacy definitions
  - from k-Anonymity to Differential Privacy
- Our Studies for Formalizing Speech Privacy
  - [**ICME20**] Voice-Indistinguishability
  - [**ICASSP23**] General or Specific? Investigating Effective Speech Privacy Protection in Federated Learning for Speech Emotion Recognition
- Open Problems and Future Directions

A Key Question: How to Define Privacy

- **(2000 ~ 2006) Early efforts on "*privacy as anonymity*"**
  - k-anonymity [4], L-diversity [5], t-closeness [6]
  - Such a privacy definition is conditioned on the attackers' knowledge

[4] Sweeney, "k-anonymity: A model for protecting privacy." Int. J. Uncertain. Fuzziness Knowl.-Based Syst, 2002.
[5] Machanavajjhala et al., "L-diversity: Privacy beyond k-anonymity." ACM TKDD 2007.
[6] Li et al., "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity." IEEE ICDE 2007.

Data Privacy in the early age (2000~2006)

- A Runining Example: Medical Data Sharing
  - Medical records is valuable for data analysis
  - But the health condition is very sensitive!

medical records

sensitive!

Name	Sex	Birth	ZIP	disease
Tom	M	1/1	1001	cardiopathy
Jack	M	1/2	1002	diabete
Bob	M	1/3	1003	HIV
Wang	F	2/1	2001	HIV
Alice	F	2/2	2002	HIV
Dua	F	2/3	2003	HIV

First thought: anonymize by removing PII

- PII = Personally Identifying Information
  - anything that identifies the person directly
  - Name, Phone number, Email, Address …

💡 Cut the link between a specific person and the medical record

medical records without PII

Name	Sex	Birth	ZIP	disease
[redacted]	M	1/1	1001	cardiopathy
[redacted]	M	1/2	1002	diabete
[redacted]	M	1/3	1003	HIV
[redacted]	F	2/1	2001	HIV
[redacted]	F	2/2	2002	HIV
[redacted]	F	2/3	2003	HIV

Is it secure to release?

Data Privacy in the early age (2000~2006)

Re-identification by Linkage Attack

- **Just removing PII is not enough**

"Anonymized" Medical records

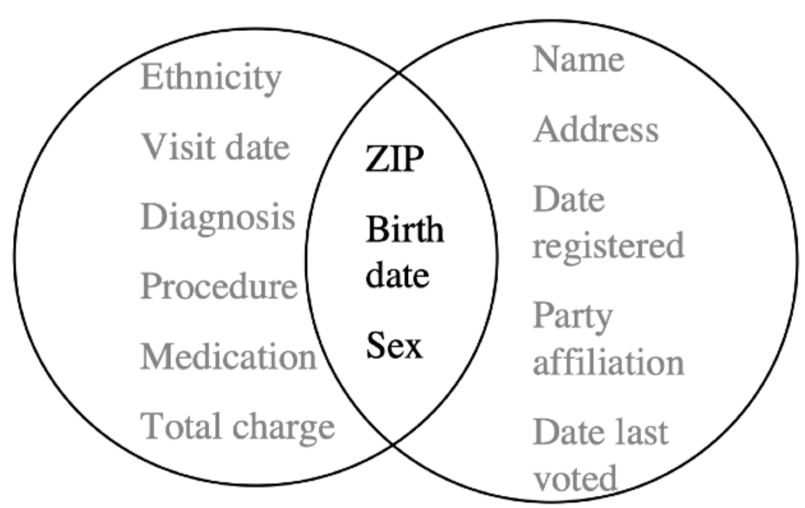
ID	Sex	Birth	ZIP	disease
r1	M	1/1	1001	cardiopathy
r2	M	1/2	1002	diabete
r3	M	1/3	1003	HIV
r4	F	2/1	2001	HIV
r5	F	2/2	2002	HIV
r6	F	2/3	2003	HIV

Attacker's Prior Knowledge

I know Bob:
{M, 1/3, 1003}
so r3 = Bob!

- **A real-world linkage attack [1]**

"Anonymized"
Massachusetts hospital
discharge dataset



Public voter dataset

L. Sweeney. 1997. Guaranteeing anonymity when sharing medical data, the Datafly System. Proc AMIA Annu Fall Symp (1997), 51–55.

Data Privacy in the early age (2000~2006)

k-Anonymity

- **Quasi-identifiers**
  - Can be used for linking anonymized dataset with other datasets

quasi-identifier (Sex, Birth, ZIP)

	Sex	Birth	ZIP	disease
	M	1/1	1001	cardiopathy
	M	1/2	1002	diabete
	M	1/3	1003	HIV
	F	2/1	2001	HIV
	F	2/2	2002	HIV
	F	2/3	2003	HIV

3-Anonymity

	Sex	Birth	ZIP	disease
	M	1/*	100*	cardiopathy
	M	1/*	100*	diabete
	M	1/*	100*	HIV
	F	2/*	200*	HIV
	F	2/*	200*	HIV
	F	2/*	200*	HIV

I know Bob:
{M, 1/3, 1003}
but which one is Bob?

Sweeney, "k-anonymity: A model for protecting privacy." Int. J. Uncertain. Fuzziness Knowl.-Based Syst, 2002.

Data Privacy in the early age (2000~2006)

L-diversity

- Hide me in a crowd of people with L-diverse sensitive data

3-Anonymity

	Sex	Birth	ZIP	disease
	M	1/*	100*	cardiopathy
	M	1/*	100*	diabete
	M	1/*	100*	HIV
	F	2/*	200*	HIV
	F	2/*	200*	HIV
	F	2/*	200*	HIV

all people in this group have HIV !

2-diversity

UID	gender	Birth	ZIP	disease
u1	male	1/*	>10	cardiopathy
u2	male	1/*	>10	diabete
u3	male	1/*	>10	HIV
u4	female	1*/*	>20	HIV
u5	female	1*/*	>20	HIV
u6	female	1*/*	>20	diabete

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, “**L-diversity: Privacy beyond k-anonymity**,” ACM Transactions on Knowledge Discovery from Data, vol. 1, no. 1, p. 3–es, Mar. 2007.

Data Privacy in the early age (2000~2006)

T-closeness

- Hide me in a group and the groups should have similar distr.

UID	gender	Birth	ZIP	disease
u1	male	1/*	>10	cardiopathy
u2	male	1/*	>10	diabete
u3	male	1/*	>10	HIV
u4	female	1*/*	>20	HIV
u5	female	1*/*	>20	HIV
u6	female	1*/*	>20	diabete

2-diversity

people in this group has high risk of HIV !

UID	gender	Birth	ZIP	disease
u1	male	1/*	>10	cardiopathy
u2	male	1/*	>10	diabete
u3	male	1/*	>10	HIV
u4	female	1*/*	>20	HIV
u5	female	1*/*	>20	cardiopathy
u6	female	1*/*	>20	diabete

0.167-closeness

similarity between the distributions of two groups

[5]N. Li, T. Li, and S. Venkatasubramanian, "**t-Closeness: Privacy Beyond k-Anonymity and l-Diversity**," in IEEE 23rd International Conference on Data Engineering, 2007. ICDE 2007, pp. 106–115.

Limitations of k-Anonymity family

- "*All these notions, however, are* ***syntactic****, in the sense that they define a property about the final "anonymized" dataset, and do not consider the algorithm or mechanism via which the output is obtained.*" [*]

- **A modern view of data privacy**: **privacy should be a property of algorithm, instead of data**.

- How can we define privacy in this way?

[*] N. Li, M. Lyu, D. Su, and W. Yang, **Differential Privacy: From Theory to Practice**. Morgan & Claypool Publishers, 2016.

Differential Privacy (DP) (2006~now)

From Semantic security to Differential Privacy

- Semantic Security [*]:

  Pr(**Attacker**(length of plaintext, ciphertext)=output)
  ≈
  Pr(**Attacker**(length of plaintext)=output)

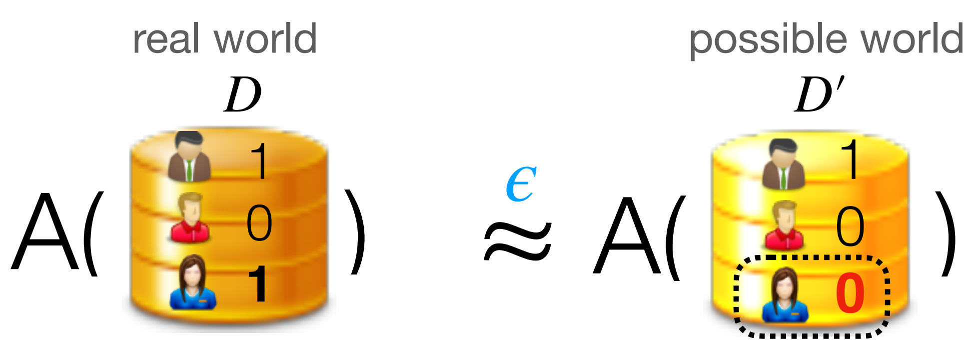
- Differential Privacy

  Pr(**M**(data with Bob)=output)
  ≈
  Pr(**M**(data without Bob)=output)

[*]S. Goldwasser, S. Micali (1982). "**Probabilistic encryption and how to play mental poker keeping secret all partial information**". Proc. 14th Symposium on Theory of Computing: *the author won Turing Award in 2012.

Differential Privacy (DP) [7]

- Randomized Algorithm A satisfies ϵ-DP over D, iff $\forall o, D, D', \frac{\Pr(A(D)=o)}{\Pr(A(D')=o)} \leq e^{\epsilon}$ where D and D' differ in any one individual record.



- Privacy parameter ϵ ($\epsilon \geq 0$): ϵ ⬆️, privacy guarantee ⬇️

- Intuitively, **DP is a constraint on algorithms:** the algorithm's <u>output should not be influenced significantly</u> by any single record of the input database

[7] Dwork, Cynthia. "Differential privacy." International Colloquium on Automata, Languages, and Programming, 2006.

DP has many variants, but all follow DP's principle

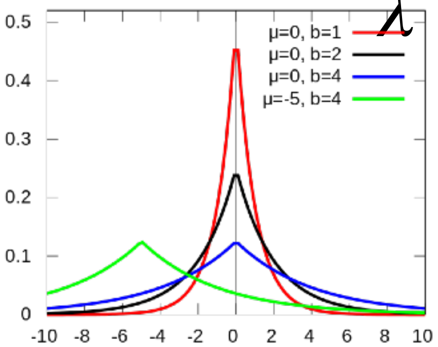
- **(ε,δ)-DP**: relaxation. Allow violation of ε-DP in probability δ
  - $\forall D, D', \Pr(o \quad D) \leq \Pr(o \quad D') * e^{\epsilon} + \delta$
- **PDP**: everyone has a personalized ε.
- **Pufferfish Privacy**: generalization of DP under constraints
- **Renyi DP**: re-place the distance of (ε,δ)-DP using Renyi divergence
- **Geo-indistinguishability**: apply DP to location data
- **Local DP**: achieve DP with an untrusted server
- **Shuffle DP**: better privacy-utility trade-off by introducing a shuffler between client and server
- **Voice-indistinguishability**: apply DP to voiceprint. our work in ICME20
- …. see [*] [**] for more details.

[*] I. Wagner and D. Eckhoff, "**Technical Privacy Metrics**: A Systematic Survey," ACM Comput. Surv., 2018.
[**] B. Pejó and D. Desfontaines, "**SoK: Differential Privacies**," in PETS, 2020.

Building blocks of DP mechanisms

- **Laplace mechanism** [*]
  - for Q(*) returns real value.
  - Adding Laplace noise lap(Δ/ε) to Q(D) → ε-DP
  - Δ is called sensitivity of Q(*), Δ=|Q(D)-Q(D')| for any D,D'.

$$Lap(x \mid \lambda) = (2\lambda)^{-1} \exp(-\frac{|x|}{\lambda})$$



- **Gaussian Mechanism**
  - for Q(*) returns real value
  - Adding Gaussian noise $\mathcal{N}(\sigma^2)$ where $\sigma = 2\Delta \log(1.25/\delta)/\epsilon^2$ to Q(D), then we have (ε,δ)-DP
  - less noise than Laplace mechanism for vector-valued functions
- **Exponential mechanism** [**]
  - For Q(*) returns categorical values
  - Return Q(D) randomly (see ** for more details)

Local DP

- **Random Response (RR)**
  - For Q(*) returns categorical values and **without (trusted) central server** to collect all user data.
  - E.g., assume d= {0,1} RR will output 1 w/ Prob. $\frac{e^\epsilon}{e^\epsilon + 1}$ if d=1; output 1 w/ Prob. $\frac{1}{e^\epsilon + 1}$ if d=0.

[*] C. Dwork, et al, Calibrating Noise to Sensitivity in Private Data Analysis, in TCC 2006.
[**] F. McSherry and K. Talwar, Mechanism Design via Differential Privacy, in FOCS, 2007.

Properties of DP

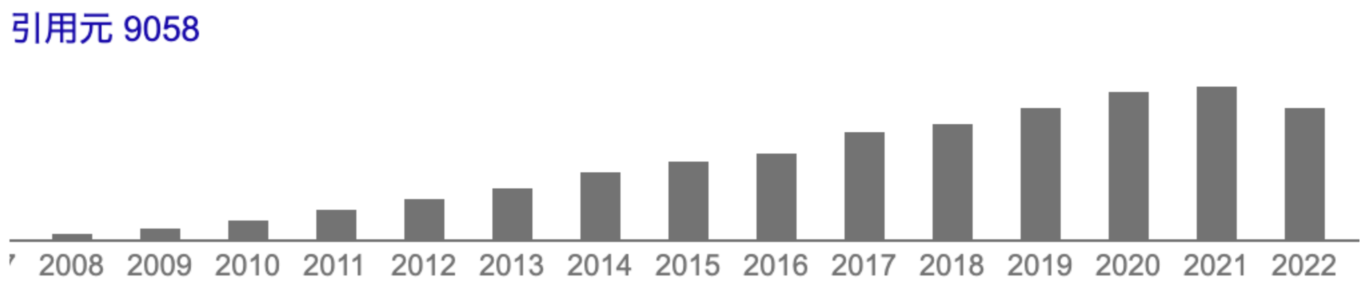
Composition Theorems & Post-processing

- Sequential composition:
  - if **M1(D)** satisfies ε1-DP and **M2(D)** satisfies ε2-DP, then we can say **M={M1,M2}** satisfies (ε1+ε2)-DP over D.
- Parallel composition:
  - Assuming D=D1∩D2 and D1, D2 are disjointed.
  - if **M1(D1)** satisfies ε1-DP and **M2(D2)** satisfies ε2-DP, then we can say **M={M1,M2}** satisfies max{ε1, ε2}-DP over D.
- Post-Processing
  - if M(D) satisfies ε-DP, for any deterministic or randomized function f, f(M(D)) satisfies ε-DP

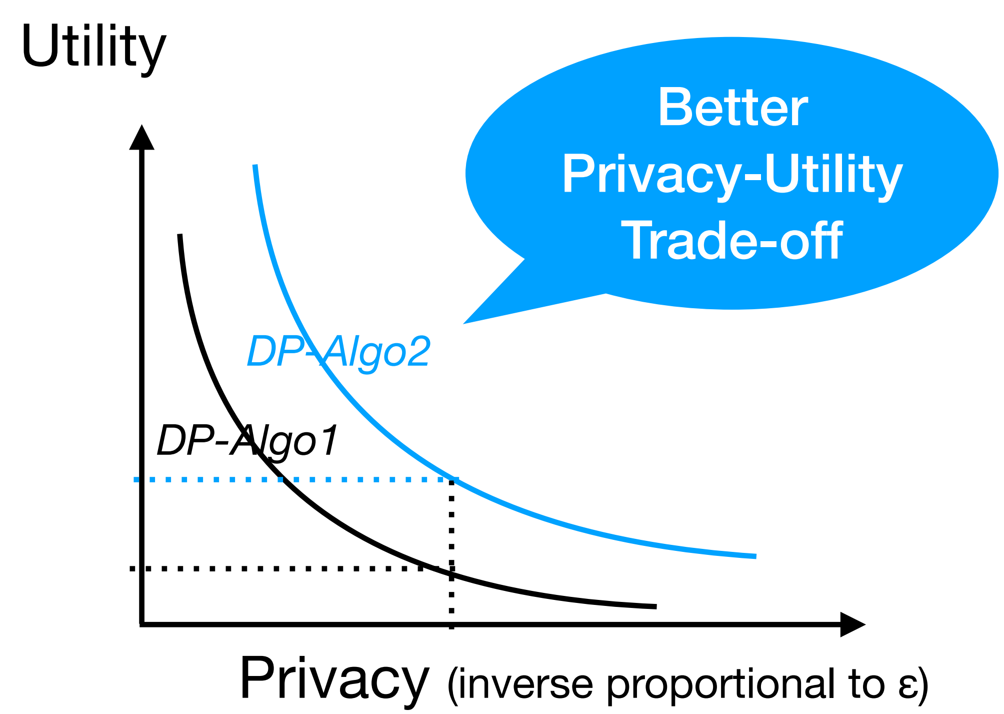
DP in Academia

- **Design "DP version" algorithms**
  - Differentially Private Data Collection [8]
  - Differentially Private Data Mining [9]
  - Differentially Private Machine Learning [10]

# of citation of Dwork's DP survey paper [11]



- Holy Grail: **Privacy-Utility Trade-off**



[8] "Differentially private data publishing and analysis: A survey." IEEE TKDE. 2017.
[9] "Data mining with differential privacy." ACM KDD 2010.
[10] "A survey on differentially private machine learning." IEEE Computational Intelligence Magazine. 2020.
[11] Dwork, Cynthia. "Differential privacy: A survey of results." Intl. conf. on theory and applications of models of computation, 2008.

DP in Industry

- **Google** - collect Chrome user click statistics (2014); release COVID-19 mobility statistics (2020)
- **Apple** - analyze App and Emoji usage (2017)
- **Microsoft** - collect Windows crash statistics (2017)
- **Facebook/Meta** - release user-sharing-url datasets (2020)
- **US Census 2020** - release demographic statistics (2020)

Outline

- Scenario and Motivation
  - why we need to formalize speech privacy?
- A brief history of privacy definitions
  - from k-Anonymity to Differential Privacy
- Our Studies for Formalizing Speech Privacy
  - [**ICME20**] Voice-Indistinguishability
  - [**ICASSP23**] General or Specific? Investigating Effective Speech Privacy Protection in Federated Learning for Speech Emotion Recognition
- Open Problems and Future Directions

IEEE ICME 2020

Voice-Indistinguishability

Protecting Voiceprint in Privacy-Preserving Speech Data Release

Yaowei Han, Sheng Li, Yang Cao, Qiang Ma, Masatoshi Yoshikawa
Department of Social Informatics, Kyoto University, Kyoto, Japan
National Institute of Information and Communications Technology, Kyoto, Japan

CONTENT

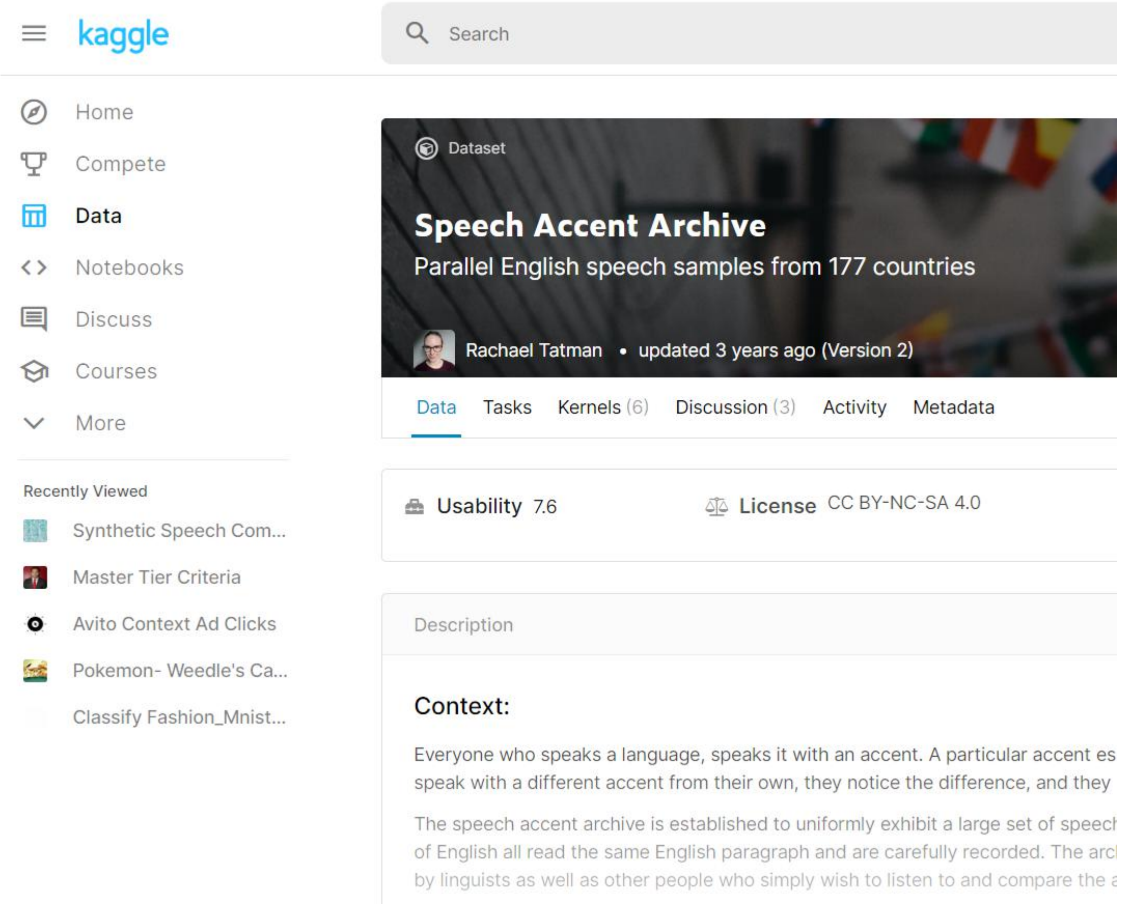
01 Motivation

Motivation - Speech Data Release

Speech Data Release

Share speech dataset with the 3rd parties

Eg. Apple collects speech data for Siri quality evaluation process, which they call grading.



Motivation - Risks of Speech Data Release

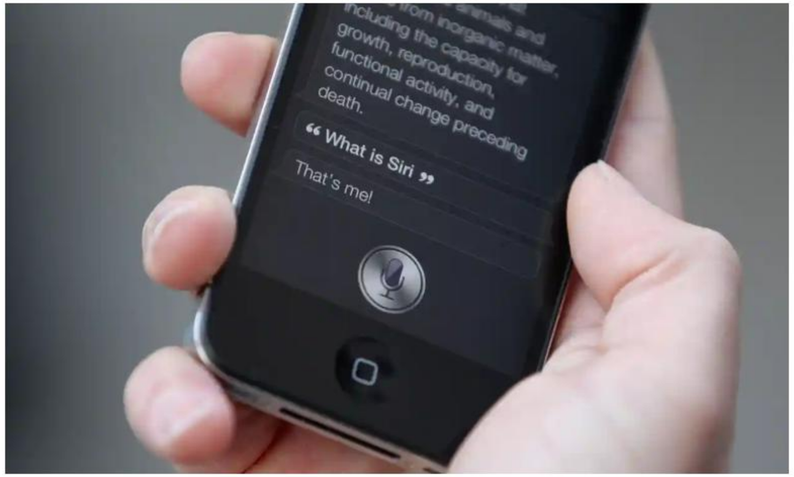
Risks of Speech Data Release

Privacy concern.

- Speech data is personal data.
- Everybody has a unique voiceprint, which is a kind of biometric identifiers.
- GDPR[1] bans the sharing of biometric identifiers.

Apple contractors 'regularly hear confidential details' on Siri recordings

Workers hear drug deals, medical details and people having sex, says whistleblower



[1] A. Nautsch and et al., “The GDPR & speech data:Reflections of legal and technology communities, firststeps towards a common understanding,” 2019.
https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings

Motivation - Risks of Speech Data Release

Risks of Speech Data Release

Security risks.

- Spoofing attacks to the voice authentication systems
- Reputation attacks (fake Obama speech[1])

How to protect privacy in speech data release?

[1] S. Suwajanakorn and et al., "Synthesizing obama: learning lip sync from audio," ACM Transactions on Graphics, 2017.

02 Related Works

Related Works

	Privacy		Voice technology
	protection level	privacy guarantee	
[1][2]	voice-level	ad-hoc	Vocal Tract Length Normalization (VTLN)
[3][4]	feature-level	k-anonymity	Speech Synthesize
[5]	model-level	ad-hoc	ASR

[1] J. Qian and et al., "Hidebehind: Enjoy voice input with voiceprint unclonability and anonymity," in ACM SenSys 2018.
[2] B. Srivastava and et al., "Evaluating voice conversion-based privacy protection against informed attackers," arXiv preprint arXiv:1911.03934, 2019.
[3] T. Justin and et al., "Speaker deidentification using diphone recognition and speech synthesis," in FG 2015.
[4] F. Fang and et al., "Speaker anonymization using X-vector and neural waveform models," in 10th ISCA Speech Synthesis Workshop, 2019.
[5] B. Srivastava and et al., "Privacy-Preserving Adversarial Representation Learning in ASR: Reality or Illusion?," in Interspeech 2019.

Related Works - Insufficiency of Existing Methods

Existing methods for protecting speech data privacy

(1) Speech2text (2) K-anonymity

However, they are insufficient because

(1) Speech2text

not useful for speech analysis

without any formal privacy guarantee

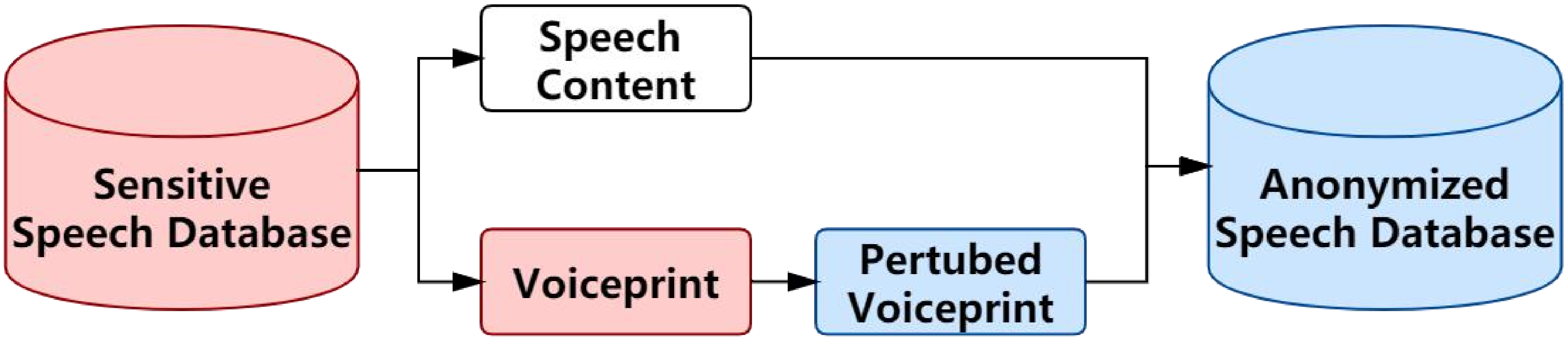
(2) K-anonymity

based on the assumption of attackers' knowledge

(= not secure under powerful attackers)

03 Problem Setting and Contributions

Problem Setting



Privacy-preserving speech data release

We focus on protecting voiceprint, i.e., user voice identity.

Contributions

1 **How to formally define voiceprint privacy?**

Voice-Indistinguishability

- The first formal privacy definition for voiceprint, not depend on attacker's background knowledge.

How to design a mechanism achieving our privacy definition? 2

Voiceprint perturbation mechanism

- Use voiceprint to present user voice identity
- Our mechnism output a anonymized voiceprint

3 **How to implement frameworks for private speech data release?**

Privacy-preserving speech synthesis

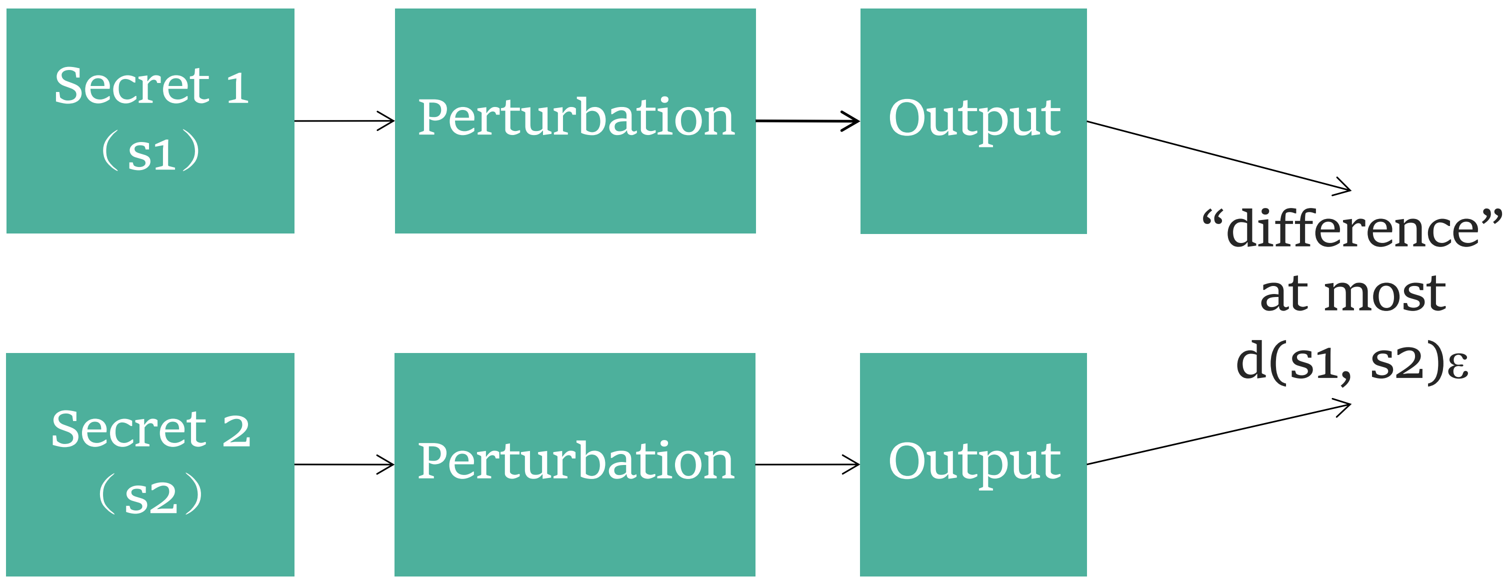
- Synthesize voice record with anonymized voiceprint

04 Our Solution

Our Solution - Metric Privacy

How to formally define voiceprint privacy?

Definition of Metric Privacy



Advantages:

1) Has no assumptions on the attackers’ background knowledge.
2) Privacy loss can be quantified.
   the bigger ε -> the better utility, the weaker privacy
3) d(s1, s2): distance metric between secrets.

Our Solution - Decision of Secrets

When applying metric privacy, we should decide secrets and distance metric.

- What's the secret?

  Voiceprint

- How to represent the voiceprint?

  x-vector[1], a widely used speaker space vector.

  For example. 512 dimensional

  [1.291081 0.9634209 ... 2.59955]

[1] D. Snyder and et al., "X-vectors: Robust dnn embeddings for speaker recognition," inProc. IEEE-ICASSP,2018, pp. 5329–5333.

Our Solution - Decision of Distance Metric

When applying metric privacy, we should decide secrets and distance metric.

- How to define the distance metric between voiceprint?

Euclidean distance? ✕

Can not well represent the distance between two x-vectors

Cosine distance? ✕

Widely used in speaker recognition but doesn't satisfy triangle inequality

Angular distance? YES

Also a kind of cosine distance but satisfies triangle inequality

Our Solution - Voice-Indistinguishablility

How to formally define voiceprint privacy?

For single user

Voice-Indistinguishability, Voice-Ind

$$\frac{\Pr(\tilde{x}|x)}{\Pr(\tilde{x}|x')} \leq e^{\epsilon d_{\mathcal{X}}(x,x')}$$

$$d_{\mathcal{X}} = \frac{\arccos(\cos\ similarity < x, x' >)}{\pi}$$

For multiple users in a speech dataset

Speech Data Release under Voice-Ind

$$\frac{\Pr(\tilde{D}|D)}{\Pr(\tilde{D}|D')} \leq e^{\epsilon d(D,D')}$$

$$d(D, D') = d_{\mathcal{X}}(x, x')$$

ε: privacy budget

privacy-utility tradeoff

bigger ε :

(1) weaker privacy

(2) better utility

n: speech database size

larger n:

(1) stronger privacy

-> later, we will verify this

Our Solution - Mechanism

How to design a mechanism achieving our privacy definition?

$$\Pr(\tilde{x}|x_0) \propto e^{-\epsilon d_{\mathcal{X}}(x_0, \tilde{x})}$$

Original \ Pertubed	A	B	C
A	$\propto e^{0}$	$\propto e^{d(A, B)}$	$\propto e^{d(A, C)}$
B	$\propto e^{d(A, B)}$	$\propto e^{0}$	$\propto e^{d(B, C)}$
C	$\propto e^{d(A, C)}$	$\propto e^{d(B, C)}$	$\propto e^{0}$

Our Solution - Privacy Guarantee

Privacy guarantee of the released private speech database.

Sensitive Speech database

Speaker	Speech Data	Attr
A	Record 1	...
B	Record 2	...
C	Record 3	...
...	...	...

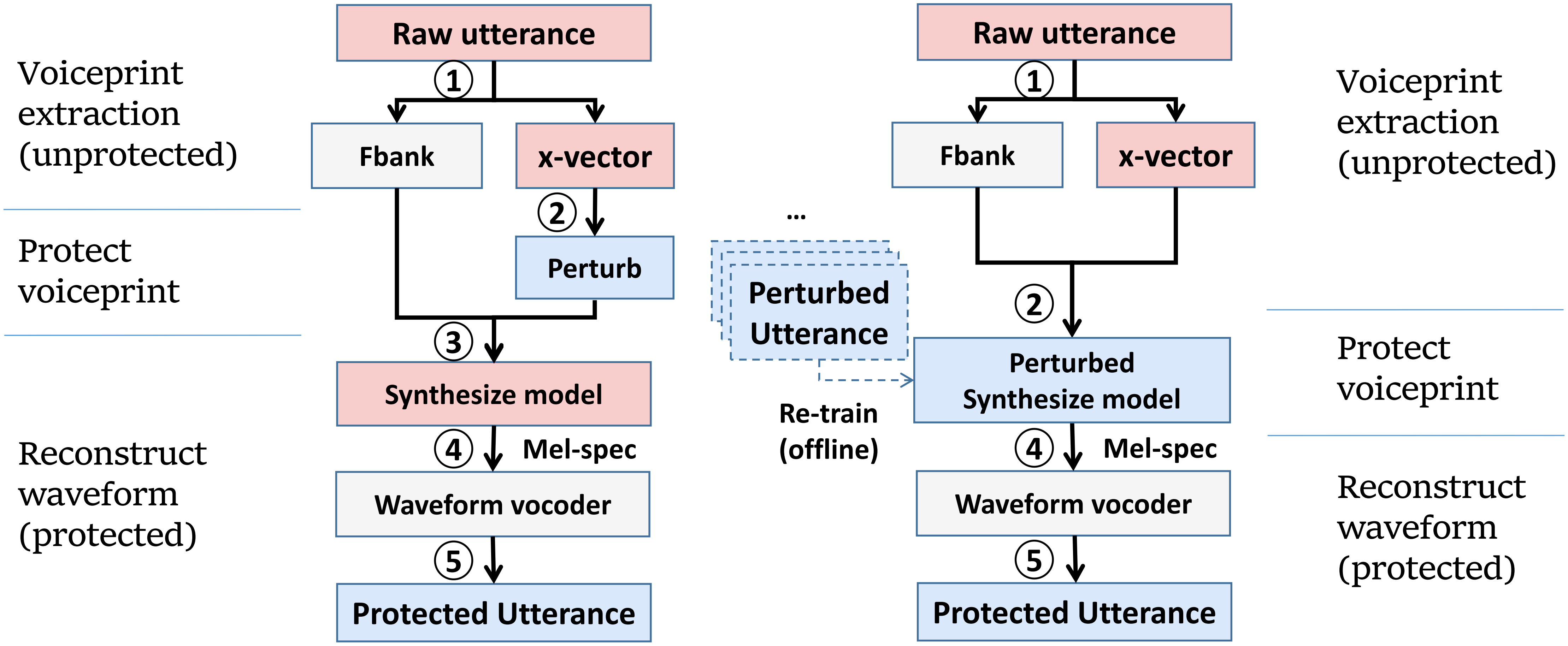
Our Method →

Anonymized Speech database

Speaker	Speech Data	Attr
A	Record 1 (with C's voiceprint)	...
B	Record 2 (with A's voiceprint)	...
C	Record 3 (with B's voiceprint)	...
...	...	...

Our Solution

How to implement frameworks for private speech data release?



(a) Feature-level (b) Model-level

05 Experiment and Conclusion

Experiment

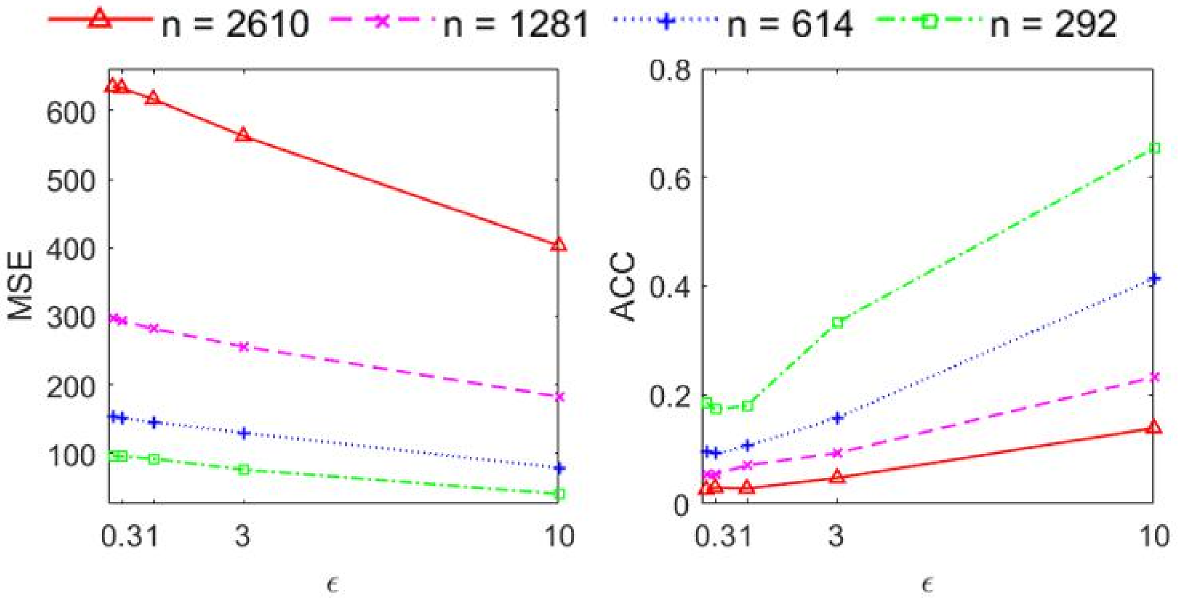
Verify the utility-privacy tradeoff of Voice-Indistinguishability.

- How does the privacy parameter ε affect the privacy and utility?
- How does the database size n affect the privacy?

Experiment

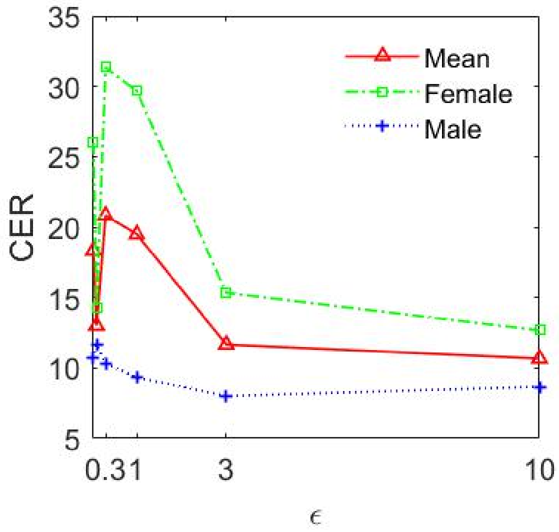
(**Objective** evaluation.)

Protected speech data with bigger ε -> (1) weaker privacy (2) better utility



MSE vs. ε

(PLDA) ACC vs. ε



CER vs. ε

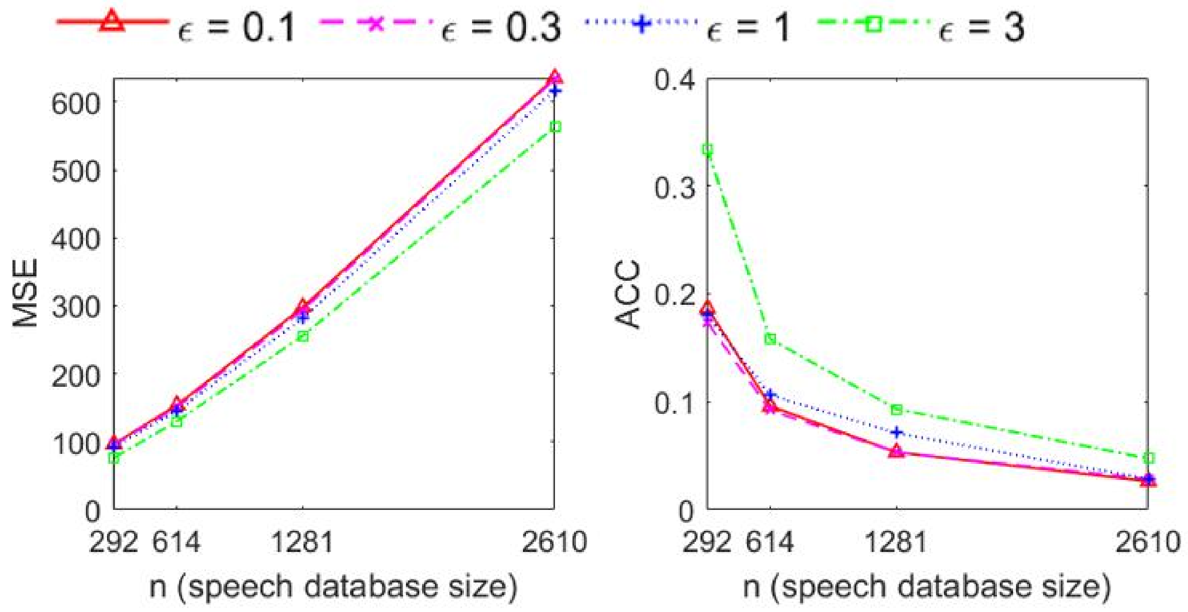
MSE: the difference before and after modification
lower MSE -> weaker privacy
(PLDA) ACC: the accuracy of speaker verification
higher ACC -> weaker privacy

CER: the performance of speech recognition
lower CER -> better utility

Experiment

(Objective evaluation.)

Protected speech data with larger n -> (1) stronger privacy



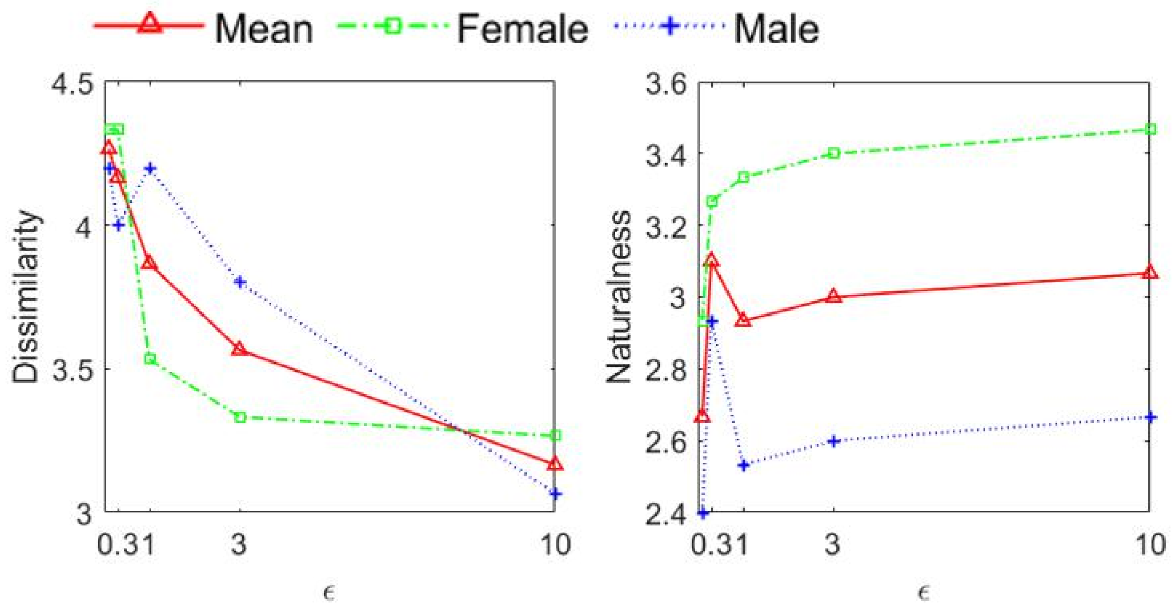
MSE vs. n (PLDA) ACC vs. n

MSE: the difference before and after modification
lower MSE -> weaker privacy
(PLDA) ACC: the accuracy of speaker verification
higher ACC -> weaker privacy

Experiment

(Subjective evaluation.) 15 speakers

Protected speech data with bigger ε -> (1) weaker privacy (2) better utility



Dissimilarity vs. ε   Naturalness vs. ε

Dissimilarity: the voice's differences between and after the modification

lower Dissimilarity -> weaker privacy

Naturalness: the naturalness of sounds that closely resemble the human voice

higher Naturalness -> better utility

Conclusion and Future work

Conclusion:

- Voice-Ind is the first formal privacy notion for voiceprint privacy.
- Our mechanism serves as a primitive to achieve voice-ind.
- Our end-to-end frameworks provide a good privacy-utility trade-off.

Future Works:

- Apply Voice-ind in Virtual Assistant, speech data processing, etc.
- Extend Voice-Ind for speech content privacy.

GENERAL OR SPECIFIC? INVESTIGATING EFFECTIVE PRIVAC PROTECTION IN FEDERATED LEARNING FOR SPEECH EMOTION RECOGNITION

Chao Tan (Kyoto U), Yang Cao (Hokkaido U), Sheng Li (NICT), Masatoshi Yoshikawa (Osaka Seike U)

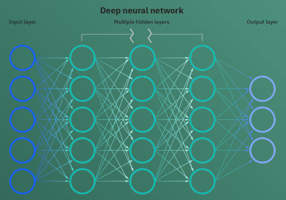
Outlines

1. Background
2. Related work
3. Proposed method
4. Empirical analysis
5. Conclusion and future work

Collecting data is harder

- Deep learning needs much more data.
- It is hard to collect data because of the law and the privacy awareness of people.
- Federated learning [1] doesn't need to collect data.

Deep learning



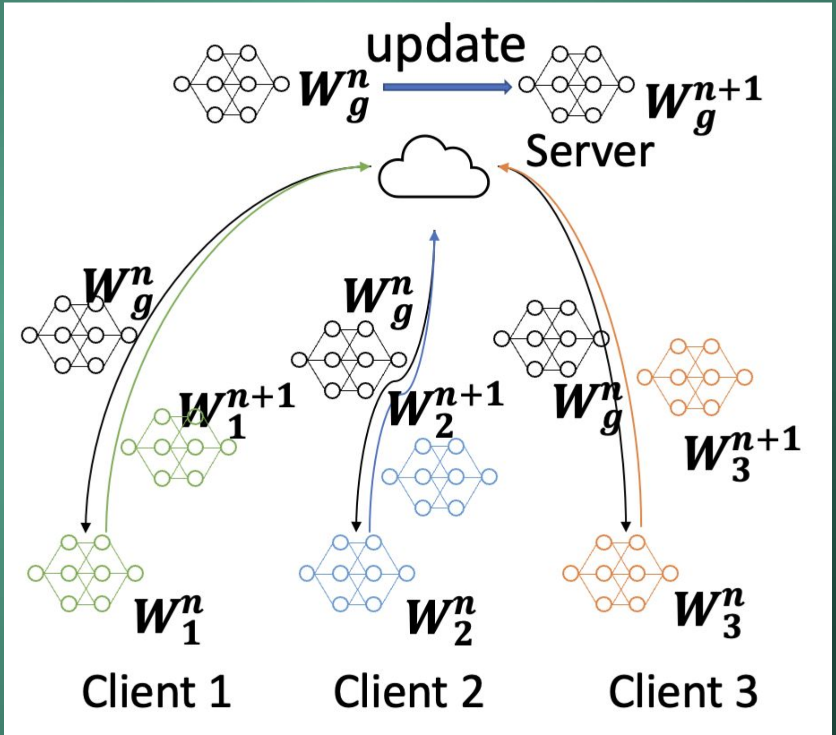
Need more data

But hard to collect data because of the law, and people are worried about their privacy.

Federated Learning

Google proposed a solution that doesn't need data leaving the user's device.

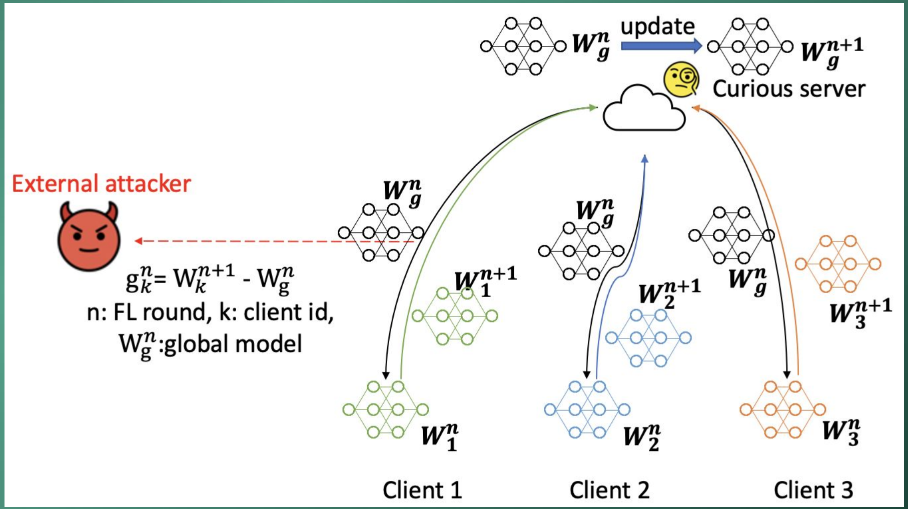
[1] Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving communication efficiency[J]. arXiv preprint arXiv:1610.05492, 2016.



Preliminary of FL [1]

- A cycle of federated learning
  - 1. Server generates a global model W_g^n
  - 2. Server distributes global model to clients.
  - 3. clients do local training and update local models to server.
  - 4. server updates global model according to these local models.
- Clients' data has never left local device.
- FL still not totally safe.

Preliminary of attack in FL



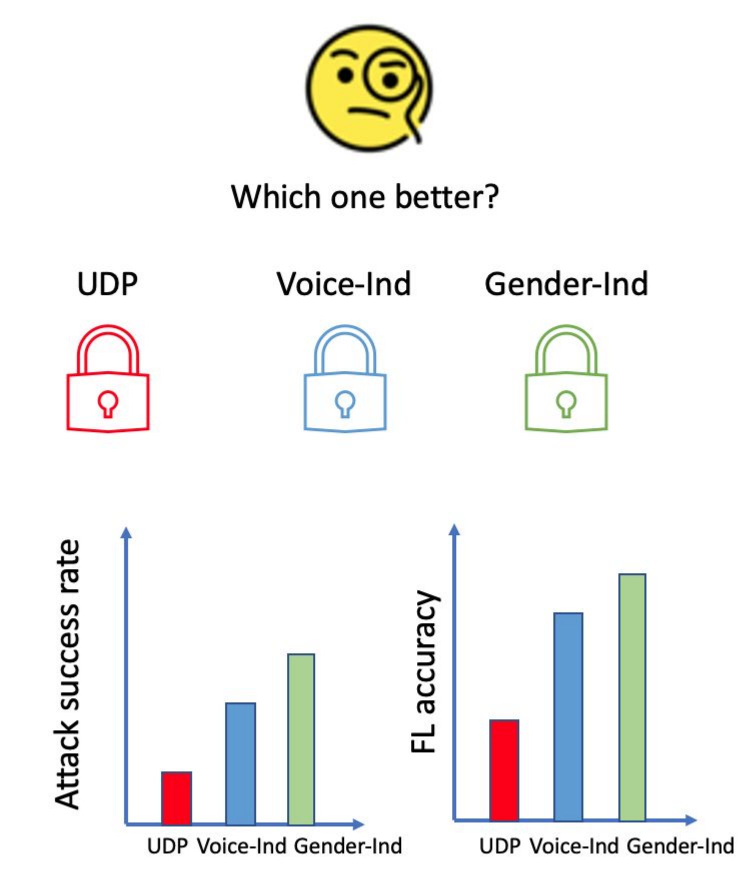
- FL does not provide strict privacy protection and still have privacy problem. [2, 3]
- Curious server and external attacker might threaten privacy.
- We focus on the Property Inference attack.

[2] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020.
[3] Melis L, Song C, De Cristofaro E, et al. Exploiting unintended feature leakage in collaborative learning[C]//2019 IEEE Symposium on Security and Privacy (SP). IEEE, 2019: 691-706.

Motivation of this work

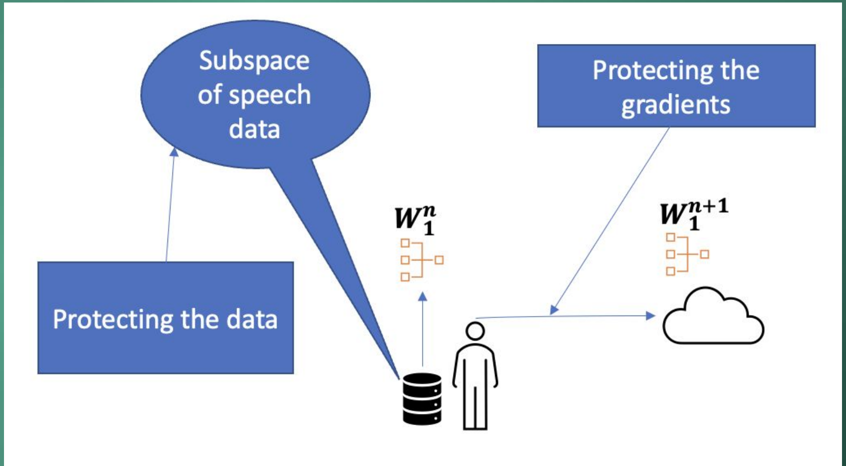
Knowledge gap on effectiveness between different privacy protection methods

- There are general methods (UDP) and specific designed methods (Voice-Ind, Gender-Ind).
- General or Specific design?
- No study told us which one is better in speech-federated learning.



Outlines

1. Background
2. Related work
3. Proposed method
4. Empirical analysis
5. Conclusion and future work



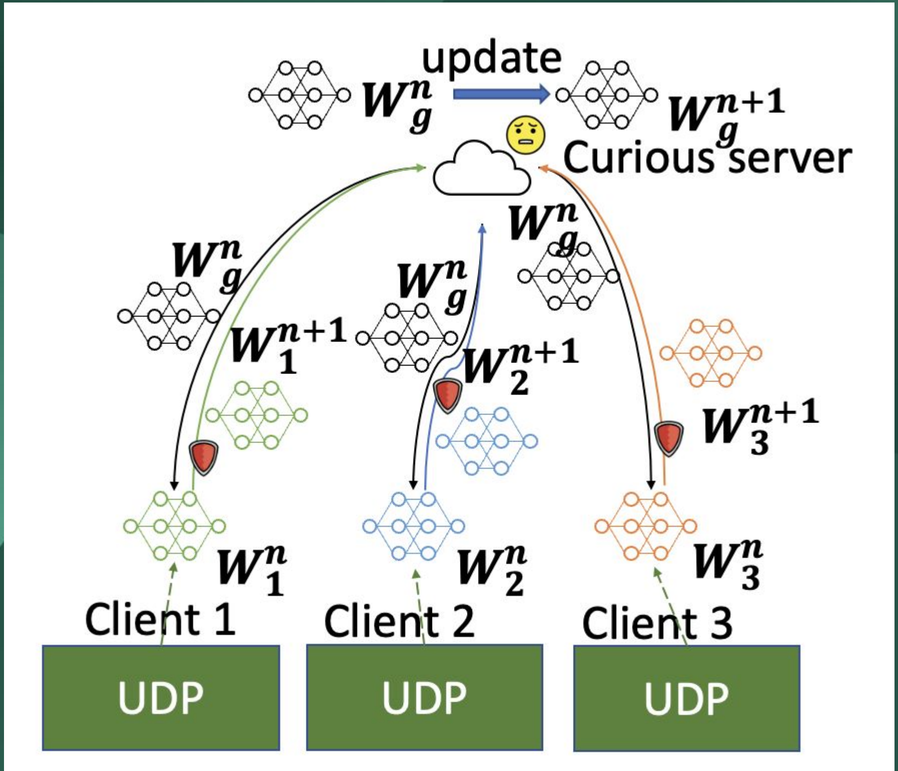
Two kinds of protection methods

- General method:
  - User-level Differential Privacy (UDP) [4]
- Specific method:
  - Voice-indistinguishability (Voice-Ind) [5]

[4] Feng T, Peri R, Narayanan S. User-Level Differential Privacy against Attribute Inference Attack of Speech Emotion Recognition in Federat
Learning[J]. arXiv preprint arXiv:2204.02500, 2022.
[5] Han Y, Li S, Cao Y, et al. Voice-indistinguishability: Protecting voiceprint in privacy-preserving speech data release[C]//2020 IEEE Internat
Conference on Multimedia and Expo (ICME). IEEE, 2020: 1-6.

UDP (User-level DP) [4]

- Step 1: Obtain parameters $\boldsymbol{w_i}$ through local training.
- Step 2: Perturb gradients $\boldsymbol{w_i}$ according to LDP parameter $(\boldsymbol{\varepsilon_i}, \boldsymbol{\delta_i})$ and some other factors to get $\widetilde{w}_i$.
- Step 3: Upload parameters $\widetilde{w}_i$.



Voice-Indistinguishability [5]

- Step 1: Separate raw utterance $\boldsymbol{s}$ to Fbank $\boldsymbol{f}$ and x-vector $\boldsymbol{x}$.
- Step 2: Change x-vector $\boldsymbol{x}$ to $\tilde{x}$ with a probability according to the cosine distance between $\boldsymbol{x}$ and x-vectors in pool $\mathcal{X}_p$.
- Step 3: Synthesis utterance $\tilde{s}$ with Fbank $\boldsymbol{f}$ and perturbed x-vector $\tilde{x}$.

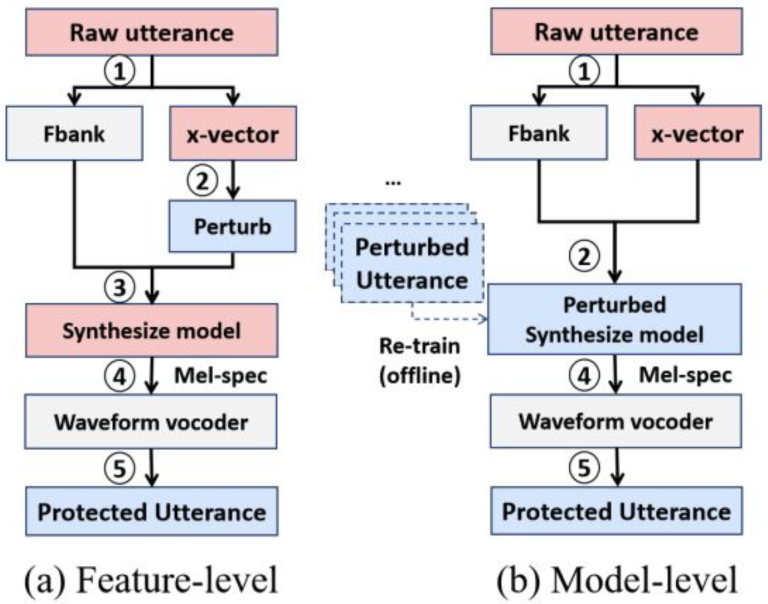


Fig. 3. The proposed frameworks.

Outlines

Privacy notion: Gender-Indistinguishability (Gender-Ind)

- A mechanism $\boldsymbol{M_g}$ satisfies ϵ-Gender-Indistinguishability if for any output gender embedding $\widetilde{\boldsymbol{h}}$ and any two possible input $\boldsymbol{h}, \boldsymbol{h'} \in \mathcal{H}$:

$$\frac{\Pr(\mathcal{M}_g(h) = \tilde{h})}{\Pr(\mathcal{M}_g(h') = \tilde{h})} \leq e^{\epsilon \boldsymbol{d}_{\mathcal{H}}(\boldsymbol{h}, \boldsymbol{h'})}$$

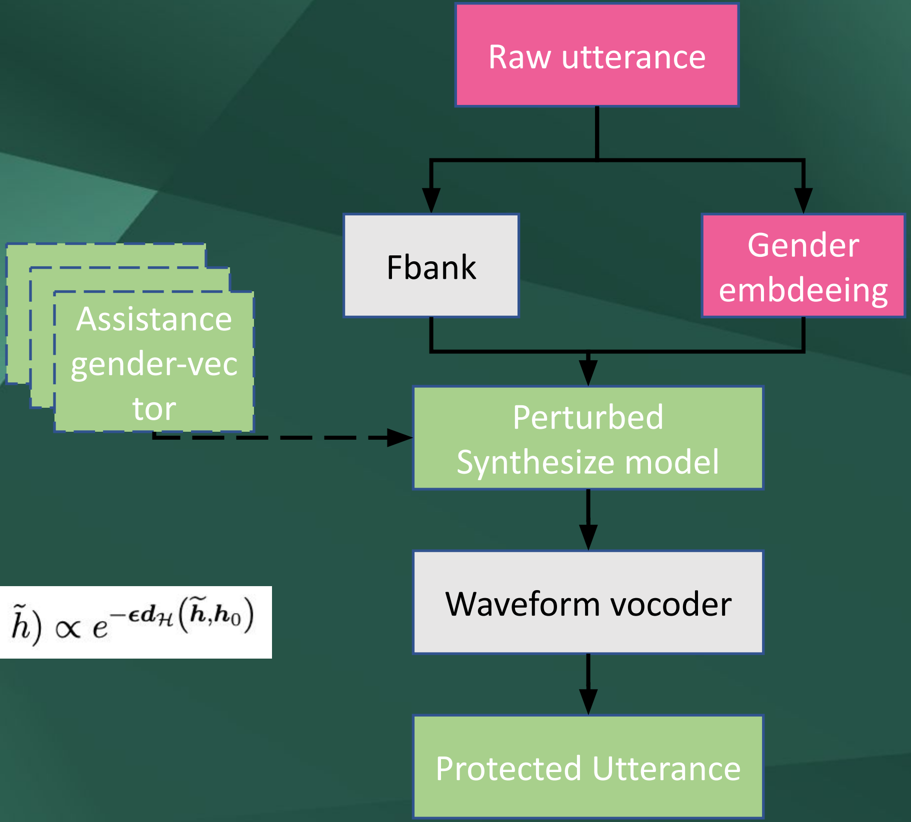
where $\mathcal{H}$ is a set of gender embedding in public datasets, $\boldsymbol{d_{\mathcal{H}}(h, h')}$ represents the angular distance between $\boldsymbol{h}$ and $\boldsymbol{h'}$.

Gender embedding protection method

- Step 1: Separate raw utterance $\boldsymbol{s}$ to Fbank $\boldsymbol{f}$ and gender embedding $\boldsymbol{h}$.
- Step 2: Change gender embedding $\boldsymbol{h}$ to $\tilde{h}$ with a probability according to the angular distance between $\boldsymbol{h}$ and gender embedding in pool $\mathcal{H}_p$.

$$\Pr(\mathcal{M}_g(h_0) = \tilde{h}) \propto e^{-\epsilon d_{\mathcal{H}}(\tilde{\boldsymbol{h}}, \boldsymbol{h}_0)}$$

- Step 3: Synthesis utterance $\tilde{s}$ with Fbank $\boldsymbol{f}$ and perturbed gender embedding $\tilde{h}$.



Outlines

1. Background

2. Related work

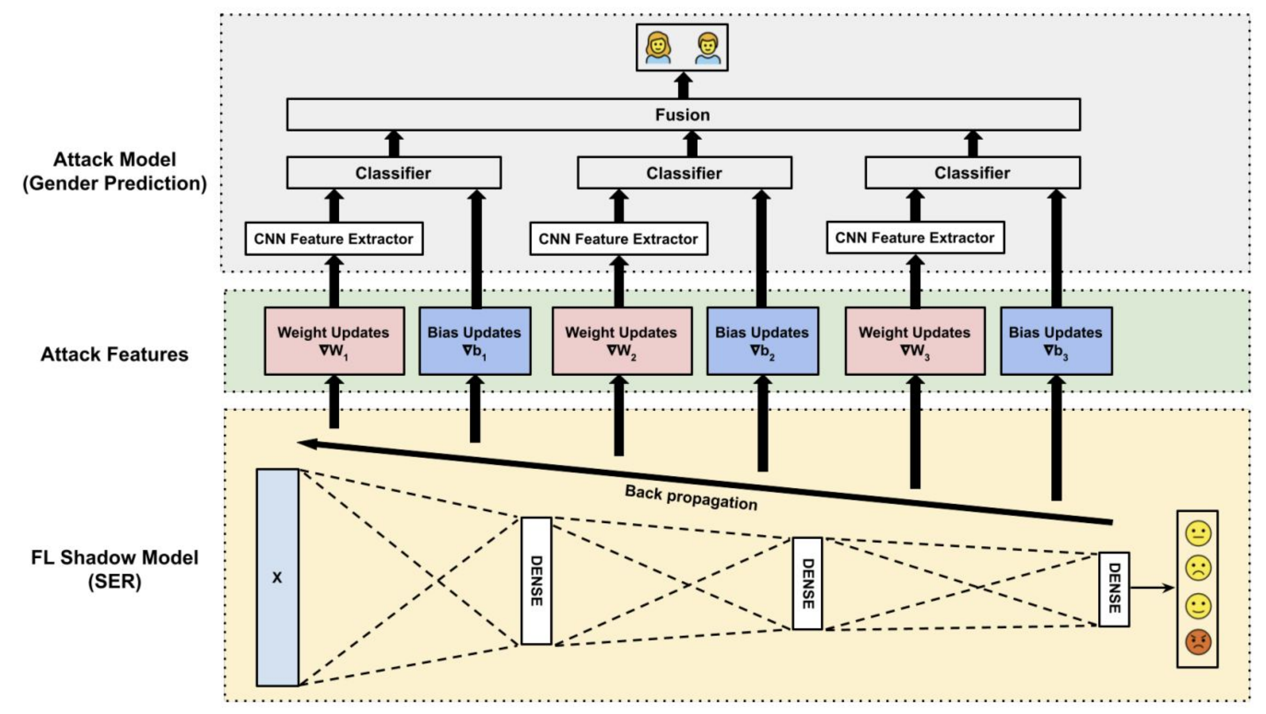
3. Proposed method

4. Empirical analysis

5. Conclusion and future work

Specifical FL and attack

- FL: SER model
- Attack: Steal personal information (gender, age) from gradients of SER-FL



Framework of FL and Attack Model [6]

[6] Feng T, Hashemi H, Hebbar R, et al. Attribute inference attack of speech emotion recognition in federated learning settings[J]. arXiv preprint arXiv:2112.13416, 2021.

Attack model success rate and FL accuracy (without protection)

Table 1. The success rate of attack and accuracy of FL-SER model without protections. (ACC: Accuracy; UAR: Unweighted Average Recall; Fold: training subsets, the random factors to order user's data; SR: Success Rate; UASR: Unweighted Average Success Recall

	Attack model		FL-SER model	
	SR	UASR	ACC	UAR
Fold1	0.837	0.829	0.663	0.595
Fold2	0.847	0.838	0.666	0.601
Fold3	0.817	0.791	0.656	0.619

$$\mathrm{ACC} = \frac{preditedReal_{true}}{predicted_{true}} * \frac{real_{true}}{total} + \frac{preditedReal_{false}}{predicted_{false}} * \frac{real_{false}}{total}$$

$$\mathrm{UAR} = \frac{preditedReal_{true}}{predited_{true}} + \frac{preditedReal_{false}}{predited_{false}}$$

$$\mathrm{SR} = \frac{preditedReal_{male}}{predited_{male}} * \frac{real_{male}}{total} + \frac{preditedReal_{female}}{predited_{female}} * \frac{real_{female}}{total}$$

$$\mathrm{UASR} = \frac{preditedReal_{male}}{predited_{male}} + \frac{preditedReal_{female}}{predited_{female}})$$

Comparison between protection methods for FL

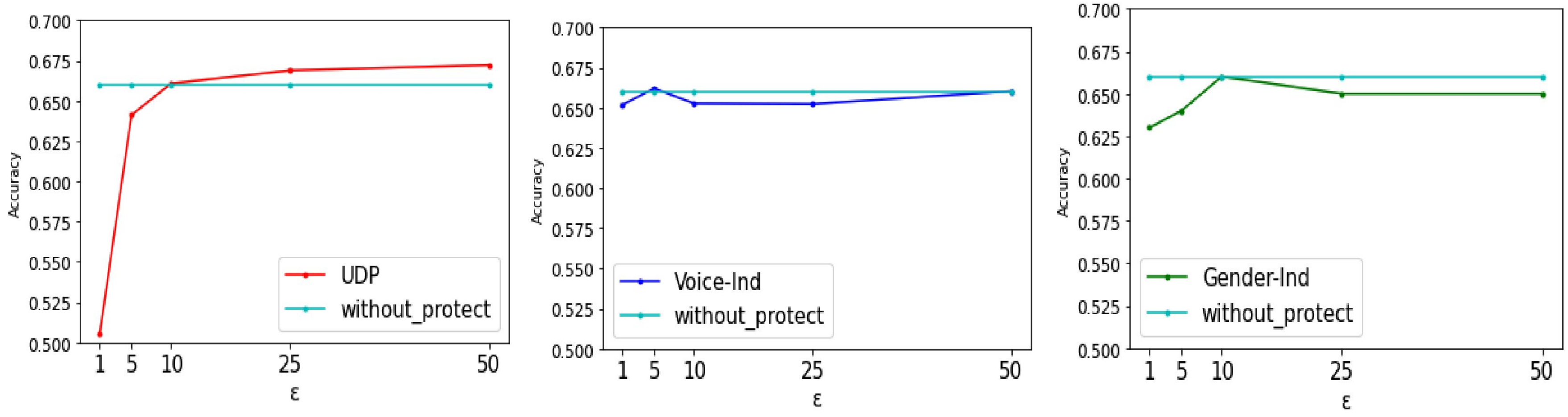


Figure 12: The accuracy of FL-SER model with UDP

Figure 13: The accuracy of FL-SER model with Voice-Ind

Figure 14: The accuracy of FL-SER model with Gender-Ind

FL accuracy

Voice-Ind and Gender-Ind have better model accuracy than UDP

Comparison between protection methods for gender attack

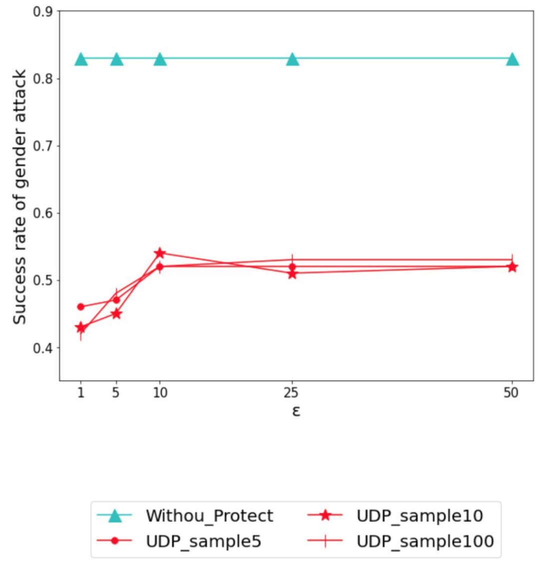


Figure 9: The success rate of gender attack model with UDP

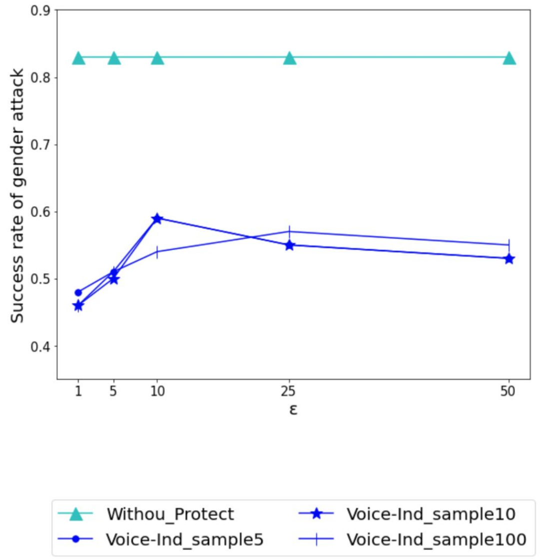


Figure 10: The success rate of gender attack model with Voice-Ind

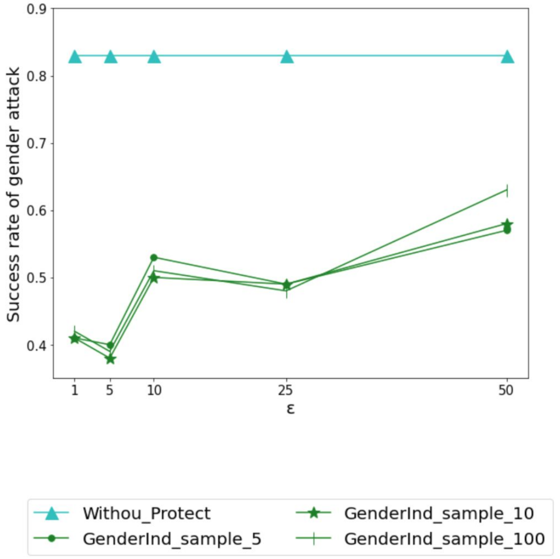


Figure 11: The success rate of gender attack model with Gender-Ind

Gender Attack model success rate

all of them decrease the attacker's success rate to around 50%, which is similar to a random guess

Outlines

Conclusion and future work

- Conclusion
  - Specifically, designed protection method gives better effectiveness in speech-FL.

- Future work
  - Expanded gender-Ind to attribute-Ind.

Outline

- Scenario and Motivation
  - why we need to formalize speech privacy?
- A brief history of privacy definitions
  - from k-Anonymity to Differential Privacy
- Our Studies for Formalizing Speech Privacy
  - [**ICME20**] Voice-Indistinguishability
  - [**ICASSP23**] General or Specific? Investigating Effective Speech Privacy Protection in Federated Learning for Speech Emotion Recognition
- Open Problems and Future Directions

Open Problems and Future Directions

- Theory of Speech Privacy
  - How to formalize privacy metrics for different types of "secrets" in speech processing?
  - Is there a Composition Theorem for speech privacy?
- Practice of Speech Privacy
  - How to understand the connection between Formal Privacy Metrics and Practical Attacks (i.e., Membership Inference Attacks, Gradient Reconstruction Attacks, etc).
  - How to define advanced private mechanisms for Formal Privacy Metrics (instead of using the building blocks like Laplace mechanisms)?

Acknowledgement

- The above two studies were primarily contributed by my collaborators and former students:
  - Dr. Sheng LI (NICT)
  - Yaowei HAN (Master student at Kyoto U) - ICME20
  - Chao TAN (Master student at Kyoto U) - ICASSP20
  - Prof. Masatoshi YOSHIKAWA (Osaka Seikei U)
  - Prof. Qiang MA (Kyoto Institute of Technology)

Thanks 😊

Q&A ❓

Looking forward to Collaborating on Speech Privacy 🤝