Privacy-Preserving Distributed Optimization via Subspace Perturbation: A General Framework

Qiongxiu Li, Richard Heusdens, Mads Græsbøll Christensen
2020/08/03

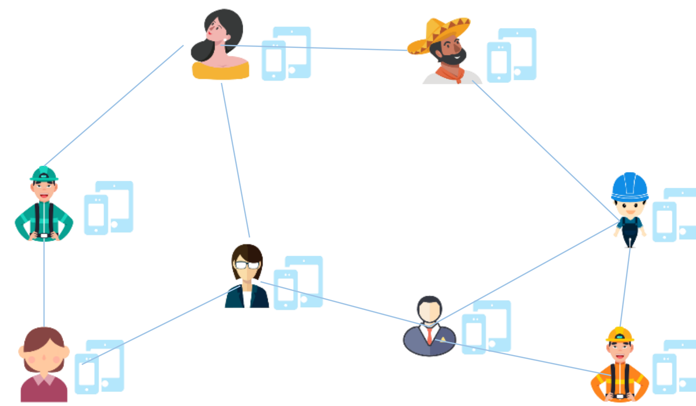
AALBORG UNIVERSITY
DENMARK

Agenda

- Motivation
- Problem setup
- State of the art
- Proposed approach
- Numerical validation
- Conclusions & Future works
- Q&A

Motivation

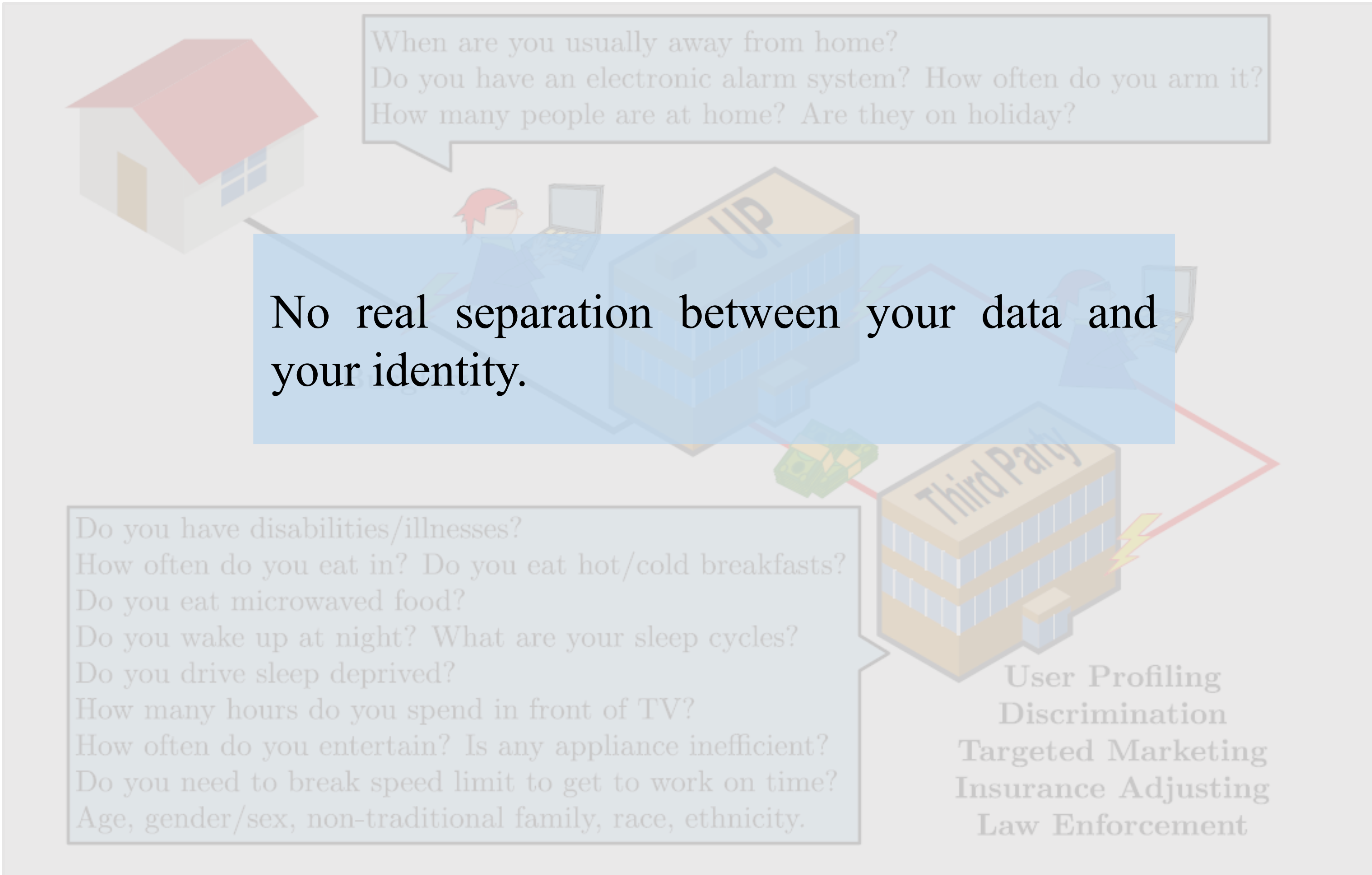
Centralized system v.s. Distributed system



☹ Totally dependent on the authority
☹ Vulnerable to malicious attack

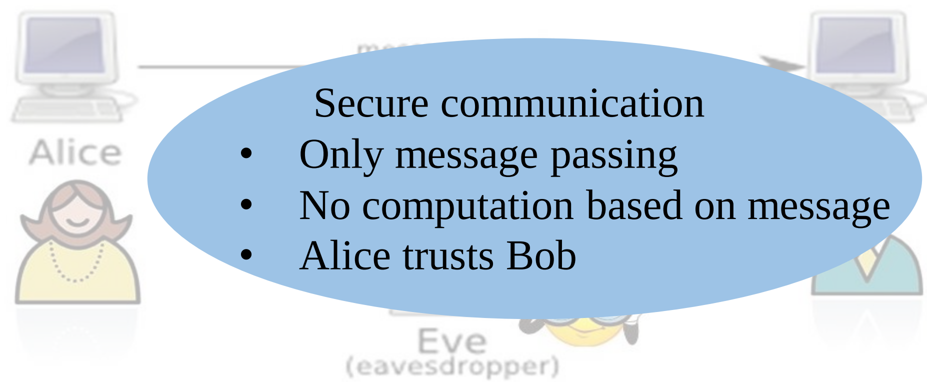
☺ No dependency on any single party
☺ More flexible system
☺ Robust to malicious attack

Privacy issue in smart meters [Giaconi, 2018]

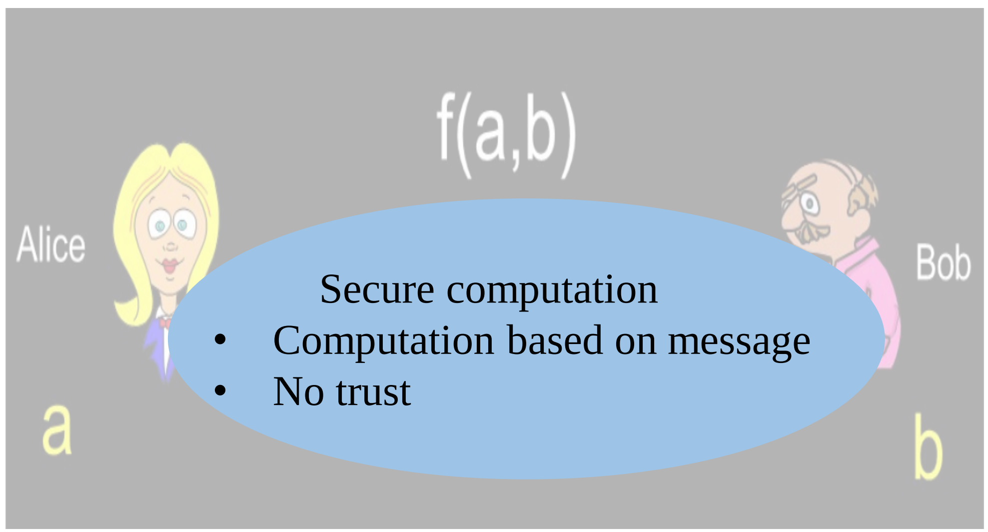


No real separation between your data and your identity.

Secure communication v.s. secure computation



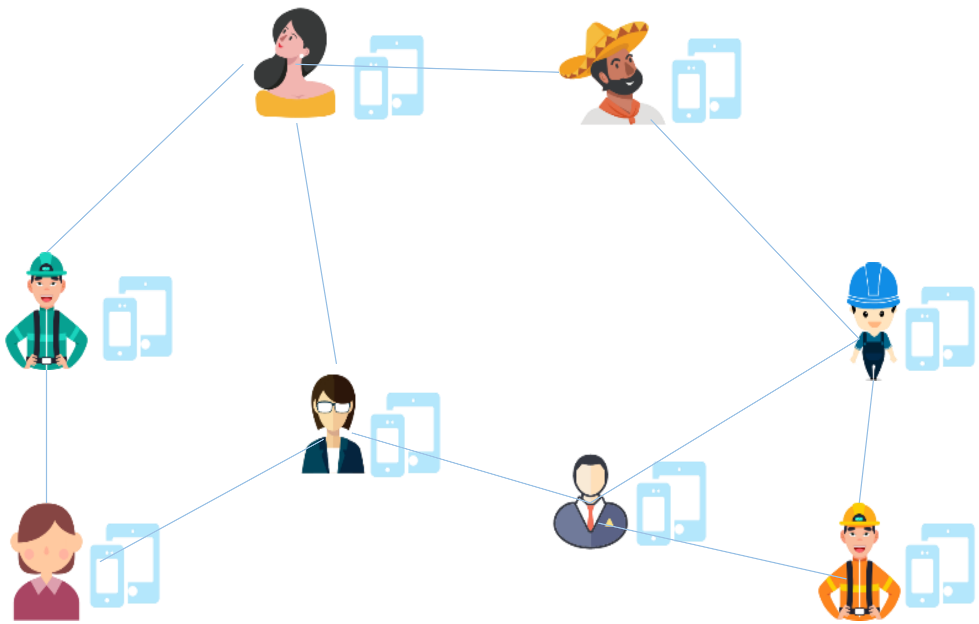
Secure communication
- Only message passing
- No computation based on message
- Alice trusts Bob



Secure computation
- Computation based on message
- No trust

Intuitive examples (1)

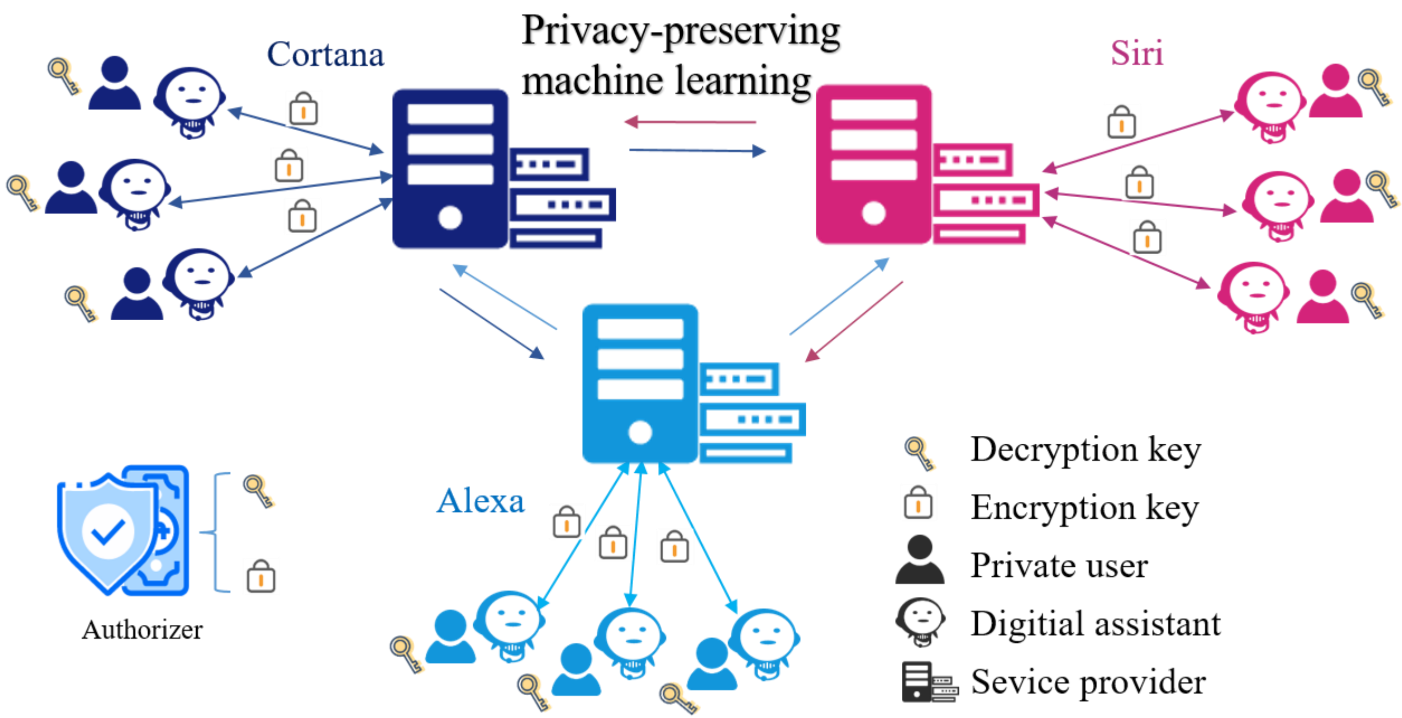
- How to securely compute the average salary over a group of people while keeping each person's own salary private from others?



AALBORG UNIVERSITY
DENMARK

Intuitive examples (2)

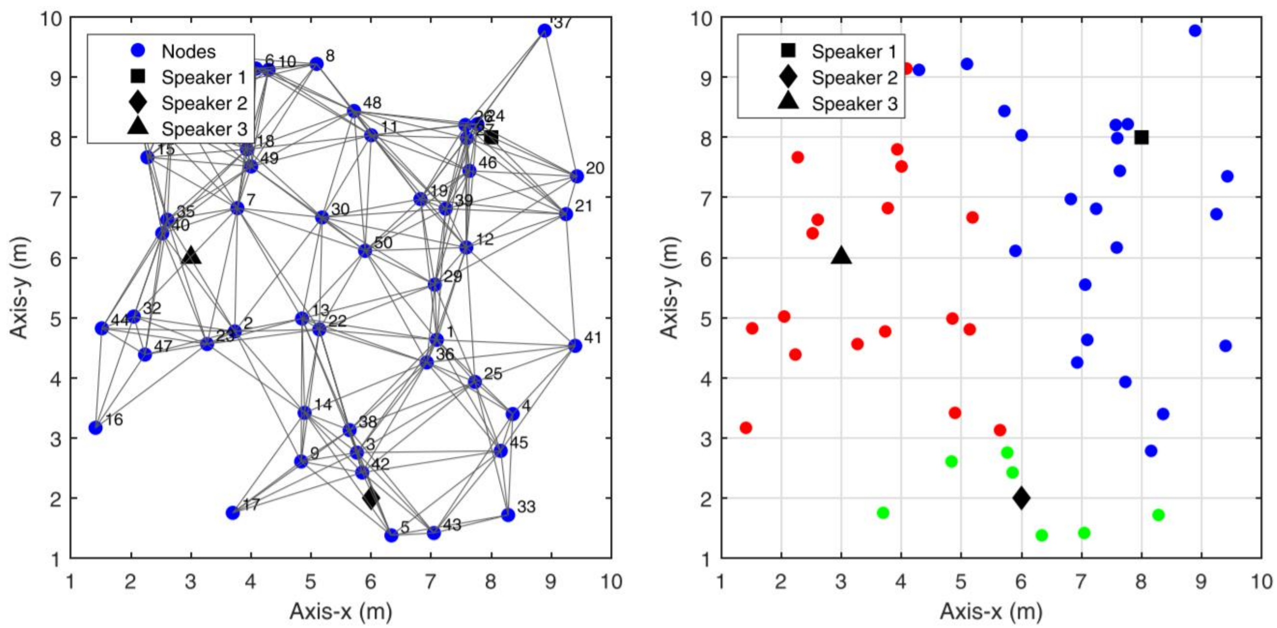
- Privacy-preserving machine learning over multiple parties
  - Collaborative learning without revealing private data



AALBORG UNIVERSITY
DENMARK

Intuitive examples (3)

- Privacy-preserving distributed acoustic environment classification in WASNs
    - Privacy-preserving distributed clustering



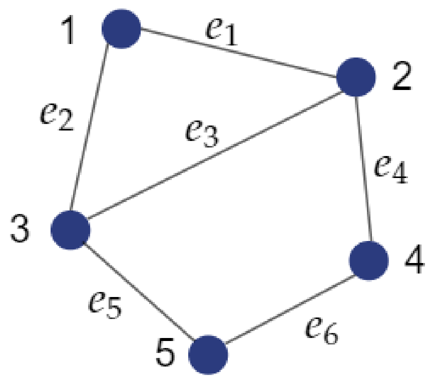
Zhao, Y., Nielsen, J. K., Chen, J., & Christensen, M. G. (2020). Model-based distributed node clustering and multi-speaker speech presence probability estimation in wireless acoustic sensor networks. *The Journal of the Acoustical Society of America*.

Problem setup

AALBORG UNIVERSITY
DENMARK

Distributed convex optimization over a network

A graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ $\mathcal{N} = \{1, 2, ..., n\}, n = |\mathcal{N}|,$ $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}, m = |\mathcal{E}|$



$$\min_{\boldsymbol{x}_i} \quad \sum_{i \in \mathcal{N}} f_i\left(\boldsymbol{x}_i, \boldsymbol{s}_i\right)$$

$$\text{s.t.} \quad \boldsymbol{B}_{i|j}\boldsymbol{x}_i + \boldsymbol{B}_{j|i}\boldsymbol{x}_j = \boldsymbol{b}_{ij} \ \forall (i,j) \in \mathcal{E}$$

$$b = \begin{pmatrix} b_{12} \\ b_{13} \\ b_{23} \\ b_{24} \\ b_{35} \\ b_{45} \end{pmatrix}$$

Incidence matrix: $\boldsymbol{B} \in \mathbb{R}^{m \times n}$

Main requirements for privacy-preserving distributed optimization

1. Output correctness: optimum result $\boldsymbol{x}_i^*$ should be achieved
2. Individual privacy: the private data $\boldsymbol{s}_i$ should not be revealed to others

AALBORG UNIVERSITY
DENMARK

Primal-Dual Methods of Multiplers (PDMM) [Sherson, 2018]

Extended augmented Lagrangian of PDMM

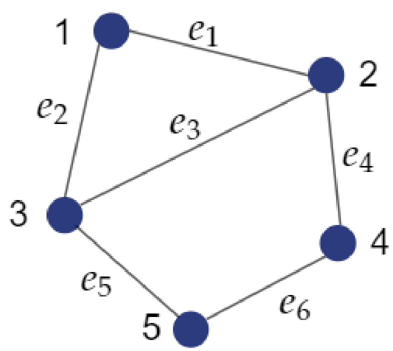
$$L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x},\boldsymbol{s}) + (\boldsymbol{P}\boldsymbol{\lambda}^{(k)})^\top \boldsymbol{C}\boldsymbol{x} + \frac{c}{2}\|\boldsymbol{C}\boldsymbol{x} + \boldsymbol{P}\boldsymbol{C}\boldsymbol{x}^{(k)} - 2\boldsymbol{d}\|_2^2$$

$c > 0$: constant for controlling convergence rate

$\boldsymbol{\lambda} \in \mathbb{R}^{2m}$: dual variables for constraints

Each edge $e_l = (i,j) \in \mathcal{E}$ corresponds two dual variables: $\lambda_{i|j}$ for node i, $\lambda_{j|i}$ for node j

$$\lambda = \begin{pmatrix} \lambda_{1|2} \\ \lambda_{1|3} \\ \lambda_{2|3} \\ \lambda_{2|4} \\ \lambda_{3|5} \\ \lambda_{4|5} \\ \lambda_{2|1} \\ \lambda_{3|1} \\ \lambda_{3|2} \\ \lambda_{4|2} \\ \lambda_{5|3} \\ \lambda_{5|4} \end{pmatrix}$$



Updating functions

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} L\left(\boldsymbol{x}, \boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)}\right)$$

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{P}\boldsymbol{\lambda}^{(k)} + c(\boldsymbol{C}\boldsymbol{x}^{(k+1)} + \boldsymbol{P}\boldsymbol{C}\boldsymbol{x}^{(k)} - 2\boldsymbol{d})$$

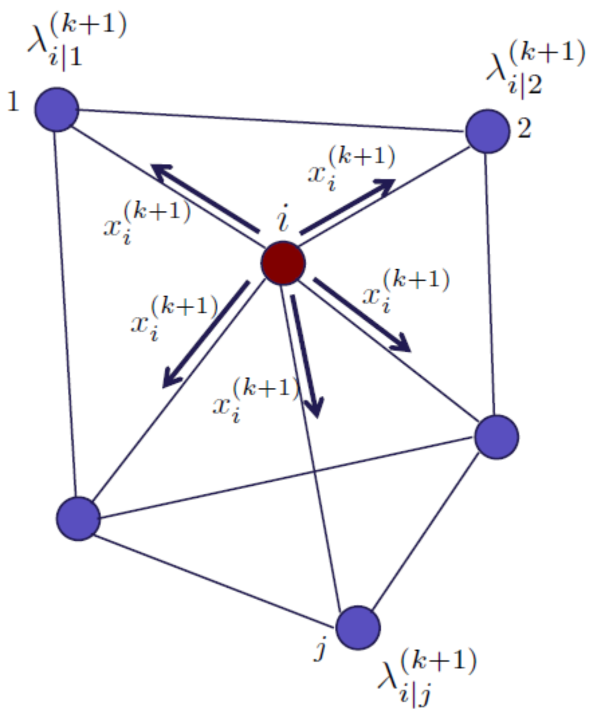
$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{B}^+ \\ \boldsymbol{B}^- \end{pmatrix}, \boldsymbol{P}\boldsymbol{C} = \begin{pmatrix} \boldsymbol{B}^- \\ \boldsymbol{B}^+ \end{pmatrix} \in \mathbb{R}^{2m \times n}$$

Why conventional approaches violate privacy?

Local updating functions of PDMM

$$\boldsymbol{x}_i^{(k+1)} = \arg\min_{\boldsymbol{x}_i} \left(\boxed{f_i(\boldsymbol{x}_i, \boldsymbol{s}_i)} + \sum_{j \in \mathcal{N}_i} \boldsymbol{\lambda}_{j|i}^{(k)\top} \boldsymbol{B}_{i|j} \boldsymbol{x}_i + \frac{c}{2} \sum_{j \in \mathcal{N}_i} \|\boldsymbol{B}_{i|j} \boldsymbol{x}_i + \boldsymbol{B}_{j|i} \boldsymbol{x}_j^{(k)} - \boldsymbol{b}_{i,j}\|_2^2 \right)$$

$$\forall j \in \mathcal{N}_i : \boldsymbol{\lambda}_{i|j}^{(k+1)} = \boldsymbol{\lambda}_{j|i}^{(k)} + c\left(\boldsymbol{B}_{i|j} \boldsymbol{x}_i^{(k+1)} + \boldsymbol{B}_{j|i} \boldsymbol{x}_j^{(k)} - \boldsymbol{b}_{i,j}\right)$$



$\partial f_i(\boldsymbol{x}_i^{(k+1)}, \boldsymbol{s}_i)$ is correlated with private data $\boldsymbol{s}_i$
Information-theoretically: $I(S_i; X_i^{(k+1)}) \neq 0$

Exchange of $\boldsymbol{x}_i^{(k+1)}$ violates individual privacy

AALBORG UNIVERSITY
DENMARK

State of the art

Existing approach (1)_Homomorphic encryption[Freris, 2016]

Main idea: all computation are conducted on encrypted data

$$Enc(x) \times Enc(y) = Enc(x + y)$$

Computational security model: based on computational hardness assumption

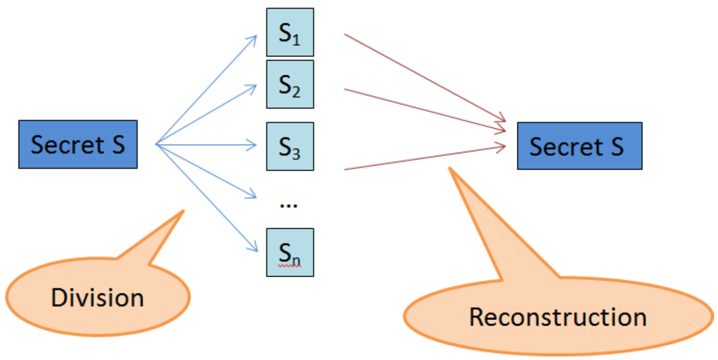
Pros and cons:

☺ No compromisation in algorithm accuracy
☹ Computationally complex
☹ The adversary is assumed computationally bounded

AALBORG UNIVERSITY
DENMARK

Existing approach (2)_Secret sharing[Tjell, 2020]

Main idea: split each message into pieces and send to different parties



Information-theoretic security model: the adversary does not have enough information to infer the secret/private data

Pros and cons:

☺ The adversary is assumed computationally unbounded
☺ Computationally simple
☺ No compromisation in algorithm accuracy
☹ Communication demanding

AALBORG UNIVERSITY
DENMARK

Existing approach (3)_Differential privacy [Nozari, 2018]

Main idea: obfuscate sensitive data before sharing to others

$$x' = x + r$$

Information-theoretic security model

Pros and cons:

☺ The adversary is assumed computationally unbounded
☺ Computationally simple
☺ Robust to n-1 number of corruptions
☹ Tradeoff between privacy and accuracy

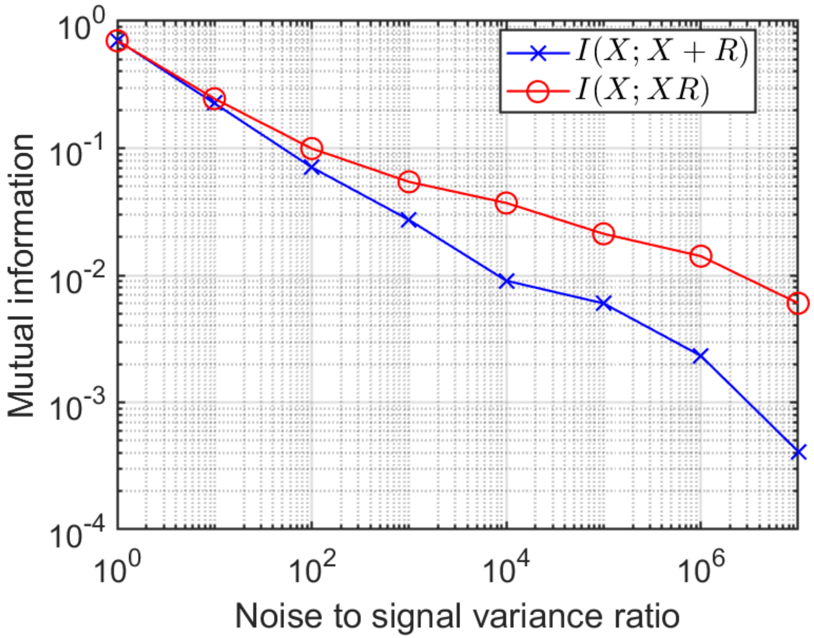
AALBORG UNIVERSITY
DENMARK

Proposed approach

AALBORG UNIVERSITY
DENMARK

Information-theoretic security using noise insertion

Private data x, inserted noise r



Normalized mutual information (i.e., information leakage) in terms of the amount of inserted noise for both additive and multiplicative cases

The more noise inserted, the less privacy leakage

Proposed approach

x-update of PDMM

$$\boldsymbol{x}_i^{(k+1)} = \arg\min_{\boldsymbol{x}_i} \left(f_i(\boldsymbol{x}_i, \boldsymbol{s}_i) + \sum_{j\in\mathcal{N}_i} \boldsymbol{\lambda}_{j|i}^{(k)\top} \boldsymbol{B}_{i|j}\boldsymbol{x}_i + \frac{c}{2}\sum_{j\in\mathcal{N}_i} \|\boldsymbol{B}_{i|j}\boldsymbol{x}_i + \boldsymbol{B}_{j|i}\boldsymbol{x}_j^{(k)} - \boldsymbol{b}_{i,j}\|_2^2 \right)$$

$$\mathbf{0} \in \partial f_i(\boldsymbol{x}_i^{(k+1)}, \boldsymbol{s}_i) + \boxed{\sum_{j\in\mathcal{N}_i} \boldsymbol{B}_{i|j}\boldsymbol{\lambda}_{j|i}^{(k)}} + c\sum_{j\in\mathcal{N}_i} (\boldsymbol{x}_i^{(k+1)} - \boldsymbol{x}_j^{(k)} - \boldsymbol{B}_{i|j}\boldsymbol{b}_{i,j})$$

Motivation:

Instead of inserting additional noise, why not exploit the dual variable as noise?

AALBORG UNIVERSITY
DENMARK

Convergence behavior of dual variable

Consider two successive λ-updates

$$\boldsymbol{\lambda}^{(t+2)} = \boldsymbol{\lambda}^{(t)} + c(\boldsymbol{C}\boldsymbol{x}^{(t+2)} + 2\boldsymbol{P}\boldsymbol{C}\boldsymbol{x}^{(t+1)} + \boldsymbol{C}\boldsymbol{x}^{(t)})$$

$$H = \mathrm{span}(\boldsymbol{C}) + \mathrm{span}(\boldsymbol{P}\boldsymbol{C}) \qquad \boldsymbol{\lambda}^{(t+2)} - \boldsymbol{\lambda}^{(t)} \in H$$

$$\boldsymbol{\lambda}^{(t)} = \Pi_H \boldsymbol{\lambda}^{(t)} + (\boldsymbol{I} - \Pi_H)\boldsymbol{\lambda}^{(t)}$$

$t \to \infty$ $\Downarrow$ converge $\Downarrow$ Only be permuted at every iteration

$$\boldsymbol{\lambda}^* \qquad \boxed{\boldsymbol{P}^{(t)}\left(\boldsymbol{I} - \Pi_{\bar{H}}\right)\boldsymbol{\lambda}^{(0)}}$$ Subspace noise

Non-convegence property will not affect the accuracy: $\boldsymbol{x} \to \boldsymbol{x}^*$
since $\left(\left(\boldsymbol{I} - \Pi_{\bar{H}}\right)\boldsymbol{\lambda}^{(0)}\right)^\top \boldsymbol{C}\boldsymbol{x} = 0$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}, \boldsymbol{s}) + \boxed{(\boldsymbol{P}\boldsymbol{\lambda}^{(k)})^\top \boldsymbol{C}\boldsymbol{x}} + \frac{c}{2}\|\boldsymbol{C}\boldsymbol{x} + \boldsymbol{P}\boldsymbol{C}\boldsymbol{x}^{(k)} - 2\boldsymbol{d}\|_2^2$$

AALBORG UNIVERSITY
DENMARK

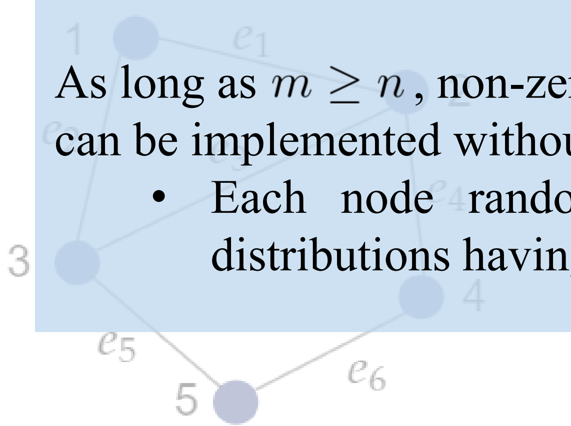
Non-empty subspace always exists

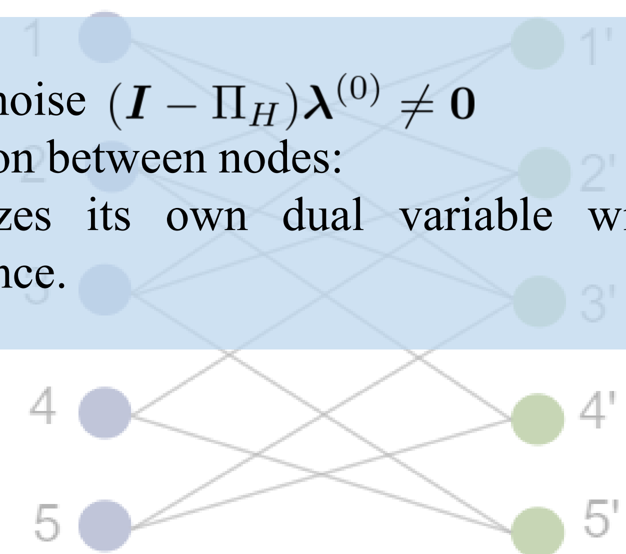
$H = \mathrm{span}(\boldsymbol{C}) + \mathrm{span}(\boldsymbol{PC})$

Incidence matrix: $\boldsymbol{B} \in \mathbb{R}^{m \times n}$

Incidence matrix:

$$[\boldsymbol{C} \quad \boldsymbol{PC}] = \begin{bmatrix} \boldsymbol{B}^+ & \boldsymbol{B}^- \\ \boldsymbol{B}^- & \boldsymbol{B}^+ \end{bmatrix} \in \mathbb{R}^{2m \times 2n}$$





As long as $m \geq n$, non-zero subspace noise $(\boldsymbol{I} - \Pi_H)\boldsymbol{\lambda}^{(0)} \neq \boldsymbol{0}$ can be implemented without coordination between nodes:

- Each node randomly initializes its own dual variable with distributions having large variance.

Incidence matrix of a graph is always rank dificient $\rightarrow \dim(H) < 2n$

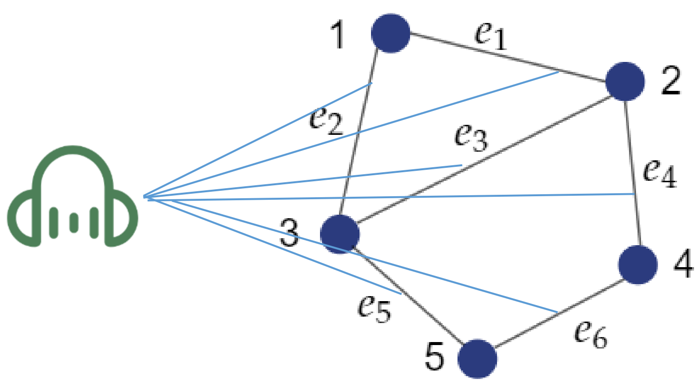
Robustness against adversary models (1)

Eavesdropping adversary model

- It eavesdrops all communication channels between nodes

Channel encryption cost (only one iteration)

- Only the transimission of initializied dual vairables needs channel encryption

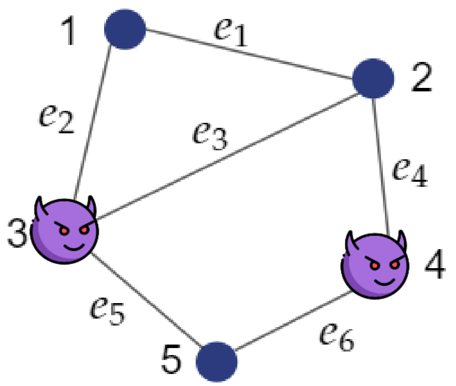


AALBORG UNIVERSITY
DENMARK

Robustness against adversary models (2)

Passive (honest-but-courious) adversary model

- Corrupted nodes follow the protocol but share information together to infer the private data of honest nodes



Conditions for privacy guarantee

- One honest neighbor is required $\mathcal{N}_{i,h} \neq \emptyset$

$$\mathbf{0} \in \partial f_i(\boldsymbol{x}_i^{(k+1)}, \boldsymbol{s}_i) + \boxed{\sum_{j \in \mathcal{N}_i} \boldsymbol{B}_{i|j} \boldsymbol{\lambda}_{j|i}^{(k)}} + c \sum_{j \in \mathcal{N}_i} (\boldsymbol{x}_i^{(k+1)} - \boldsymbol{x}_j^{(k)} - \boldsymbol{B}_{i|j} \boldsymbol{b}_{i,j})$$

$$\sum_{j \in \mathcal{N}_i} \boldsymbol{B}_{i|j} \boldsymbol{\lambda}_{j|i}^{(k)} = \boxed{\sum_{j \in \mathcal{N}_{i,c}} \boldsymbol{B}_{i|j} \boldsymbol{\lambda}_{j|i}^{(k)}} + \boxed{\sum_{j \in \mathcal{N}_{i,h}} \boldsymbol{B}_{i|j} \boldsymbol{\lambda}_{j|i}^{(k)}}$$

Known to the corrupted nodes

Unknown to the corrupted nodes

What if the adversary knows the subspace?

The dual variables of honest nodes cannot be inferred even though the subspace (whole graph toplogy) is known to the adversary:

$$\boldsymbol{\lambda}^{(0)} \notin H \Longrightarrow \{\boldsymbol{\lambda}_{j|i}\}_{(i,j)\in\mathcal{N}_h\times\mathcal{N}_h,(i,j)\in\mathcal{E}} \text{ cannot be reconstructed}$$

The proposed approach still preserves privacy even if the subspace is known to the adversary

AALBORG UNIVERSITY
DENMARK

How about ADMM?

Augmented Lagrangian of ADMM

$$L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{z}) = f(\boldsymbol{x}) + \boldsymbol{\nu}^\top(\boldsymbol{M}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{z}) + \frac{c}{2}\|\boldsymbol{M}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{z} - 2\boldsymbol{d}\|^2$$

Updating functions

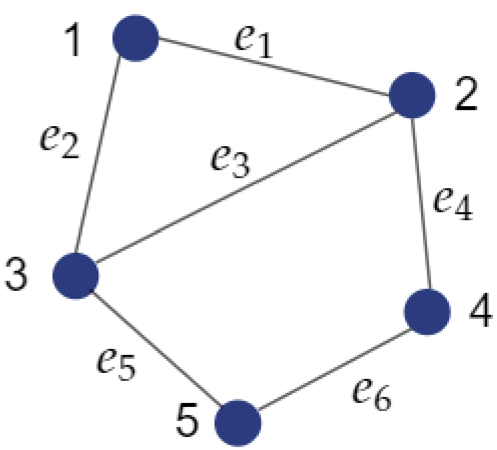
$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} L\left(\boldsymbol{x}, \boldsymbol{z}^{(k)}, \boldsymbol{\nu}^{(k)}\right))$$

$$\boldsymbol{z}^{(k+1)} = \arg\min_{\boldsymbol{z}} L\left(\boldsymbol{x}^{(k+1)}, \boldsymbol{z}, \boldsymbol{\nu}^{(k)}\right)$$

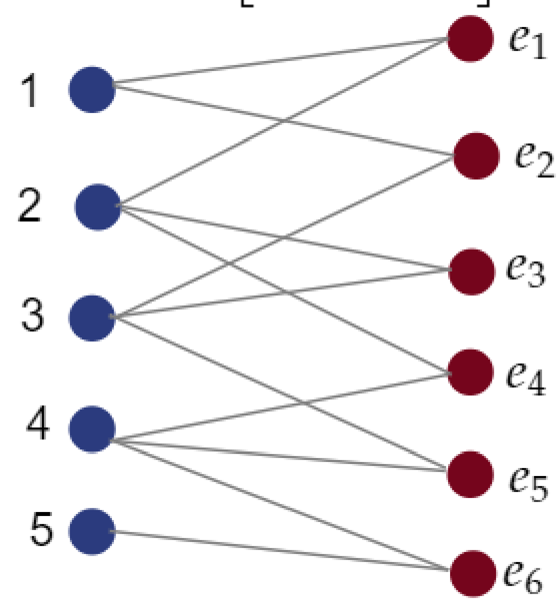
$$\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + c\left(\boldsymbol{M}\boldsymbol{x}^{(k+1)} + \boldsymbol{W}\boldsymbol{z}^{(k+1)} - 2\boldsymbol{d}\right)$$

Incidence matrix

$$\begin{bmatrix}\boldsymbol{M} & \boldsymbol{W}\end{bmatrix} = \begin{bmatrix}\boldsymbol{B}^+ & -\boldsymbol{I} \\ -\boldsymbol{B}^- & -\boldsymbol{I}\end{bmatrix} \in \mathbb{R}^{2m\times(m+n)}$$



Bipartite graph



The same applies to Dual ascent

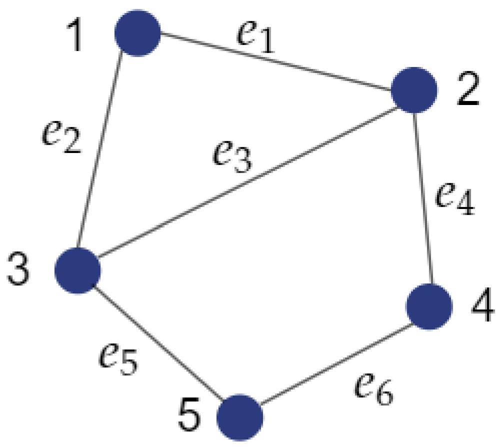
Lagrangian of dual ascent

$$L(\boldsymbol{x}, \boldsymbol{u}) = f(\boldsymbol{x}) + \boldsymbol{u}^\top (\boldsymbol{B}\boldsymbol{x} - \boldsymbol{b})$$

Updating functions

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} L\left(\boldsymbol{x}, \boldsymbol{u}^{(k)}\right)$$

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + t^{(k)}\left(\boldsymbol{B}\boldsymbol{x}^{(k+1)} - \boldsymbol{b}\right)$$



Graphs of dual ascent, ADMM and PDMM

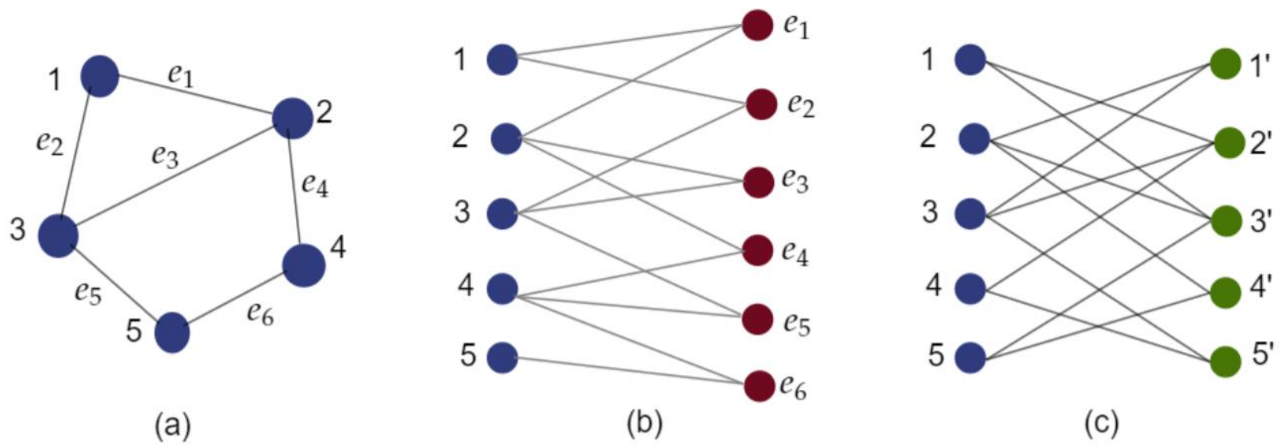


Fig. 1: An example of graph topologies associated with dual ascent, ADMM and PDMM with $u = 1$: (a) A graph with $n = 5$ nodes and $m = 6$ edges. (b) The bipartite graph constructed by ADMM with $n + m$ nodes and $2m$ edges. (c) The bipartite graph constructed by PDMM with $2n$ nodes and $2m$ edges.

The proposed subspace perturbation also applies to other optimizers like ADMM and Dual Ascent

Applications:

applicable to all convex problems

AALBORG UNIVERSITY
DENMARK

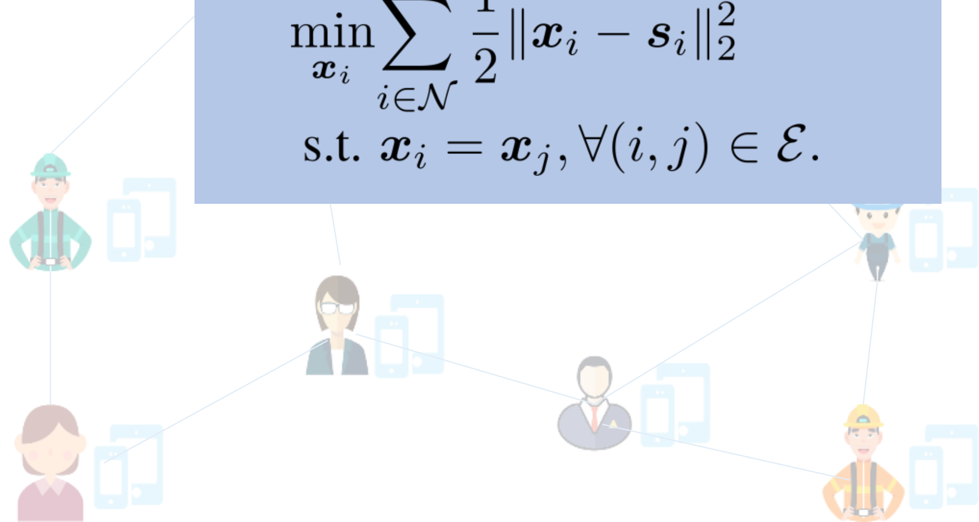
Applications(1)_consensus

- How to securely compute the average salary over a group of people while keeping each person's own salary private from others?

Distributed average consensus

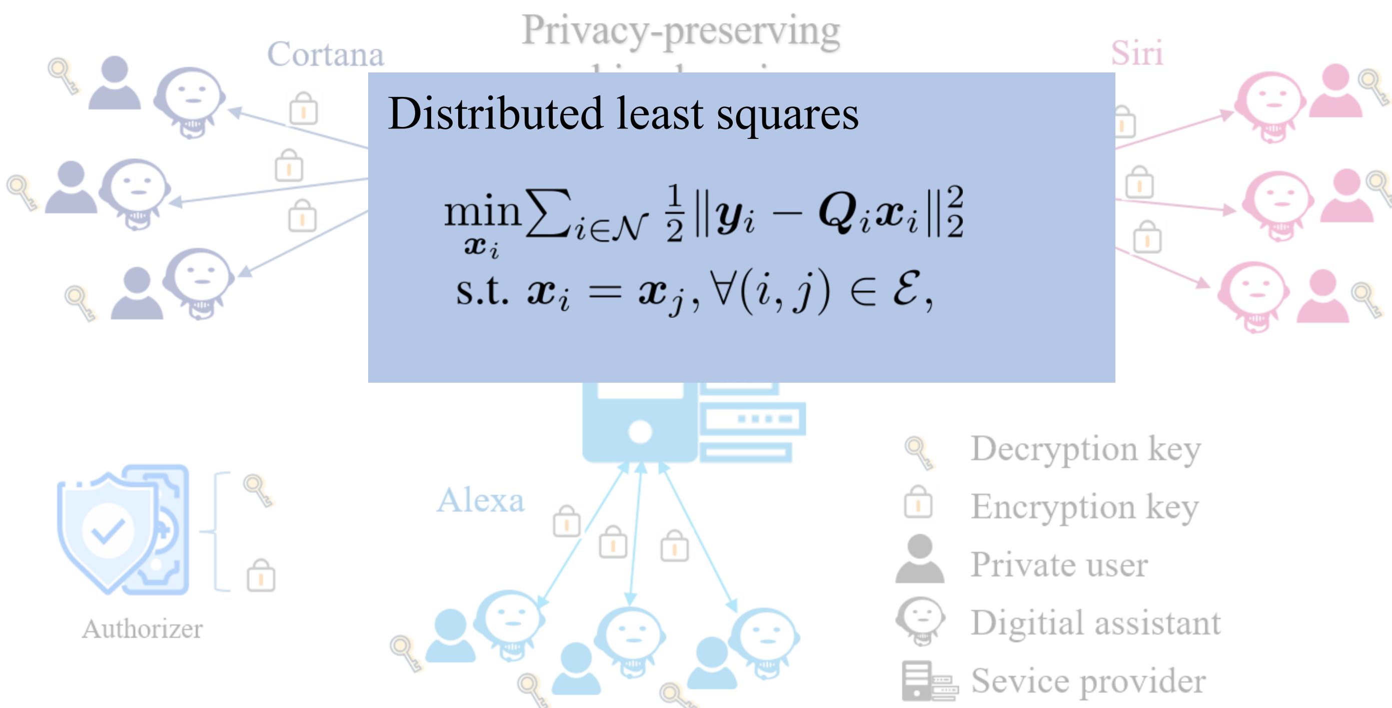
$$\min_{\boldsymbol{x}_i} \sum_{i \in \mathcal{N}} \frac{1}{2} \|\boldsymbol{x}_i - \boldsymbol{s}_i\|_2^2$$

$$\text{s.t. } \boldsymbol{x}_i = \boldsymbol{x}_j, \forall (i,j) \in \mathcal{E}.$$



Applications(2)_machine learning

- Privacy-preserving machine learning over multiple parties
  - Collaborative learning without revealing private data



Distributed least squares

$$\min_{\boldsymbol{x}_i} \sum_{i\in\mathcal{N}} \frac{1}{2}\|\boldsymbol{y}_i - \boldsymbol{Q}_i\boldsymbol{x}_i\|_2^2$$
$$\text{s.t. } \boldsymbol{x}_i = \boldsymbol{x}_j, \forall (i,j) \in \mathcal{E},$$

Applications(3)_sparsity related

- Privacy-preserving distributed compressed sensing

Distributed Lasso

$$\min_{\boldsymbol{x}_i} \sum_{i \in \mathcal{N}} \frac{1}{2} \|\boldsymbol{y}_i - \boldsymbol{Q}_i \boldsymbol{x}_i\|_2^2$$
$$\text{s.t. } \boldsymbol{x}_i = \boldsymbol{x}_j, \forall (i,j) \in \mathcal{E},$$

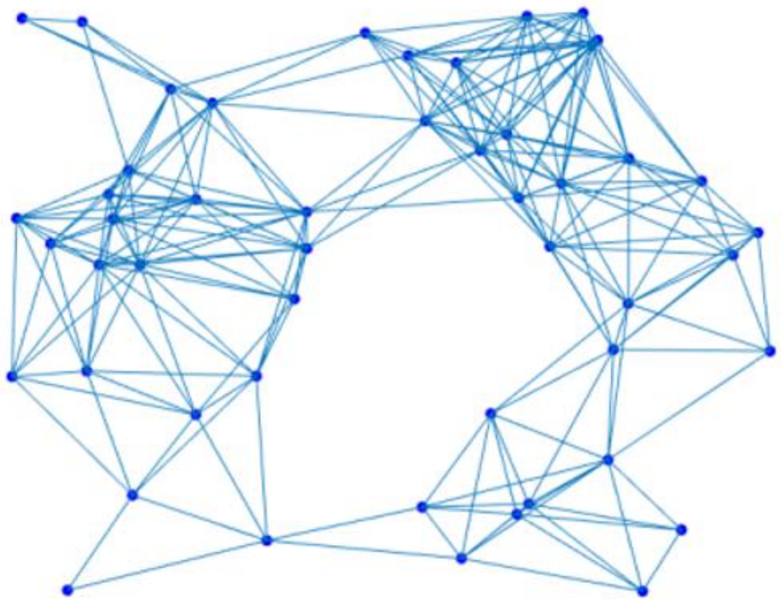
*Distributed least sqaures and Lasso have similar problem setup, the former assumes an overdetermined system and the latter assumes an underdetermined one.

Numerical results

AALBORG UNIVERSITY
DENMARK

Network setup

The connectivity of nodes is enabled if their distance is within a radius $2\sqrt{\frac{\log n}{n}}$ to have a connected graph with high probability



A random connected geometric graph with 50 nodes

AALBORG UNIVERSITY
DENMARK

Experimental results (Average consensus)

Proposed approaches

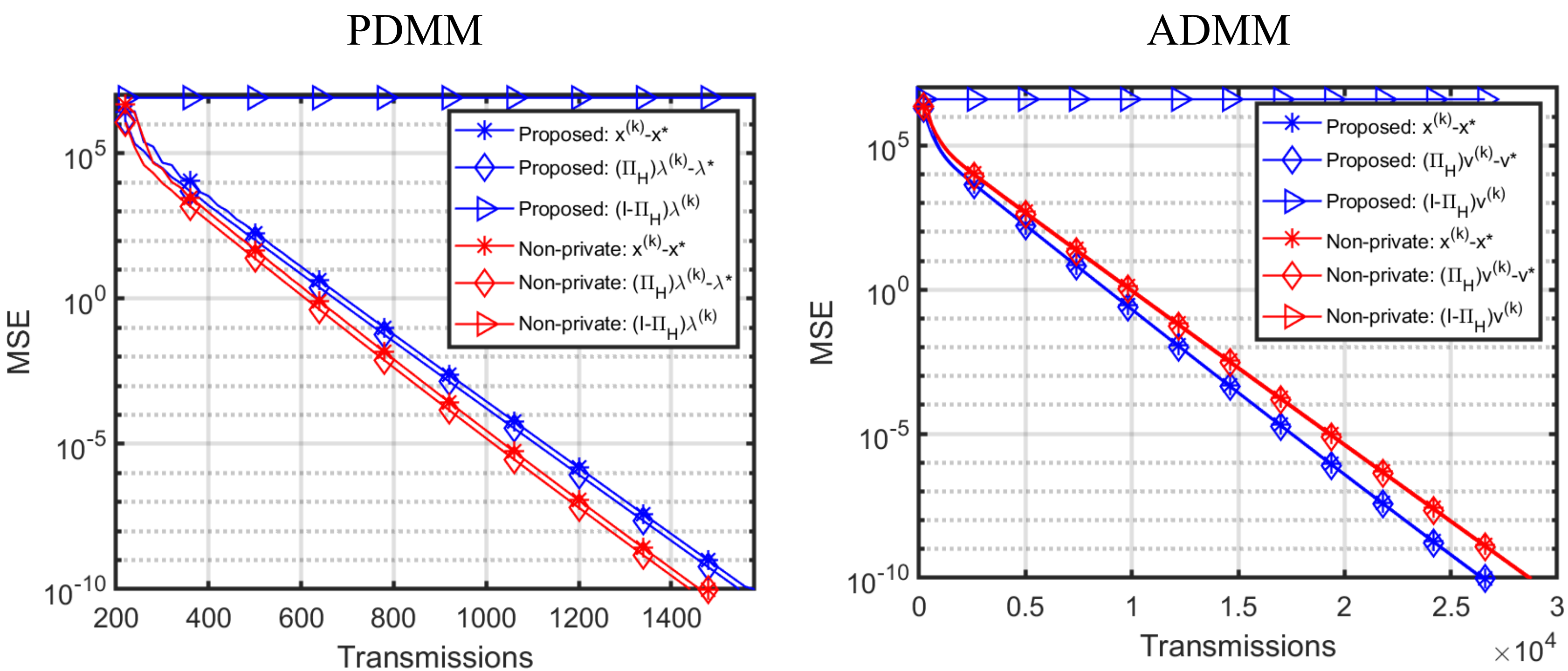


Fig.2: Convergence of the primal variable, the converging component and non-converging component of the dual variable in PDMM and ADMM with two different initializations.

AALBORG UNIVERSITY
DENMARK

Experimental results (Least square & Lasso)

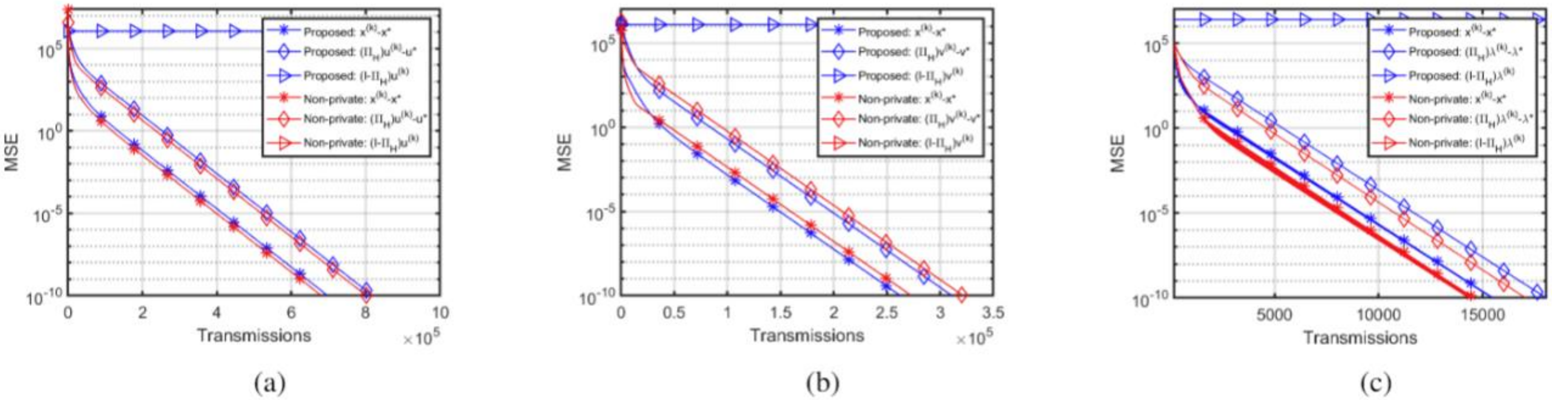


Fig. 3: Distributed least squares with two different initializations of the dual variable with a variance of 10^6: convergence of the optimization variable, the convergent and non-convergent component of the dual variable of (a) dual ascent, (b) ADMM and (c) PDMM.

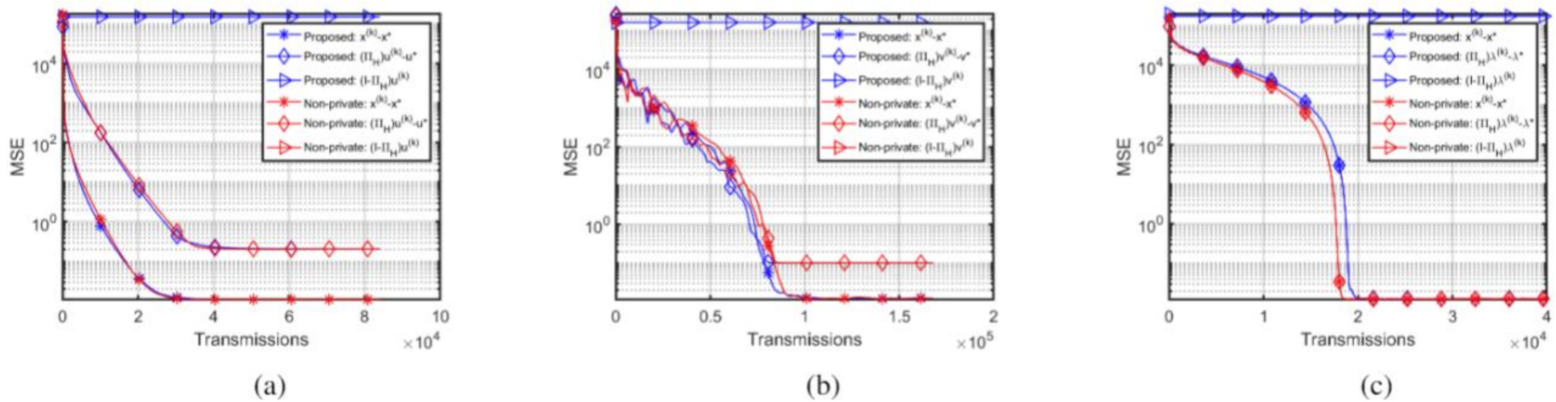
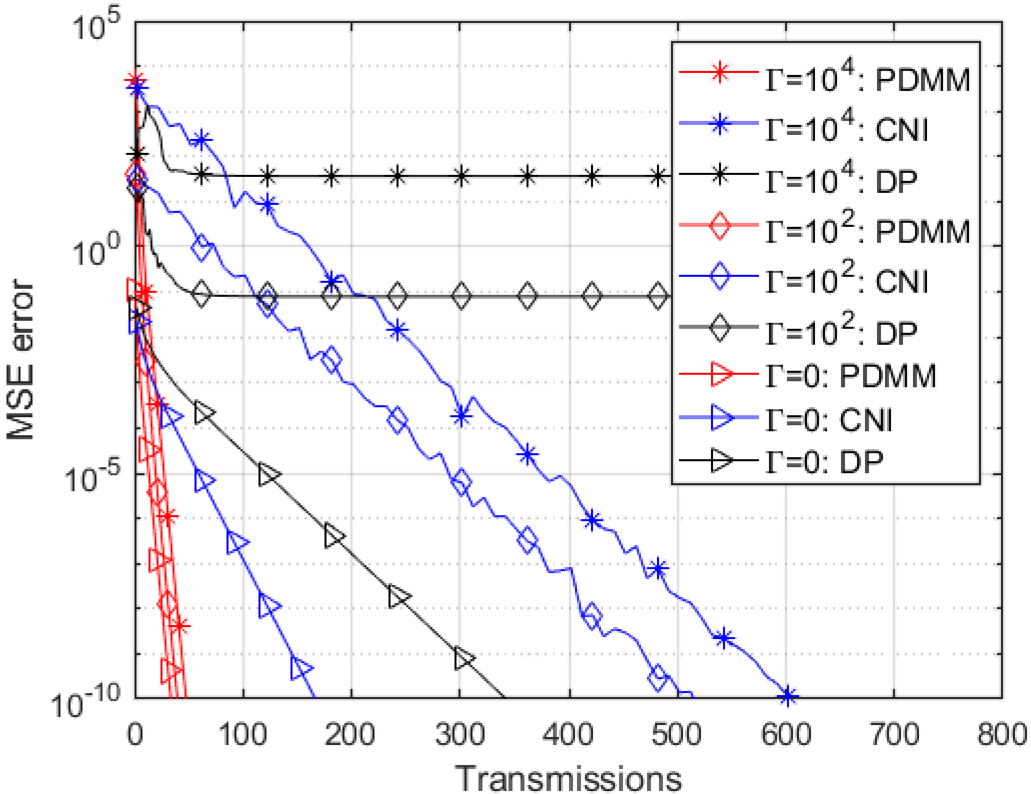


Fig. 4: Distributed LASSO with two different initializations of the dual variable with a variance of 10^6: convergence of the optimization variable, the convergent and non-convergent component of the dual variable of (a) dual ascent, (b) ADMM and (c) PDMM.

Comparison with existing methods



DP: differential privacy [Nozari, 2017]

CNI: correlated noise insertion [He, 2019]

Convergence of the proposed PDMM and state-of-the-art algorithms under three different noise levels for distributed average consensus.

AALBORG UNIVERSITY
DENMARK

Lower bound of information leakage

Sometimes it is impossible to have zero information leakage

- The optimum solution itself may reveal some private information (unavoidable if perfect accuracy is preserved)

Distributed average consensus

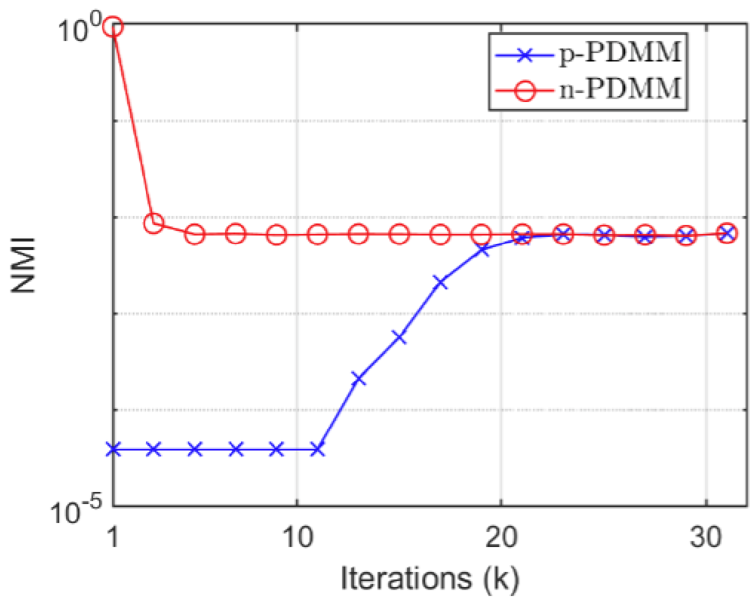


Fig. 7: Normalized mutual information of an arbitrary node i (i.e., $\frac{I(S_i;X_i^{(k+1)})}{I(S_i;S_i)}$) using the proposed p-PDMM and non-private PDMM (n-PDMM) for each iteration.

Zero privacy leakage and perfect accuracy are sometimes impossible to achieve

Conclusions and future works

AALBORG UNIVERSITY
DENMARK

Conclusions &future works

Conclusions

- A new subspace perturbation approach based on distributed convex optimization
- Generally applicable to all convex problems
- Both computationally and communication efficient (compared to SMPC)
- No tradeoff between privacy and accuracy (compared to differential privacy)
- Convergence rate is not affected
- Require one honest neighbor

Future works:

- Optimization in terms of practical constraints for example quantization
- Apply to distributed federated learning

Q&A

AALBORG UNIVERSITY
DENMARK