Secure and Resilient Distributed Machine Learning Algorithms Under Adversarial Environment

Rui Zhang Quanyan Zhu

Department of Electrical and Computer Engineering

Polytechnic School of Engineering

New York University

July 6, 2015

Motivations

Machine learning algorithms are widely used.

They are vulnerable to attackers who can create "misleading" data.

Distributed algorithms are developed to solve large-scale problems.

However, the system becomes more vulnerable.

- Enlarge attack surface.
- Network effects.

Objective of the talk

We study secure and resilient distributed machine learning algorithms under adversarial environment.

Related Works

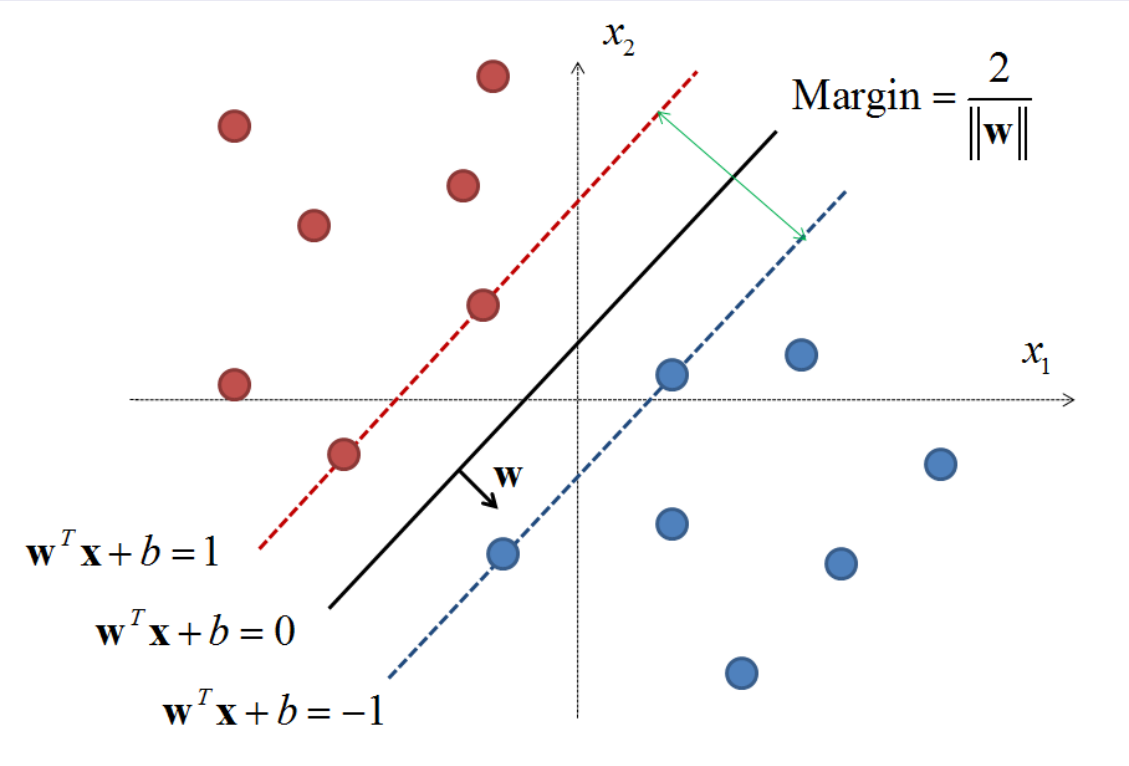
- Nilesh Dalvi, Pedro Domingos, Sumit Sanghai, and Deepak Verma. Adversarial classification. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 99-108. ACM, 2004.

- Wei Liu, and Sanjay Chawla. A game theoretical model for adversarial learning. *Proceedings of IEEE International Conference on Data Mining Workshops,* pp. 25-30. IEEE, 2009.

- Marco Barreno, Blaine Nelson, Anthony Joseph, and Doug Tygar. The security of machine learning. *Machine Learning* pp. 121-148. 2010.

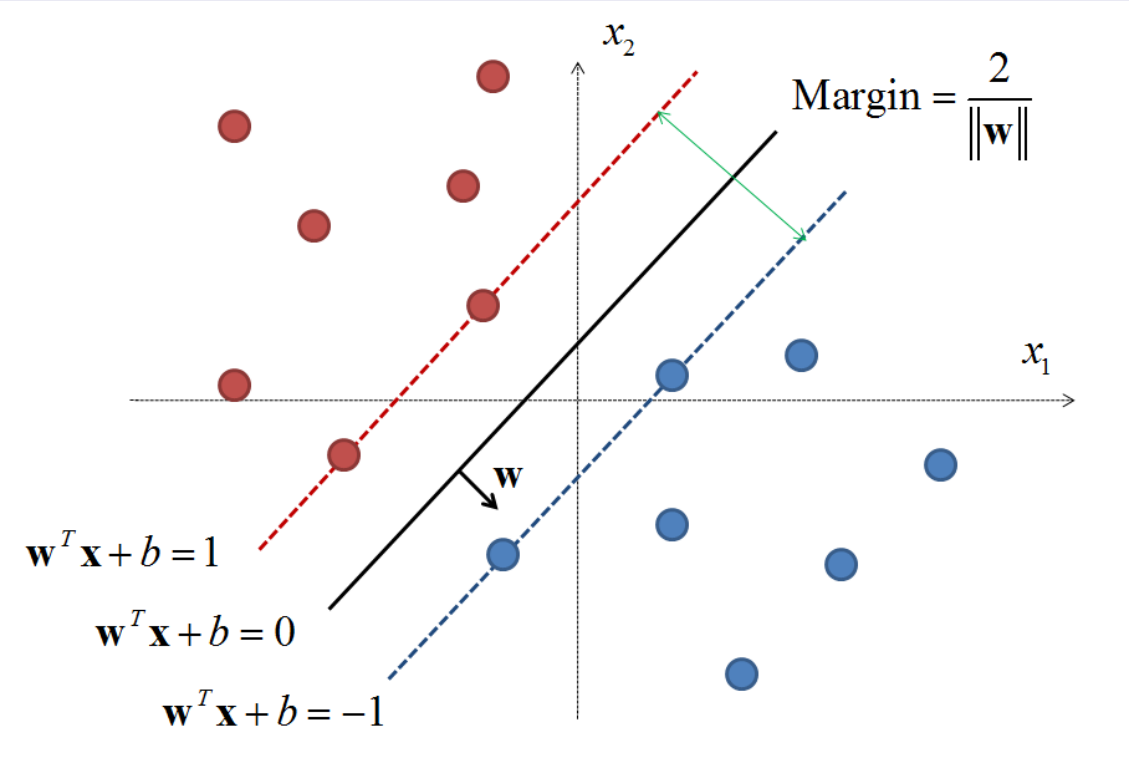
Distributed Support Vector Machines

Introduction to Support Vector Machines (SVMs)



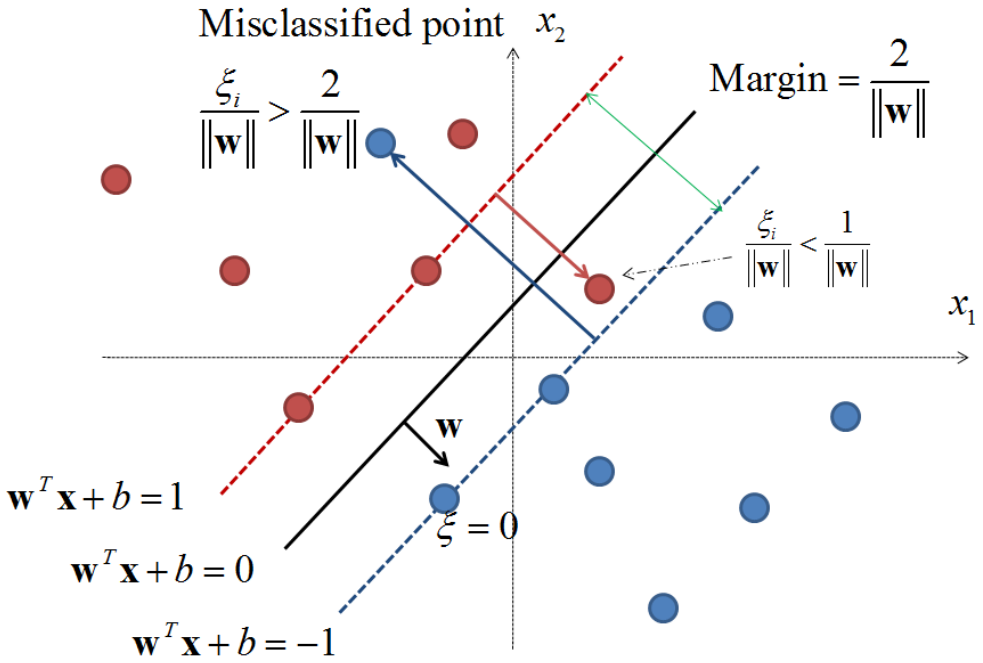
- SVMs are supervised learning models used for classification.
- The goal is to find the hyperplane with largest margins between different sets of samples [Cortes and Vapnik, 1995].

Introduction to Support Vector Machines (SVMs)



$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{s.t.} \quad y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1, \quad i = 1, ..., N.$$

Distributed Support Vector Machines

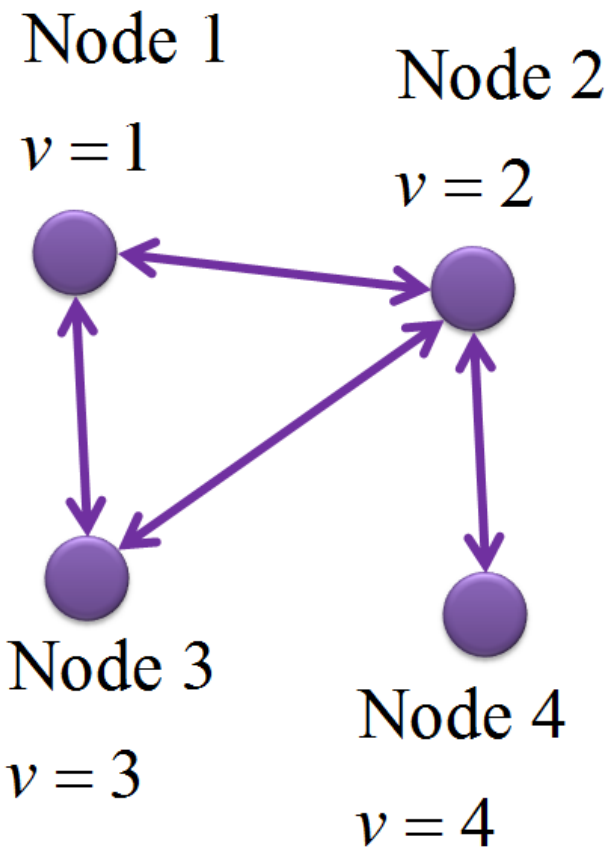


$$\min_{\mathbf{w},\{\xi_i\}} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1-\xi_i, \qquad i=1,...,N;$$

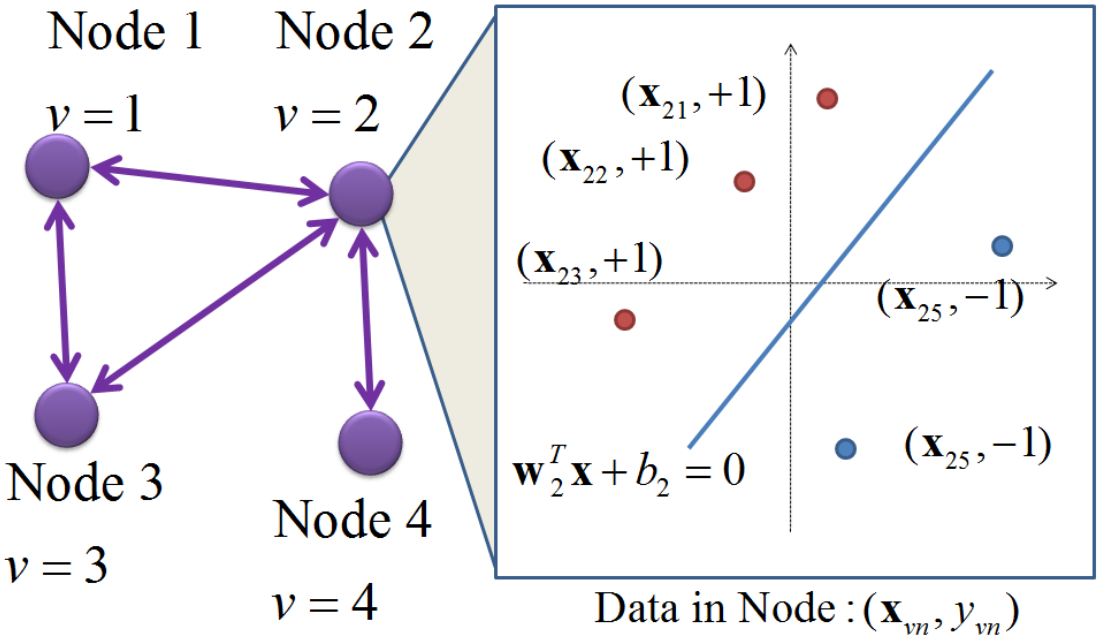
$$\xi_i \geq 0, \qquad i=1,...,N.$$

Distributed Support Vector Machines



- A network consists of 4 nodes.
- Nodes can communicate with their neighbors.
- Nodes do not have to be fully connected.

Distributed Support Vector Machines

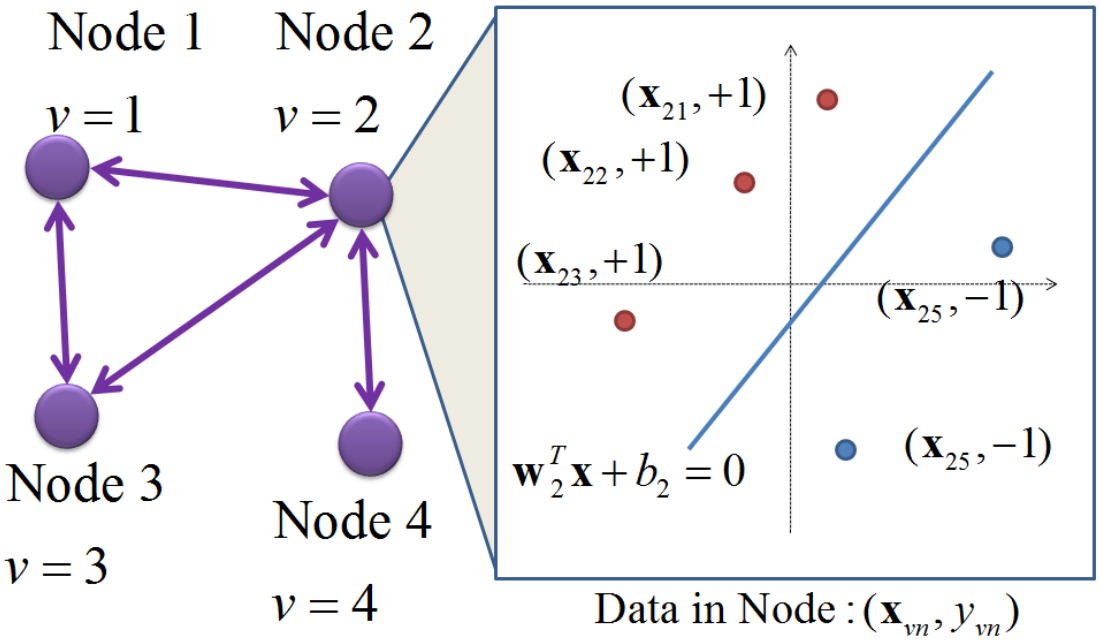


- Each node contains a labeled training set

$$\mathcal{D}_v := \{(\mathbf{x}_{vn}, y_{vn}) : n = 1, ..., N_v\}.$$

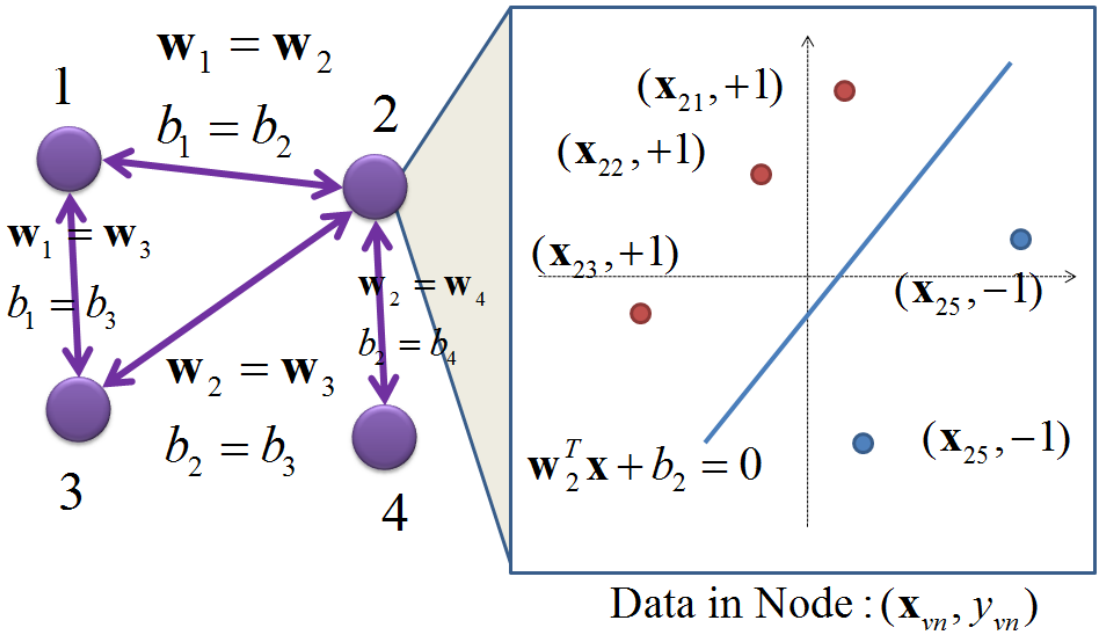
- $\mathbf{x}_{vn}$ is the n-th data in node v.
- y_{vn} is the label of n-th data in node v.

Distributed Support Vector Machines



- Pedro A. Forero, Alfonso Cano, and Georgios B. Giannakis. Consensus-based distributed support vector machines. *The Journal of Machine Learning Research* 11 (2010): 1663-1707.

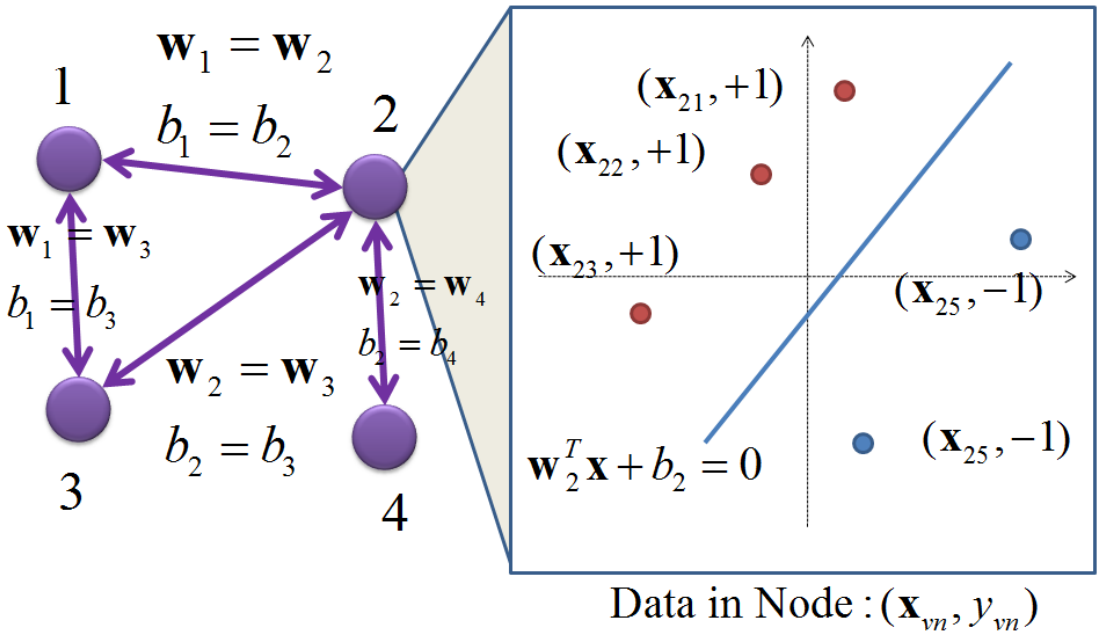
Distributed Support Vector Machines



$$\min_{\{\mathbf{w}_v, b_v, \{\xi_{vn}\}\}} \frac{1}{2} \sum_{v=1}^{V} \|\mathbf{w}_v\|^2 + V C_l \sum_{v=1}^{V} \sum_{n=1}^{N_v} \xi_{vn}$$

$$\text{s.t.} \quad \begin{array}{ll} y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v) \geq 1 - \xi_{vn}, & \forall v \in \mathcal{V}, n = 1, ..., N_v; \\ \xi_{vn} \geq 0, & \forall v \in \mathcal{V}, n = 1, ..., N_v; \\ \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, & \forall v \in \mathcal{V}, u \in \mathcal{B}_v. \end{array} \tag{1}$$

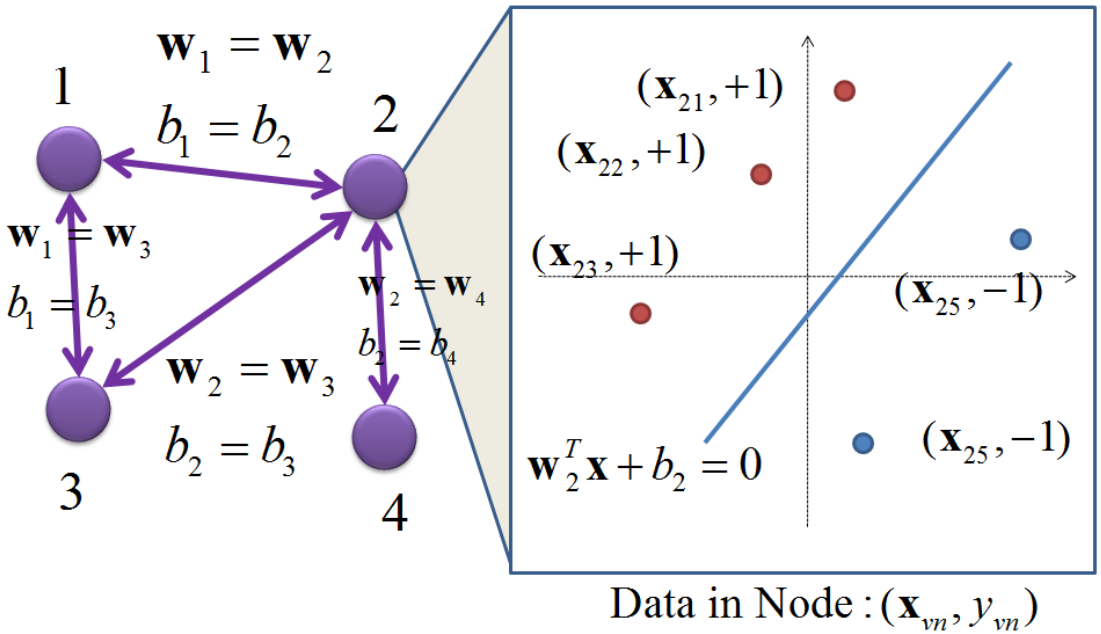
Distributed Support Vector Machines



$$\min_{\{\mathbf{w}_v, b_v, \{\xi_{vn}\}\}} \frac{1}{2} \sum_{v=1}^{V} \|\mathbf{w}_v\|^2 + V C_l \sum_{v=1}^{V} \sum_{n=1}^{N_v} \xi_{vn}$$

$$\begin{aligned} & y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v) \geq 1 - \xi_{vn}, && \forall v \in \mathcal{V}, n = 1, ..., N_v; \\ \text{s.t.} \quad & \xi_{vn} \geq 0, && \forall v \in \mathcal{V}, n = 1, ..., N_v; \\ & \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, && \forall v \in \mathcal{V}, u \in \mathcal{B}_v. \end{aligned}$$

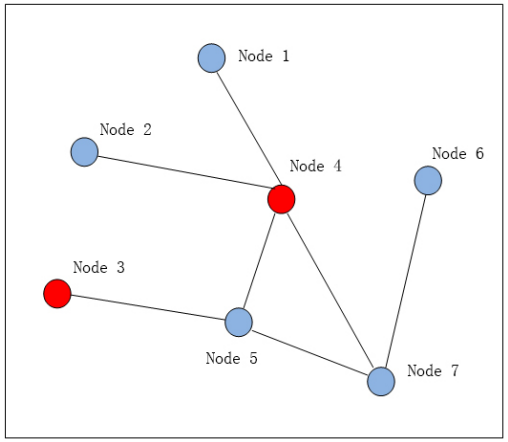
Distributed Support Vector Machines



$$\min_{\{\mathbf{w}_v, b_v\}} \frac{1}{2} \sum_{v=1}^{V} \|\mathbf{w}_v\|^2 + V C_l \sum_{v=1}^{V} \sum_{n=1}^{N_v} \left[1 - y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+ \tag{2}$$

$$\text{s.t.} \quad \mathbf{w}_v = \mathbf{w}_u, \quad b_v = b_u, \quad \forall v \in \mathcal{V}, u \in B_v.$$

Problem Statement: Overview



- Each node contains a labeled training set $\mathcal{D}_v := \{(\mathbf{x}_{vn}, y_{vn}) : n = 1, ..., N_v\}$.
- Node 4 can communicate with its 4 neighbors. An attacker can take over a subset of the nodes.

System Model

- Aim of Learner: High Accuracy
- Aim of Attacker: Low Accuracy

Attack Model

- Attacker knows learning algorithm.
- Attacker can take over a set of nodes.
- Attacker can modify the training data.

A game between a learner and an attacker

Problem Statement: Mathematical Model

Mathematical Model for Learner

$$\min_{\{\mathbf{w}_v, b_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)} + \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \widehat{\mathbf{x}}_{vn} + b_v)\right]_+}_{(c)} \tag{3}$$

$$\text{s.t.} \quad \mathbf{w}_v = \mathbf{w}_u, \quad b_v = b_u, \quad \forall v \in \mathcal{V}, u \in \mathcal{B}_v.$$

- Term (a) is the sum of inverse of the distance of margin.
- Term (b) is the error penalty of nodes without attack.
- Term (c) is the error penalty of nodes under attack.
- $\widehat{\mathbf{x}}_{vn}$ represents data modified by the attacker.

Problem Statement: Mathematical Model

Mathematical Model for Learner

$$\min_{\{\mathbf{w}_v, b_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - y_{vn}(\mathbf{w}_v^T \widehat{\mathbf{x}}_{vn} + b_v)\right]_+}_{(c)}$$
$$\text{s.t.} \quad \mathbf{w}_v = \mathbf{w}_u, \quad b_v = b_u, \quad \forall v \in \mathcal{V}, u \in \mathcal{B}_v.$$

- Term (a) is the sum of inverse of the distance of margin.
- Term (b) is the error penalty of nodes without attack.
- Term (c) is the error penalty of nodes under attack.
- $\widehat{\mathbf{x}}_{vn}$ represents data modified by the attacker.

Problem Statement: Mathematical Model

Mathematical Model for Learner

$$\min_{\{\mathbf{w}_v, b_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_I C_I \sum_{v=1}^{V_I}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_I \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \widehat{\mathbf{x}}_{vn} + b_v)\right]_+}_{(c)}$$
$$\text{s.t.} \quad \mathbf{w}_v = \mathbf{w}_u, \quad b_v = b_u, \quad \forall v \in \mathcal{V}, u \in \mathcal{B}_v.$$

- Term (a) is the sum of inverse of the distance of margin.
- Term (b) is the error penalty of nodes without attack.
- Term (c) is the error penalty of nodes under attack.
- $\widehat{\mathbf{x}}_{vn}$ represents data modified by the attacker.

Problem Statement: Mathematical Model

Mathematical Model for Learner

$$\min_{\{\mathbf{w}_v, b_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \widehat{\mathbf{x}}_{vn} + b_v)\right]_+}_{(c)}$$
$$\text{s.t.} \quad \mathbf{w}_v = \mathbf{w}_u, \quad b_v = b_u, \quad \forall v \in \mathcal{V}, u \in \mathcal{B}_v.$$

- Term (a) is the sum of inverse of the distance of margin.
- Term (b) is the error penalty of nodes without attack.
- Term (c) is the error penalty of nodes under attack.
- $\widehat{\mathbf{x}}_{vn}$ represents data modified by the attacker.

Problem Statement: Mathematical Model

Mathematical Model for Learner

$$\min_{\{\mathbf{w}_v, b_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \widehat{\mathbf{x}}_{vn} + b_v)\right]_+}_{(c)}$$
$$\text{s.t.} \quad \mathbf{w}_v = \mathbf{w}_u, \quad b_v = b_u, \quad \forall v \in \mathcal{V}, u \in \mathcal{B}_v.$$

- Term (a) is the sum of inverse of the distance of margin.
- Term (b) is the error penalty of nodes without attack.
- Term (c) is the error penalty of nodes under attack.
- $\widehat{\mathbf{x}}_{vn}$ represents data modified by the attacker.

Problem Statement: Mathematical Model

Mathematical Model for the Attacker

- The attacker changes $\mathbf{x}_{vn}$ into $\widehat{\mathbf{x}}_{vn} = \mathbf{x}_{vn} - \delta_{vn}$, where $\{\delta_{vn}\} \in \mathcal{U}$, and $\mathcal{U}$ is the attacker's action set [Xu et al., 2009].
- Here

$$\mathcal{U} = \left\{ (\delta_{v1}, \delta_{v2}, ..., \delta_{vN_v}) \left| \sum_{n=1}^{N_v} \|\delta_{vn}\| \leq C_\delta \right. \right\}$$

- Also we introduce

$$\mathcal{U}_0 = \{\delta_v \,|\|\delta_v\| \leq C_\delta\}$$

- Notice that C_δ does not have to be same in different nodes.

Problem Statement: Mathematical Model

Mathematical Model for Attacker

$$\max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1-\mathrm{y}_{vn}(\mathbf{w}_v^T\mathbf{x}_{vn}+b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1-\mathrm{y}_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn}-\delta_{vn})+b_v)\right]_+}_{(c)} \underbrace{- C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$
$$\text{s.t. } (\delta_{v1},...,\delta_{vN_v}) \in \mathcal{U}_v, \quad \forall v \in \mathcal{V}_a. \tag{4}$$

- Terms (a)(b)(c) are the objectives of the learner's min-problem.
- Term (d) is the cost function for the attacker.
- l_0 norm is a total number of nonzero elements in a vector.

Problem Statement: Mathematical Model

Mathematical Model for Attacker

$$\max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1-\mathrm{y}_{vn}(\mathbf{w}_v^T\mathbf{x}_{vn}+b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1-\mathrm{y}_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn}-\delta_{vn})+b_v)\right]_+}_{(c)} - \underbrace{C_a\sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$
$$\text{s.t. } (\delta_{v1},...,\delta_{vN_v}) \in \mathcal{U}_v, \quad \forall v \in \mathcal{V}_a.$$

- Terms (a)(b)(c) are the objectives of the learner's min-problem.
- Term (d) is the cost function for the attacker.
- l_0 norm is a total number of nonzero elements in a vector.

Problem Statement: Mathematical Model

Mathematical Model for Attacker

$$\max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1-\mathrm{y}_{vn}(\mathbf{w}_v^T\mathbf{x}_{vn}+b_v)\right]_+}_{(b)}$$

$$\underbrace{+ V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1-\mathrm{y}_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn}-\delta_{vn})+b_v)\right]_+}_{(c)} \underbrace{- C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$

$$\text{s.t. } (\delta_{v1},...,\delta_{vN_v}) \in \mathcal{U}_v, \quad \forall v \in \mathcal{V}_a.$$

- Terms (a)(b)(c) are the objectives of the learner's min-problem.
- Term (d) is the cost function for the attacker.
- l_0 norm is a total number of nonzero elements in a vector.

Problem Statement: Mathematical Model

Mathematical Model for Attacker

$$\max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\Big[1-\mathrm{y}_{vn}(\mathbf{w}_v^T\mathbf{x}_{vn}+b_v)\Big]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\Big[1-\mathrm{y}_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn}-\delta_{vn})+b_v)\Big]_+}_{(c)} \underbrace{- C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$
$$\text{s.t. } (\delta_{v1},...,\delta_{vN_v}) \in \mathcal{U}_v, \quad \forall v \in \mathcal{V}_a.$$

- Terms (a)(b)(c) are the objectives of the learner's min-problem.
- Term (d) is the cost function for the attacker.
- l_0 norm is a total number of nonzero elements in a vector.

Mathematical Model:Min-Max Problem

$$\min_{\{\mathbf{w}_v, b_v\}} \max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$

$$\underbrace{+ V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn} - \delta_{vn}) + b_v)\right]_+}_{(c)} \underbrace{- C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$

$$\text{s.t.} \quad \begin{array}{ll} \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, & \forall v \in \mathcal{V}, u \in \mathcal{B}_v; \\ (\delta_{v1}, ..., \delta_{vN_v}) \in \mathcal{U}_v, & \forall v \in \mathcal{V}_a. \end{array} \tag{5}$$

- Terms $(a)(b)(c)$ are related to the min-problem for the learner.
- Terms $(a)(b)(c)(d)$ are related to the max-problem for the attacker.
- A nonzero-sum game.

Mathematical Model:Min-Max Problem

$$\min_{\{\mathbf{w}_v, b_v\}} \max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$

$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - y_{vn}(\mathbf{w}_v^T (\mathbf{x}_{vn} - \delta_{vn}) + b_v)\right]_+}_{(c)} - \underbrace{C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$

$$\text{s.t.} \quad \begin{array}{ll} \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, & \forall v \in \mathcal{V}, u \in \mathcal{B}_v; \\ (\delta_{v1}, ..., \delta_{vN_v}) \in \mathcal{U}_v, & \forall v \in \mathcal{V}_a. \end{array}$$

- Terms $(a)(b)(c)$ are related to the min-problem for the learner.
- Terms $(a)(b)(c)(d)$ are related to the max-problem for the attacker.
- A nonzero-sum game.

Mathematical Model:Min-Max Problem

$$\min_{\{\mathbf{w}_v, b_v\}} \max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$

$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - \mathrm{y}_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn} - \delta_{vn}) + b_v)\right]_+}_{(c)} \underbrace{- C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$

$$\text{s.t.} \quad \begin{array}{ll} \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, & \forall v \in \mathcal{V}, u \in \mathcal{B}_v; \\ (\delta_{v1}, ..., \delta_{vN_v}) \in \mathcal{U}_v, & \forall v \in \mathcal{V}_a. \end{array}$$

- Terms $(a)(b)(c)$ are related to the min-problem for the learner.
- Terms $(a)(b)(c)(d)$ are related to the max-problem for the attacker.
- A nonzero-sum game.

Mathematical Model:Min-Max Problem

$$\min_{\{\mathbf{w}_v, b_v\}} \max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V_l C_l \sum_{v=1}^{V_l}\sum_{n=1}^{N_v}\left[1 - y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$

$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\left[1 - y_{vn}(\mathbf{w}_v^T(\mathbf{x}_{vn} - \delta_{vn}) + b_v)\right]_+}_{(c)} - \underbrace{C_a \sum_{v=1}^{V_a}\sum_{n=1}^{N_v}\|\delta_{vn}\|_0}_{(d)}$$

$$\text{s.t.} \quad \begin{array}{ll} \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, & \forall v \in \mathcal{V}, u \in \mathcal{B}_v; \\ (\delta_{v1}, ..., \delta_{vN_v}) \in \mathcal{U}_v, & \forall v \in \mathcal{V}_a. \end{array}$$

- Terms $(c)(d)$ can be simplified further by using the definition of $\mathcal{U}_v$ and $\mathcal{U}_{v0}$.

Solutions and Algorithms:Proposition 1

With $\mathcal{U}_v = \left\{ (\delta_{v1}, ..., \delta_{vN_v}) \,\middle|\, \sum_{n=1}^{N_v} \|\delta_{vn}\| \leq C_\delta \right\}$ and $\mathcal{U}_{v0} = \{\delta_v \,|\, \|\delta_v\| \leq C_\delta\}$, the min-max problem is equivalent to the following problem:

$$\min_{\{\mathbf{w}_v, b_v\}} \max_{\{\delta_{vn}\}} \underbrace{\frac{1}{2} \sum_{v=1}^{V} \|\mathbf{w}_v\|^2}_{(a)} + \underbrace{V C_l \sum_{v=1}^{V} \sum_{n=1}^{N_v} \left[1 - y_{vn}(\mathbf{w}_v^T \mathbf{x}_{vn} + b_v)\right]_+}_{(b)}$$
$$+ \underbrace{V_a C_l \sum_{v=1}^{V_a} \mathbf{w}_v^T \delta_v}_{(c)} - \underbrace{C_a \sum_{v=1}^{V_a} \|\delta_v\|_0}_{(d)}, \tag{6}$$
$$\text{s.t.} \quad \begin{array}{ll} \mathbf{w}_v = \mathbf{w}_u, b_v = b_u, & \forall v \in \mathcal{V}, u \in \mathcal{B}_v; \\ \delta_v \in \mathcal{U}_{v0}, & \forall v \in \mathcal{V}_a. \end{array}$$

Solutions and Algorithms:Problem Reformulation

- Decision variables space $\mathbf{r}_v := [\mathbf{w}_v^T, b_v]^T$,
- Augmented matrix $\mathbf{X}_v := [(\mathbf{x}_{v1}, ..., \mathbf{x}_{vN_v})^T, \mathbf{1}_v]$,
- Diagonal label matrix $\mathbf{Y}_v := diag([y_{v1}, ..., y_{vN_v}])$.
- Note that $\mathbf{w}_v = (\mathbf{I}_{p+1} - \Pi_{p+1})\mathbf{r}_v$, where $\mathbf{I}_{p+1} - \Pi_{p+1}$ is a $(p+1) \times (p+1)$ identity matrix with 0 at $(p+1, p+1)$-st entry.

Thus, problem (6) can be rewritten as

$$\begin{aligned} \min_{\{\mathbf{r}_v, \omega_{vu}\}} \max_{\{\delta_v\}} & \frac{1}{2} \sum_{v=1}^{V} \mathbf{r}_v^T (\mathbf{I}_{p+1} - \Pi_{p+1}) \mathbf{r}_v + V C_l \sum_{v=1}^{V} [\mathbf{1}_v - \mathbf{Y}_v \mathbf{X}_v \mathbf{r}_v]_+ \\ & + V_a C_l \sum_{v=1}^{V_a} \mathbf{r}_v^T (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v - C_a \sum_{v=1}^{V_a} \|\delta_v\|_0 \qquad (7) \\ \text{s.t.} \quad & \mathbf{r}_v = \omega_{vu}, \omega_{vu} = \mathbf{r}_u, \qquad \forall v \in \mathcal{V}, \forall u \in \mathcal{B}_v; \\ & \delta_v \in \mathcal{U}_{v0}, \qquad \forall v \in \mathcal{V}_a. \end{aligned}$$

Solutions and Algorithms: Problem Reformulation

$$\min_{\{\mathbf{r}_v,\omega_{vu}\}} \max_{\{\delta_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\mathbf{r}_v^T(\mathbf{I}_{p+1}-\Pi_{p+1})\mathbf{r}_v}_{(a)} + \underbrace{VC_l\sum_{v=1}^{V}[\mathbf{1}_v-\mathbf{Y}_v\mathbf{X}_v\mathbf{r}_v]_+}_{(b)}$$
$$\underbrace{+V_aC_l\sum_{v=1}^{V_a}\mathbf{r}_v^T(\mathbf{I}_{p+1}-\Pi_{p+1})\delta_v}_{(c)} \underbrace{-C_a\sum_{v=1}^{V_a}\|\delta_v\|_0}_{(d)}$$
$$\text{s.t.} \quad \mathbf{r}_v=\omega_{vu},\omega_{vu}=\mathbf{r}_u, \quad \forall v\in\mathcal{V},\forall u\in\mathcal{B}_v;$$
$$\delta_v\in\mathcal{U}_{v0}, \quad \forall v\in\mathcal{V}_a.$$

- Terms $(c)(d)$ are related to the max-problem.
- Terms $(a)(b)(c)$ are related to the min-problem.
- A zero-sum game.

Solutions and Algorithms: Problem Reformulation

$$\min_{\{\mathbf{r}_v,\omega_{vu}\}} \max_{\{\delta_v\}} \underbrace{\frac{1}{2}\sum_{v=1}^{V}\mathbf{r}_v^T(\mathbf{I}_{p+1}-\Pi_{p+1})\mathbf{r}_v}_{(a)} + \underbrace{VC_l\sum_{v=1}^{V}[\mathbf{1}_v-\mathbf{Y}_v\mathbf{X}_v\mathbf{r}_v]_+}_{(b)}$$
$$\underbrace{+V_aC_l\sum_{v=1}^{V_a}\mathbf{r}_v^T(\mathbf{I}_{p+1}-\Pi_{p+1})\delta_v}_{(c)} - \underbrace{C_a\sum_{v=1}^{V_a}\|\delta_v\|_0}_{(d)}$$

$$\text{s.t.} \quad \begin{array}{ll} \mathbf{r}_v=\omega_{vu}, \omega_{vu}=\mathbf{r}_u, & \forall v\in\mathcal{V}, \forall u\in\mathcal{B}_v; \\ \delta_v\in\mathcal{U}_{v0}, & \forall v\in\mathcal{V}_a. \end{array}$$

- Terms $(c)(d)$ are related to the max-problem.
- Terms $(a)(b)(c)$ are related to the min-problem.
- A zero-sum game.

Construct Min-Problem and Max-Problem

- To solve the min-max problem, we use best response dynamics to construct the best response for the min-problem and max-problem separately.
- For fixed $\{\mathbf{r}_v^*\}$, the max-problem,

$$\begin{aligned} \max_{\{\delta_v\}} & \sum_{v=1}^{V_a} \left(V_a C_l \mathbf{r}_{\mathbf{v}}^{*T} (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v - C_a \left\| \delta_v \right\|_0 \right) \\ \text{s.t.} \quad & \delta_v \in \mathcal{U}_{v0}, \quad \forall v \in \mathcal{V}_a. \end{aligned} \tag{8}$$

- For a fixed $\{\delta_v^*\}$, the min-problem,

$$\begin{aligned} \min_{\{\mathbf{r}_v, \omega_{vu}\}} & \frac{1}{2} \sum_{v=1}^{V} \mathbf{r}_v^T (\mathbf{I}_{p+1} - \Pi_{p+1}) \mathbf{r}_v + V C_l \sum_{v=1}^{V} \left[\mathbf{1}_v - \mathbf{Y}_v \mathbf{X}_v \mathbf{r}_v \right]_+ \\ & + \sum_{v=1}^{V_a} V_a C_l \mathbf{r}_v^T (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v^* \\ \text{s.t.} \quad & \mathbf{r}_v = \omega_{vu}, \omega_{vu} = \mathbf{r}_u, \qquad \forall v \in \mathcal{V}, \forall u \in \mathcal{B}_v. \end{aligned} \tag{9}$$

Solutions and Algorithms

The Max-Problem

For fixed $\{\mathbf{r}_v^*, \delta\}$, the max-problem,

$$\begin{aligned} \max_{\{\delta_v\}} \sum_{v=1}^{V_a} \left(V_a C_I \mathbf{r_v^*}^T (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v - C_a \left\| \delta_v \right\|_0 \right) \\ \text{s.t.} \quad \delta_v \in \mathcal{U}_{v0}, \quad \forall v \in \mathcal{V}_a. \end{aligned} \tag{10}$$

Solutions to the Max-Problem

We relax l_0 norm to l_1 norm to represent the cost function of the attacker. By writing the dual form of the l_1 norm, we arrive at

$$\begin{aligned} \max_{\{\delta_v, s_v\}} \; & V_a C_I \mathbf{r_v^*}^T (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v - \mathbf{1}^T s_v \\ & C_a \delta_v \leq s_v, \qquad \forall v \in \mathcal{V}_a; \\ \text{s.t.} \quad & C_a \delta_v \geq -s_v, \qquad \forall v \in \mathcal{V}_a; \\ & \delta_v \in \mathcal{U}_{v0}, \qquad \forall v \in \mathcal{V}_a. \end{aligned} \tag{11}$$

Solutions and Algorithms

The Max-Problem

For fixed $\{\mathbf{r}_v^*, \delta\}$, the max-problem,

$$\max_{\{\delta_v\}} \sum_{v=1}^{V_a} \left(V_a C_I \mathbf{r}_{\mathbf{v}}^{*T} (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v - C_a \|\delta_v\|_0 \right)$$
$$\text{s.t.} \quad \delta_v \in \mathcal{U}_{v0}, \quad \forall v \in \mathcal{V}_a.$$

Solutions to the Max-Problem

We relax l_0 norm to l_1 norm to represent the cost function of the attacker. By writing the dual form of the l_1 norm, we arrive at

$$\max_{\{\delta_v, s_v\}} \; V_a C_I \mathbf{r}_{\mathbf{v}}^{*T} (\mathbf{I}_{p+1} - \Pi_{p+1}) \delta_v - \mathbf{1}^T s_v$$
$$\begin{aligned} & C_a \delta_v \le s_v, && \forall v \in \mathcal{V}_a; \\ \text{s.t.} \quad & C_a \delta_v \ge -s_v, && \forall v \in \mathcal{V}_a; \\ & \delta_v \in \mathcal{U}_{v0}, && \forall v \in \mathcal{V}_a. \end{aligned}$$

Solutions and Algorithms

The Min-Problem

For fixed $\{\delta_v^*\}$, the min-problem,

$$\begin{aligned}\min_{\{\mathbf{r}_v,\xi_v,\omega_{vu}\}} \ & \frac{1}{2}\sum_{v=1}^{V}\mathbf{r}_v^T(\mathbf{I}_{p+1}-\Pi_{p+1})\mathbf{r}_v + VC_l\sum_{v=1}^{V}[\mathbf{1}_v-\mathbf{Y}_v\mathbf{X}_v\mathbf{r}_v]_+ \\ & + \sum_{v=1}^{V_a} V_a C_l \mathbf{r}_v^T(\mathbf{I}_{p+1}-\Pi_{p+1})\delta_v^* \\ \text{s.t.} \quad & \mathbf{r}_v=\omega_{vu}, \omega_{vu}=\mathbf{r}_u, \qquad \forall v\in\mathcal{V}, \forall u\in\mathcal{B}_u.\end{aligned} \tag{12}$$

Solution to the Min-Problem

We use the alternating direction method of multipliers (ADMoM).

ADMoM for the Min-Problem

ADMoM is a distributed optimization algorithm [Boyd et al.,2010].

$$\begin{aligned} &\min_{\{\mathbf{r}_v,\omega\}} \sum_{v=1}^{V} f_v(\mathbf{r}_v) \\ &\text{s.t.} \quad \mathbf{r}_v - \omega = 0, \quad v = 1, ..., V. \end{aligned} \tag{13}$$

where f_v are convex functions.

With α_v denotes the Lagrange multipliers, the ADMoM solves problem (13) by the update rules below:

$$\mathbf{r}_v(t+1) \in \arg\min_{\mathbf{r}_v \in \mathcal{P}_1} f_v(\mathbf{r}_v) + \alpha_v^T(\mathbf{r}_v - \omega) + \frac{\eta}{2}\|\mathbf{r}_v - \omega\|^2. \tag{14}$$

$$\omega(t+1) \in \arg\min_{\omega \in \mathcal{P}_2} \sum_{v=1}^{V} f_v(\mathbf{r}_v) + \alpha_v^T(\mathbf{r}_v - \omega) + \frac{\eta}{2}\|\mathbf{r}_v - \omega\|^2. \tag{15}$$

$$\alpha_v(t+1) = \alpha_v(t) + \eta(\mathbf{r}_v(t+1) - \omega(t+1)). \tag{16}$$

Solutions to the Min-Problem

- Apply ADMoM to the Min-Problem, and use iterative methods to find solutions.

Lemma Each node iterates $\lambda_v(t), \mathbf{r}_v(t)$ and $\alpha_v(t)$, given by

$$\begin{aligned}\lambda_v(t+1) \in \arg \max_{\mathbf{0} \le \lambda_v \le V C_l \mathbf{1}_v} & -\tfrac{1}{2}\lambda_v^T \mathbf{Y}_v \mathbf{X}_v \mathbf{U}_v^{-1} \mathbf{X}_v^T \mathbf{Y}_v \lambda_v \\ & + (\mathbf{1}_v + \mathbf{Y}_v \mathbf{X}_v \mathbf{U}_v^{-1} \mathbf{f}_v(t))^T \lambda_v,\end{aligned} \tag{17}$$

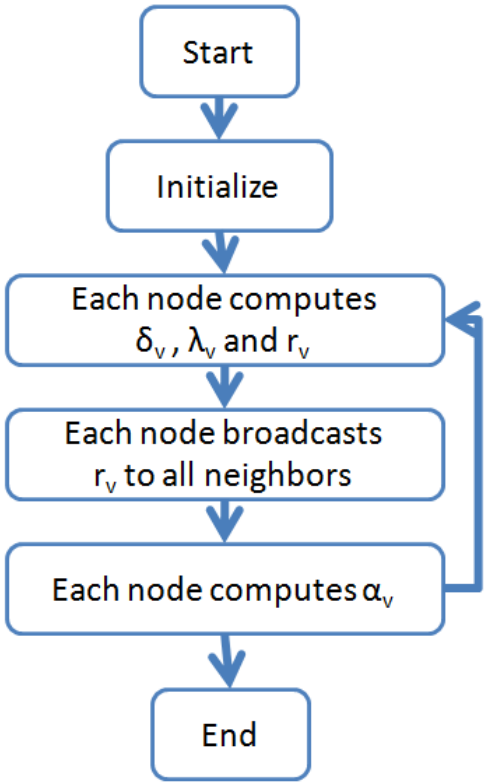
$$\mathbf{r}_v(t+1) = \mathbf{U}_v^{-1} \left(\mathbf{X}_v^T \mathbf{Y}_v \lambda_v(t+1) - \mathbf{f}_v(t) \right), \tag{18}$$

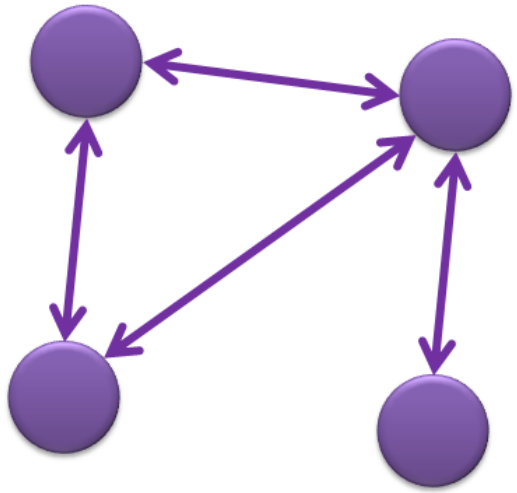
$$\alpha_v(t+1) = \alpha_v(t) + \frac{\eta}{2} \sum_{u \in \mathcal{B}_v} [\mathbf{r}_v(t+1) - \mathbf{r}_u(t+1)], \tag{19}$$

- $\mathbf{U}_v = (\mathbf{I}_{p+1} - \Pi_{p+1}) + 2\eta|\mathcal{B}_v|\mathbf{I}_{p+1}$
- $\mathbf{f}_v(t) = V_a C_l \delta_v^* + 2\alpha_v(t) - \eta \sum_{u \in \mathcal{B}_v} (\mathbf{r}_v(t) + \mathbf{r}_u(t)), \eta > 0.$

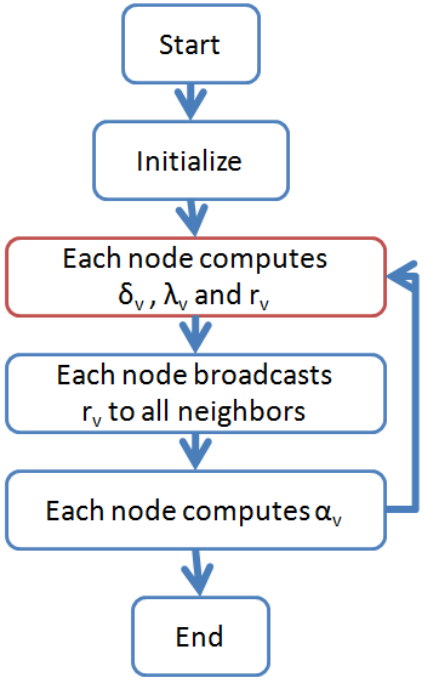
Secure and Resilient Algorithms

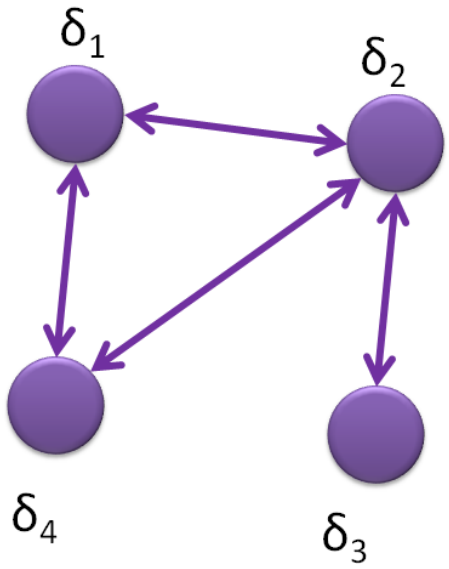
- Combine the solutions of Max-Problem with the solutions of Min-Problem.





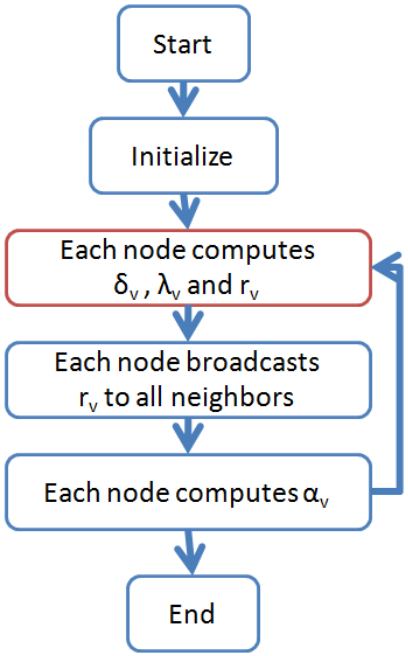
Algorithms:Step 1 Each Node Computes δ_v

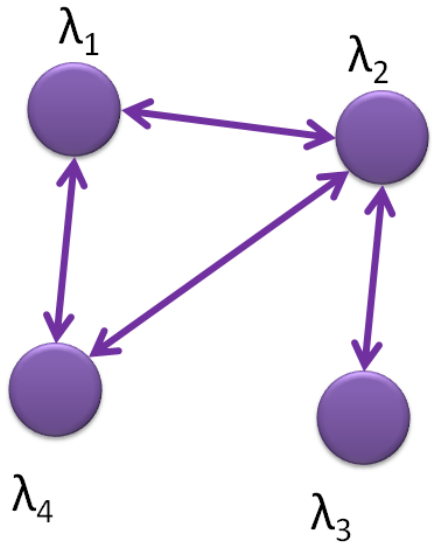




$$
\begin{aligned}
\delta_v(t+1) \in \arg \max_{\{\delta_v, s_v\}} \; & V_a C_I \mathbf{r}_v^T(t)(\mathbf{I}_{p+1} - \Pi_{p+1})\delta_v - \mathbf{1}^T s_v \\
& C_a \delta_v \leq s_v, \qquad \forall v \in \mathcal{V}_a; \\
\text{s.t.} \quad & C_a \delta_v \geq -s_v, \qquad \forall v \in \mathcal{V}_a; \\
& \delta_v \in \mathcal{U}_{v0}, \qquad \forall v \in \mathcal{V}_a.
\end{aligned} \tag{20}
$$

Algorithms:Step 2 Each Node Computes λ_v





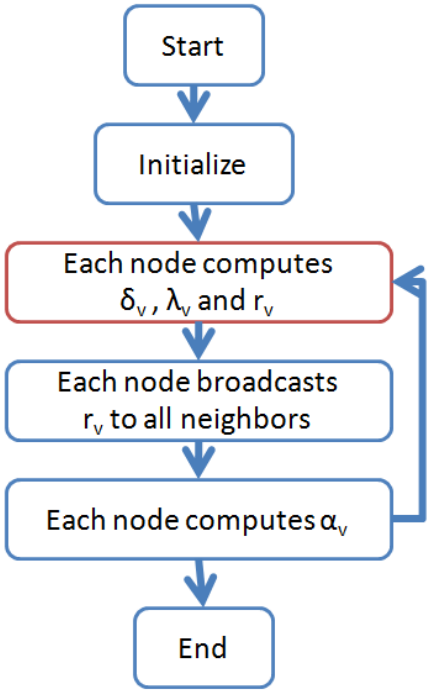
$$\lambda_v(t+1) \in \arg \max_{\mathbf{0} \le \lambda_v \le VC_l\mathbf{1}_v} -\frac{1}{2}\lambda_v^T \mathbf{Y}_v \mathbf{X}_v \mathbf{U}_v^{-1} \mathbf{X}_v^T \mathbf{Y}_v \lambda_v + \left(\mathbf{1}_v + \mathbf{Y}_v \mathbf{X}_v \mathbf{U}_v^{-1} \mathbf{f}_v(t)\right)^T \lambda_v, \tag{21}$$

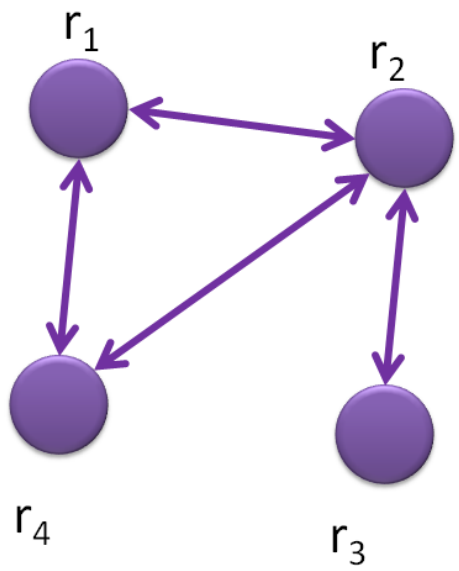
where

$$\mathbf{U}_v = (\mathbf{I}_{p+1} - \Pi_{p+1}) + 2\eta|\mathcal{B}_v|\mathbf{I}_{p+1},$$

$$\mathbf{f}_v(t) = V_a C_l \delta_v(t) + 2\alpha_v(t) - \eta \sum_{u \in \mathcal{U}_v} (\mathbf{r}_v(t) + \mathbf{r}_u(t)).$$

Algorithms:Step 3 Each Node Computes $\mathbf{r}_v$





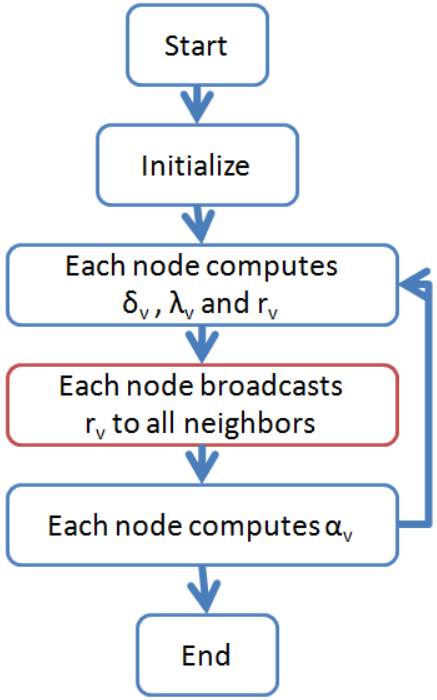
$$\mathbf{r}_v(t+1) = \mathbf{U}_v^{-1}\left(\mathbf{X}_v^T\mathbf{Y}_v\lambda_v(t+1) - \mathbf{f}_v(t)\right), \tag{22}$$

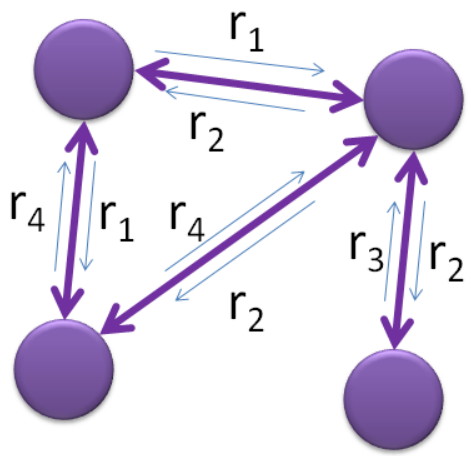
where

$$\mathbf{U}_v = (\mathbf{I}_{p+1} - \Pi_{p+1}) + 2\eta|\mathcal{B}_v|\mathbf{I}_{p+1},$$

$$\mathbf{f}_v(t) = V_a C_l \delta_v(t) + 2\alpha_v(t) - \eta \sum_{u\in\mathcal{U}_v}(\mathbf{r}_v(t) + \mathbf{r}_u(t)).$$

Algorithms:Step 4 Each Node Broadcasts $\mathbf{r}_v$ to Neighbors





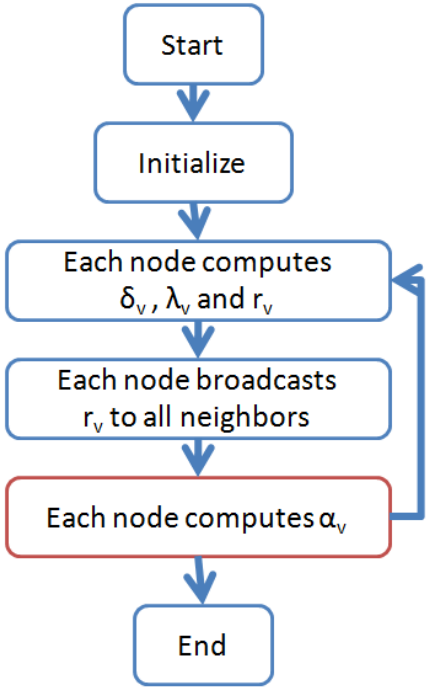
$$\mathbf{r}_v(t+1) = \mathbf{U}_v^{-1}\left(\mathbf{X}_v^T\mathbf{Y}_v\lambda_v(t+1) - \mathbf{f}_v(t)\right), \tag{23}$$

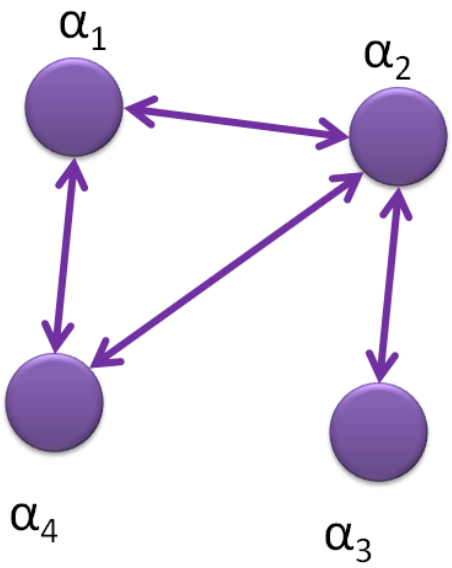
where

$$\mathbf{U}_v = (\mathbf{I}_{p+1} - \Pi_{p+1}) + 2\eta|\mathcal{B}_v|\mathbf{I}_{p+1},$$

$$\mathbf{f}_v(t) = V_a C_l \delta_v(t) + 2\alpha_v(t) - \eta \sum_{u\in\mathcal{U}_v}(\mathbf{r}_v(t) + \mathbf{r}_u(t)).$$

Algorithms:Step 5 Each Node Computes α_v

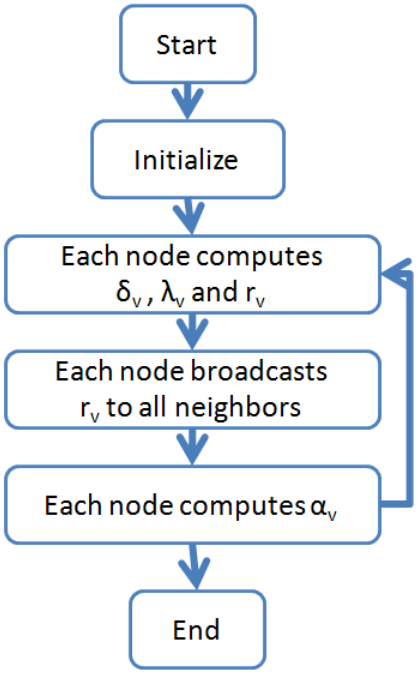


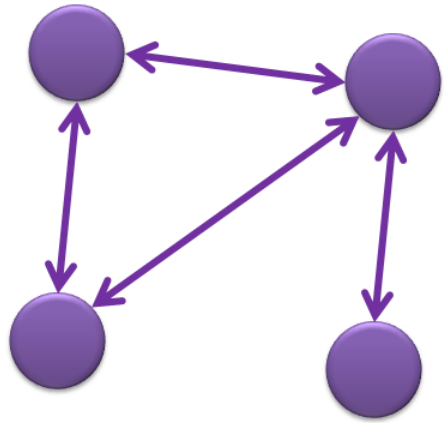


$$\alpha_v(t+1) = \alpha_v(t) + \frac{\eta}{2} \sum_{u \in \mathcal{B}_v} [\mathbf{r}_v(t+1) - \mathbf{r}_u(t+1)], \tag{24}$$

- **Return to Step 1, until convergence.**

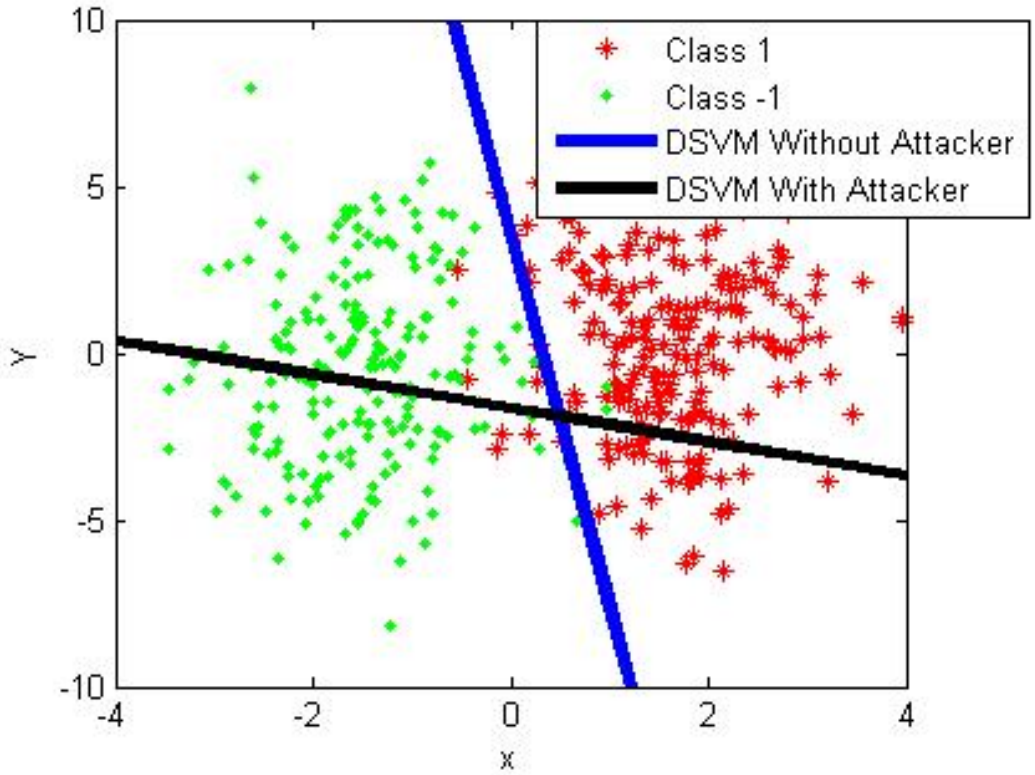
Secure and Resilient Algorithms





- It is **a fully decentralized network operation**, and it does not require exchanging training data or the value of decision functions.
- It provides the **reduced communication overhead** and **privacy preservation** at the same time.

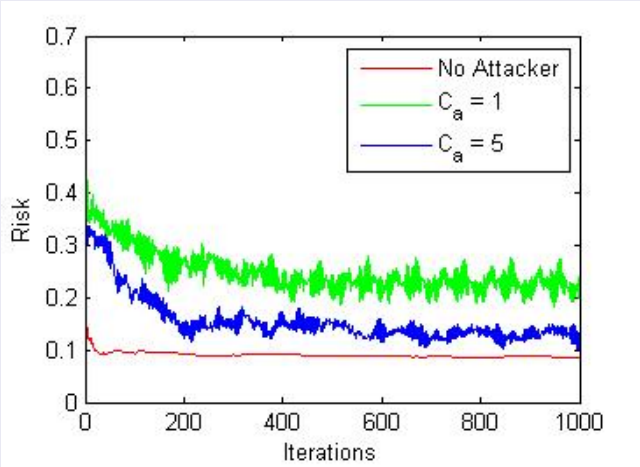
Numerical Experiment: Testing Experiment

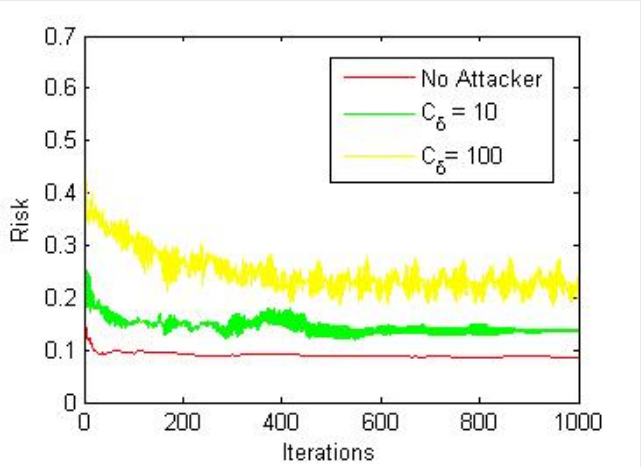


- A network with 4 nodes, each node contains 50 labeled 2-dimension training samples and 50 testing samples from global training set which is shown in this figure.
- The attacker uses the atomic action set with $C_\delta = 100$ and $C_a = 1$.

Effect of Parameters C_a and C_δ

Evolution of the risk of ADMoM-DSVM. Assume that attacker can attack any nodes. Network: 10 nodes, each node has 3 neighbors.

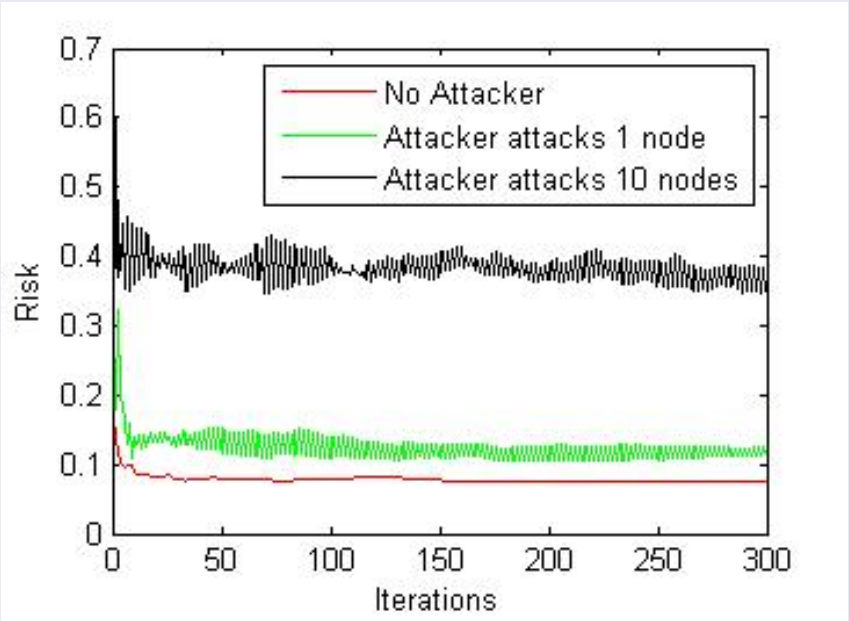




- The left figure shows that the **attacker has less influence** on learner's performance if we **increase the cost** of the attacker (C_a).
- The right figure shows that the **attacker has more influence** on learner if we **increase the size of attacker's action sets** (C_δ).

Effect of the Number of Nodes the Attacker Can Take Over

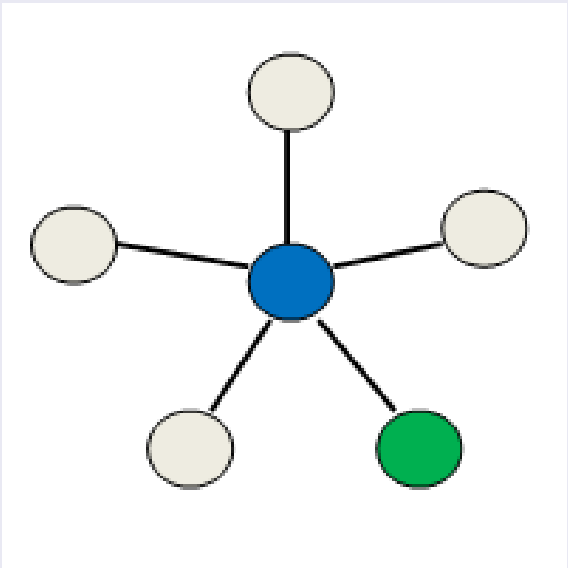
Evolution of the risk under different numbers of nodes the attacker can take over.



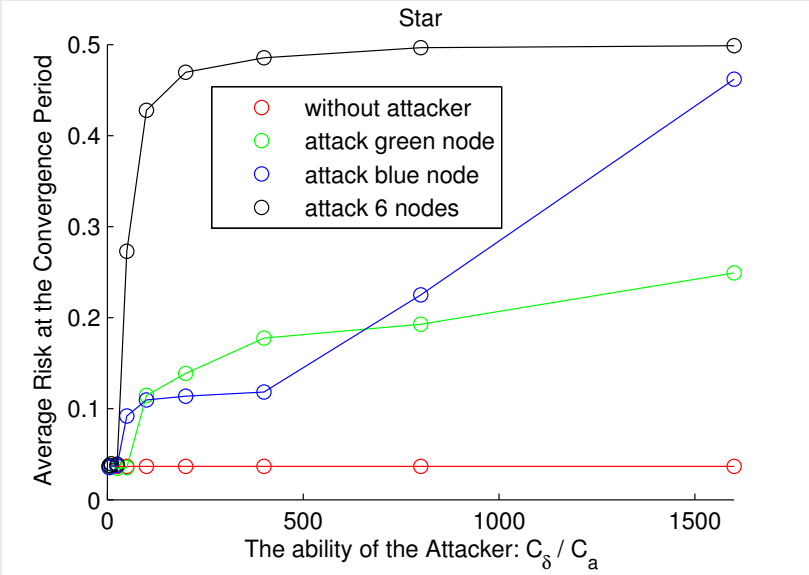
- We can conclude that if the **attacker** has the ability of **taking over more nodes**, he will create a **higher impact** on the learner.

Effect of Network Topology

Average risk in *star* network.



(c) Star



(d) Result:Star

- Fig. (c) shows the network topology.
- Fig. (d) shows the average risk.
- Nodes with **more neighbors (blue)** turn out to have **more influence** on performance as we increase the ability of the attacker (C_δ/C_a).

Conclusions

Contributions

- Capture the attacker's objective and constrained capabilities in a game-theoretic framework.
- Develop a nonzero-sum game to model the strategic interactions between an attacker and a learner with a distributed set of nodes.
- Show the strategic equivalence between the original nonzero-sum game and a zero-sum game.
- Demonstrate that network topology plays an important role in resilience to adversary behaviors.

Future works

- Design network topologies to achieve a desirable level of resiliency.
- Extend to distributed nonlinear SVMs, and other machine learning algorithms.