Trust and Resilience in Distributed Consensus Cyberphysical Systems

Michal Yemini
Joint work with Angelia Nedić, Stephanie Gil and Andrea Goldsmith

Princeton University

October 18, 2021

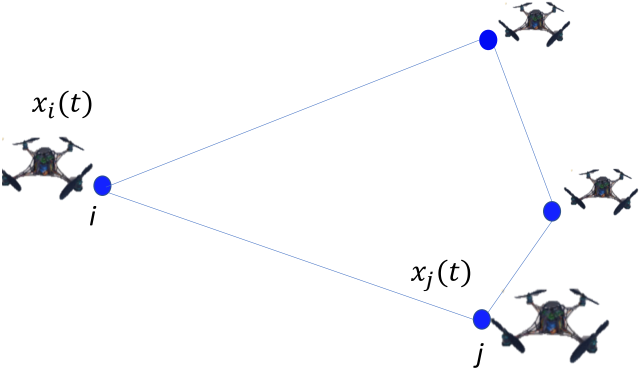
Outline

1. Distributed Consensus Systems.
2. Malicious Agents in Distributed Consensus Systems.
3. Agents' Trust Values in Cyberphysical Systems.
4. Characterizing Trust-Based Resilience in Distributed Consensus Systems.
5. Numerical Results.
6. Conclusions and Future Work.

Distributed Consensus Systems

Leaderless coordination and control for multi-agent systems.

- Robotic and drone networks (rendezvous problem).
- Sensor networks (data fusion - temperature measurement).
- Social networks (reaching a common opinion).



Mathematical of Modeling Distributed Consensus Systems

A **connected** graph $G = (\mathbb{V}\,, \mathbb{E})$, a **stochastic weight** matrix W and **initial vector** values $x(0)$.
For all $t \geq 0$

$$x_i(t+1) = w_{ii}x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij}x_j(t),$$

where $\mathcal{N}_i = \{j \in \mathbb{V} \mid \{i\,, j\} \in \mathbb{E}\}$ and

$$w_{ii} > 0\,, \qquad w_{ij} > 0 \quad \text{for all } j \in \mathcal{N}_i,$$

M. DeGroot 1970's (opinion dynamics), J. Tsitsiklis 1980's (distributed optimization)

It follows that

$$\lim_{t\to\infty} x_i(t) = \left[\lim_{t\to\infty} W^t x(0)\right]_i = [\mathbf{1}v'x(0)]_i = \lim_{t\to\infty} x_j(t)\,, \forall i\,, j$$

where v' is the Perron-Frobenius left-eigenvector of W.

Malicious Agents in Distributed Consensus Systems

In practice not all agents are legitimate (truthful), some are malicious and strategically input malicious values to either:

- prevent consensus,
- deviate the consensus from its true value.

The Classical Bound

The maximal number malicious agents that can be tolerated:

Legitimate agents can reach consensus iff the number of malicious agents is *less than 1/2 of the network connectivity*[1].

Proofs:

- Lamport, Pease and Shostak 1980, D. Dolev 1981 (Byzantine, fault tolerance, an additional condition),
- F. Pasqualetti, A. Bicchi and F. Bullo 2012 (control theory).

Both proofs assume that every legitimate agent knows the topology of G, and cannot detect malicious agents that only lie about their initial input values.

[1] The connectivity of a graph is the maximum number of disjoint paths between any two vertices of the graph.

Agents' Trust Values in Cyberphysical Systems I

Prior works have used the **data values** to overcome/detect malicious behavior. The **physical** aspects of the problem have not been considered. Namely, the **wireless communication channels**.

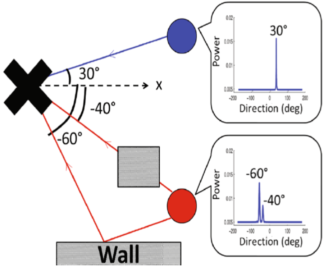
In cyberphysical systems:

- Malicious agents can lie about their location.
- A malicious agent can create many fictitious identities (Sybil attack).

Agents' Trust Values in Cyberphysical Systems II

Each transmitted signal leads to a received signal characteristics:

- Number of paths, delays.
- Angles of arrival.
- Power order of the angles of arrival.
- Power of the received signals.



*Guaranteeing spoof-resilient multi-robot networks, S. Gil *et al* 2017.

Agents' Trust Values in Cyberphysical Systems III

We can generate trust values that captures the event that an agent

- lies about its location
  - Location Verification Systems for VANETs in Rician Fading Channels, S. Yan *et al* 2016.
- uses a Sybil attack and creates multiple fictitious agents
  - Detecting Colluding Sybil Attackers in Robotic Networks using Backscatters Y. Huang *et al* 2021.

    (Limited to single antenna malicious agents.)
  - Guaranteeing spoof-resilient multi-robot networks, S. Gil *et al* 2017.

    (Limited to single antenna malicious agents.)
  - The Mason Test: A Defense Against Sybil Attacks in Wireless Networks Without Trusted Authorities, Liu *et al* 2015.

    (Assumes limited mobility of malicious agents and no beamforming).

We denote by $\alpha_{ij}(t) \in [0\,,1]$ the instantaneous single sample trust agent i gives agent j at time a t.

The Trust Based Distributed Consensus Model

Consider the system

$$\begin{bmatrix} X_{\mathcal{L}}(t+1) \\ X_{\mathcal{M}}(t+1) \end{bmatrix} = \begin{bmatrix} W_{\mathcal{L}}(t) & W_{\mathcal{M}}(t) \\ \Theta(t) & \Omega(t) \end{bmatrix} \begin{bmatrix} X_{\mathcal{L}}(t) \\ X_{\mathcal{M}}(t) \end{bmatrix},$$

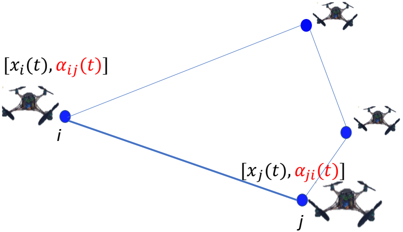
where $|x_i(t)| \leq \eta$ for every $i\,,j \in \mathcal{L} \cup \mathcal{M}$ and $t \geq 0$.

For every $i \in \mathcal{L}$:

$$x_i(t+1) = \underbrace{[1 - \sum_{j \in \mathcal{N}_i} W(i\,,j\,,t\,,\beta_{ij}(t))]}_{w_{ii}(t)}\, x_i(t) + \textstyle\sum_{j \in \mathcal{N}_i} \underbrace{W(i\,,j\,,t\,,\beta_{ij}(t))}_{w_{ij}(t)}\, x_j(t)$$

where

- $\beta_{ij}(t) = f(\alpha_{ij}(0)\,,\ldots,\alpha_{ij}(t))$



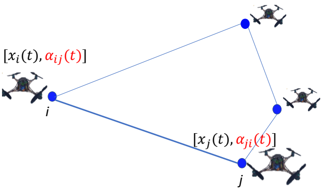
The Trust Based Distributed Consensus Model

For every $i \in \mathcal{L}$:

$$x_i(t+1) = \underbrace{\left[1 - \sum_{j \in \mathcal{N}_i} W(i\,,j\,,t\,,\beta_{ij}(t))\right]}_{w_{ii}(t)} x_i(t) + \sum\nolimits_{j \in \mathcal{N}_i} \underbrace{W(i\,,j\,,t\,,\beta_{ij}(t))}_{w_{ij}(t)} x_j(t)$$

where

- $\beta_{ij}(t) = f(\alpha_{ij}(0)\,,\ldots,\alpha_{ij}(t))$
- $w_{ii}(t) > 0$, $w_{ij}(t) \geq 0\,, j \in \mathcal{N}_i$, $\sum_{j \in \mathcal{N}_i} w_{ij} = 1$
- $w_{ij}(t) > 0\,, j \in \mathcal{N}_i \cap \mathcal{M}$ finitely many times a.s.
- $w_{ij}(t) = 0\,, j \in \mathcal{N}_i \cap \mathcal{L}$ finitely many times a.s.



Research Objectives

Objective I - Finite correct classification time

Establish characteristics of $\alpha_{ij}(t)$, and functions $\beta_{ij}(t)$ that lead to a **finite detection time** for the correct classification of legitimate and malicious agents **almost surely**.

Objective II - Convergence of the consensus protocol

Choose weights $W(i\,,j\,,t\,,\beta_{ij}(t))$ that allow **convergence** in spite of the presence of adversarial attacks.

Objective III - Bounded deviation for average consensus

We bound the **deviation** from the **true** consensus value, $\Delta\,(\delta)$ that can be achieved with a probability at least $1-\delta$.

Cumulative Trust Values

We assume that:

- $\alpha_{ij}(t)$ are statistically independent.
- There exist scalars $c < 0$ and $d > 0$ such that[2]

$$c = c_{ij} = E(\alpha_{ij}(t)) - 1/2 \qquad \text{for all } i \in \mathcal{L}\,,\ j \in \mathcal{N}_i \cap \mathcal{M},$$

$$d = d_{ij} = E(\alpha_{ij}(t)) - 1/2 \qquad \text{for all } i \in \mathcal{L}\,,\ j \in \mathcal{N}_i \cap \mathcal{L}.$$

To capture the **history** of observations $\alpha_{ij}(t)$, we define:

$$\beta_{ij}(t) = \sum_{k=0}^{t} (\alpha_{ij}(k) - 1/2) \ \text{ for } t \geq 0\,, i \in \mathcal{L}\,, j \in \mathcal{N}_i.$$

Agent i classifies agent j as legitimate if $\beta_{ij}(t) \geq 0$ and malicious otherwise.

[2] For the sake of simplicity of presentation.

Finite Correct Classification Time I

Lemma

For every $t \geq 0$ *and* $i \in \mathcal{L}$

$$\Pr\left(\beta_{ij}(t) < 0\right) \leq \exp(-2(t+1)d^2)\,,\, j \in \mathcal{N}_i \cap \mathcal{L},$$
$$\Pr\left(\beta_{ij}(t) \geq 0\right) \leq \exp(-2(t+1)c^2)\,,\, j \in \mathcal{N}_i \cap \mathcal{M}.$$

This is an immediate result of the Chernoff-Hoeffding Inequality.

Proposition

There exists a (random) finite time instant $T_f > 0$ *such that every legitimate agent* i *correctly classifies its neighbors for all* $t \geq T_f$ *almost surely.*

This proposition follows by the Borel-Cantelli Lemma

The Modified Trust Based Weights

Define the time dependent **trusted neighborhood** for agent i:

$$\mathcal{N}_i(t) = \{j \in \mathcal{N}_i : \ \beta_{ij}(t) \geq 0\},$$

We choose for all $i \in \mathcal{L}$,

$$w_{ij}(t) = \begin{cases} \mathbb{1}_{\{t \geq T_0 - 1\}} \cdot \min\left\{\frac{1}{\kappa}, \frac{1}{|\mathcal{N}_i(t)|+1}\right\} & \text{if } , j \in \mathcal{N}_i(t), \\ 0 & \text{if } , j \notin \mathcal{N}_i(t) \cup \{i\}, \\ 1 - \sum_{m \in \mathcal{N}_i} w_{im}(t) & \text{if } j = i. \end{cases},$$

where $\kappa > 0$ is a limiting effect constant.
Up to time T_0 agents measure the trust values of their neighbors but don't update their data values.

The Data Values of the Legitimate Agents

Recall that: $$\begin{bmatrix} X_{\mathcal{L}}(t+1) \\ X_{\mathcal{M}}(t+1) \end{bmatrix} = \begin{bmatrix} W_{\mathcal{L}}(t) & W_{\mathcal{M}}(t) \\ \Theta(t) & \Omega(t) \end{bmatrix} \begin{bmatrix} X_{\mathcal{L}}(t) \\ X_{\mathcal{M}}(t) \end{bmatrix}.$$

Thus,

$$x_{\mathcal{L}}(t) = \tilde{x}_{\mathcal{L}}(t) + \phi_{\mathcal{M}}(t),$$

where[3]

$$\tilde{x}_{\mathcal{L}}(t) = \left(\prod_{k=T_0-1}^{t-1} W_{\mathcal{L}}(k) \right) x_{\mathcal{L}}(0),$$

and

$$\phi_{\mathcal{M}}(t) = \sum_{k=T_0-1}^{t-1} \left(\prod_{l=k+1}^{t-1} W_{\mathcal{L}}(l) \right) W_{\mathcal{M}}(k) x_{\mathcal{M}}(k).$$

[3]Note that $W_{\mathcal{L}}(k)$ can be substochastic.

Convergence of the Consensus Protocol I

Define a matrix $\overline{W}_{\mathcal{L}}$ such that for every $i\,,j \in \mathcal{L}$,

$$[\overline{W}_{\mathcal{L}}]_{ij} = \begin{cases} \min\left\{\frac{1}{\kappa}\,, \frac{1}{|\mathcal{N}_i|+1}\right\} & \text{if } j \in \mathcal{N}_i \cap \mathcal{L}, \\ 1 - \min\left\{\frac{|\mathcal{N}_i \cap \mathcal{L}|}{\kappa}\,, \frac{|\mathcal{N}_i \cap \mathcal{L}|}{|\mathcal{N}_i|+1}\right\} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

Then, almost surely there exists a (random) finite time T_f such that

$$\prod_{k=T_0-1}^{\infty} W_{\mathcal{L}}(k) = \underbrace{\lim_{k\to\infty} \overline{W}_{\mathcal{L}}^{k-\max\{T_f,T_0\}}}_{\mathbf{1}v'} \prod_{k=T_0-1}^{\max\{T_f,T_0\}-1} W_{\mathcal{L}}(k),$$

and $W_{\mathcal{M}}(t) = \mathbf{0}$ for every $t > T_f$.

Convergence of the Consensus Protocol II

Proposition

Almost surely, there exists a random variable $z(T_0)$ *such that*

$$\lim_{t\to\infty} x_{\mathcal{L}}(t) = z(T_0)\mathbf{1},$$

where $z(T_0)$ *is in the convex hull of the initial values* $x_i(0)\,, i \in \mathcal{L} \cup \mathcal{M}$*, and its distribution depends on the starting time* T_0 *of the data passing phase.*

The Deviation from Nominal Consensus Value

Theorem

Given an error level $\delta > 0$, *we have the following result*

$$\Pr\left(\max_{i\in\mathcal{L}} \limsup_{t\to\infty} \left|\left[x_{\mathcal{L}}(t) - \mathbf{1}v' x_{\mathcal{L}}(0)\right]_i\right| \leq \Delta_{max}(T_0\,,\delta)\right) \geq 1-\delta,$$

where $\Delta_{max}(T_0\,,\delta) = 2\left[\tilde{g}_{\mathcal{L}}(T_0\,,\delta) + \tilde{g}_{\mathcal{M}}(T_0\,,\delta)\right],$

$$\tilde{g}_{\mathcal{L}}(\delta) = \frac{\eta|\mathcal{L}|^2}{\delta}\cdot\frac{\exp(-2T_0 d^2)}{1-\exp(-2d^2)} + \frac{\eta|\mathcal{L}||\mathcal{M}|}{\delta}\cdot\frac{\exp(-2T_0 c^2)}{1-\exp(-2c^2)},$$

and

$$\tilde{g}_{\mathcal{M}}(T_0\,,\delta) = \frac{\eta|\mathcal{L}||\mathcal{M}|}{\delta\cdot\kappa}\cdot\frac{\exp(-2T_0 c^2)}{1-\exp(-2c^2)}.$$

$x_{\mathcal{L}}(t) = \tilde{x}_{\mathcal{L}}(t) + \phi_{\mathcal{M}}(t) \Rightarrow$
$|x_{\mathcal{L}}(t) - \mathbf{1}v' x_{\mathcal{L}}(0)|_i \leq |\tilde{x}_{\mathcal{L}}(t) - \mathbf{1}v' x_{\mathcal{L}}(0)|_i + |\phi_{\mathcal{M}}(t)|_i.$

A Few Words regarding the Expected Convergence Time I

Proposition

Assume that $j \in \mathcal{N}_i \Leftrightarrow i \in \mathcal{N}_j$ (symmetric connectivity of legitimate agents). Then, for every $T_0 \geq 0$ and $t \geq T_0$, we have

$$
\begin{aligned}
& E\left(\left\|x_{\mathcal{L}}(t) - \mathbf{1}v' x_{\mathcal{L}}(0)\right\|_v\right) \\
& \leq 2\left(\frac{t-T_0}{2} + 1\right) \rho_2^{\frac{t-T_0}{2}} \eta + \left(\frac{|\mathcal{L}|^2 \exp(-(t+T_0+2)d^2)}{1-\exp(-2d^2)}\right. \\
& \quad \left. + \frac{|\mathcal{L}||\mathcal{M}| \exp(-(t+T_0+2)c^2)}{1-\exp(-2c^2)}\right) 2\eta \\
& = O\left(|\mathcal{L}| \cdot \max\left\{|\mathcal{L}|, |\mathcal{M}|\right\} \cdot t e^{-\gamma t}\right),
\end{aligned}
$$

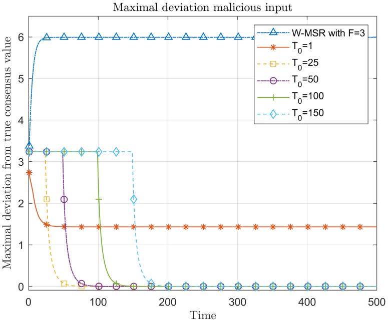
where $\rho_2 < 1$ is the second largest eigvenvalue modulus of $\overline{W}_{\mathcal{L}}$ and $v > \mathbf{0}$ be the stochastic Perron vector satisfying $v'\overline{W}_{\mathcal{L}} = v'$.

Numerical Results

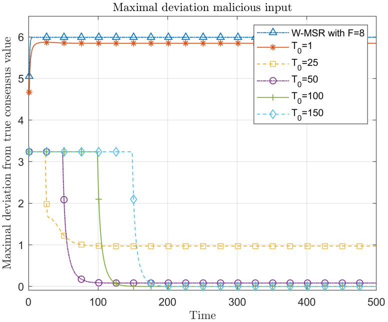
- $|\mathcal{L}|= 15$ legitimate agents
- $|\mathcal{M}|= 5\,, 15\,, 30$
- $\eta = 5$, $\kappa = 10$;
- $E(\alpha_{ij}) = 0.55$ for $i \in \mathcal{L}\,,\ j \in \mathcal{N}_i \cap \mathcal{L}$,
- $E(\alpha_{ij}) = 0.45$ for $i \in \mathcal{L}\,,\ j \in \mathcal{N}_i \cap \mathcal{M}$,
- $\alpha_{ij} \sim U\left[E(\alpha_{ij}) - \frac{\ell}{2}\,, E(\alpha_{ij}) + \frac{\ell}{2}\right]$
- $\ell = 0.2\,, 0.4\,, 0.6$

Classical bound must fulfill $|\mathcal{M}|< \frac{3+|\mathcal{M}|}{2} \Rightarrow |\mathcal{M}|< 3$.

Numerical Results - Maximum Deviation Input I

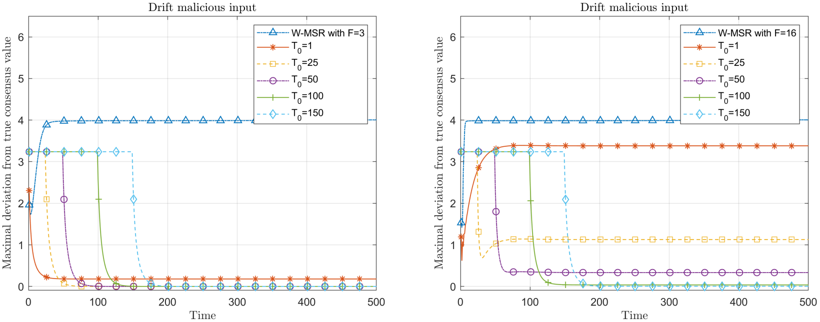


(a) $|\mathcal{M}| = 5$, $\ell = 0.2$



(b) $|\mathcal{M}| = 15$, $\ell = 0.4$

Numerical Results - Drift Input I



(a) $|\mathcal{M}| = 5$, $\ell = 0.2$ (b) $|\mathcal{M}| = 30$, $\ell = 0.6$

Conclusions and Future Work

- Physical based trust values to brake the current known bound
- Modified weight matrix - based on trust values
- Finite detection time a.s., convergence, deviation from true consensus value
- Future work

Thank You!

Questions? Collaboration ideas?
Email: myemini@princeton.edu