Intelligent Physical Attacks against Mobile Robotic Networks

Yushan Li, **Jianping He**, Xuda Ding, Lin Cai and Xinping Guan

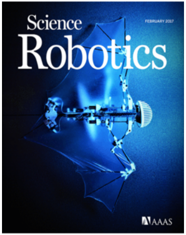
Shanghai Jiao Tong University
jphe@sjtu.edu.cn

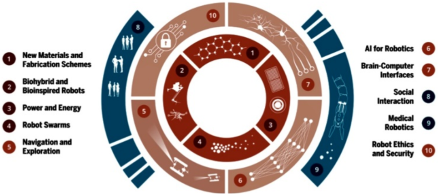
May 17, 2021

Outline

Introduction

- **What is mobile robotic network (MRN)**
  A networked system of multiple mobile robots, where the robots interact and cooperate with each other to achieve well defined tasks
- **Why adopt MRN**
  - Higher flexibility and robustness than single robot
  - Parallel operation in spatio-temporal tasks
  - Coordinated ability of acquiring and processing information

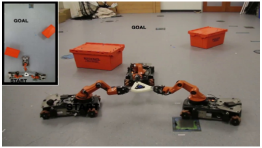




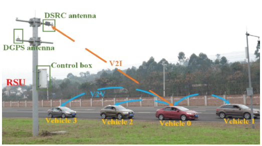
Source: [1] G.-Z.Yang, et al., Science Robotics, 2018.

Applications

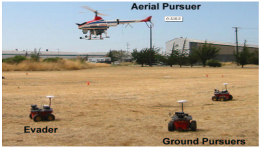
- MRN is widely deployed in military and industrial applications



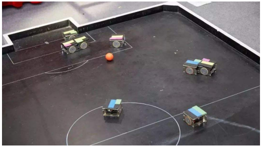
(c) Manipulation[2]



(d) Platoon[3]



(e) Pursuit-evasion[4]



(f) Combat[5]

(g) Military surveillance[6]

(h) UAV swarm[7]

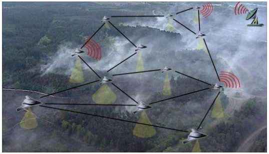
Local sense + Information interaction + Action decision ⇒ Cooperation

Source: [2] J.A. Mora et al., Int. J. Rob. Res., 2017. [3] Y. Li et al., IEEE Trans. Intell. Transp. Syst., 2019. [4] R. Vidal et al., IEEE Trans. Rob. and Autom., 2002. [5] FIRA Cup, 1997. [6] www.joao-valente.com/doku.php?id=wiki:research. [7] Article: www.prophecynewswatch.com/article.cfm?recent_news_id=3782

Vulnerabilities in Interaction

- Interaction is critical for MRNs, however there are situations where
  - sensor reading is interfered
  - communication is monitored or even hijacked
  - certain robot is corrupted as an adversary

(i) disturb sensors[8]



(j) communication leak[9]

(k) mislead the swarm[10]

- Interaction can be maliciously utilized, causing severe threats
- Urgent and vital to tackle the security vulnerabilities of MRNs

Source: [8] ICRA DJI RobotMaster Competition. [9] www.sohu.com/a/241170554_358040 [10] www.sohu.com/a/240072583_465915

Related Work

- The research about security of MRNs mainly focus on two aspects

Table 1 Related work

	characteristics	representative works
Cyber aspects	mainly focus on defense design of common cyber attacks	DoS, replay attacks false data injection (see [11]-[14] for review)
Physical aspects	against specific transducer straightforward to implement	alter gyroscopic sensor[15] disturb GPS readings[16] heat up memory cell[17]

[11] F. Pasqualetti et al., IEEE TAC, 2013. [12] Y. Mo et al., IEEE TAC, 2015. [13] H. Sandberg et al., IEEE Control Syst. Mag., 2015. [14] H.S. Sanchez et al., Annual Reviews in Control, 2019. [15] Y. Son et al., USENIX Security Symposium, 2015. [16] N.O. Tippenhauer et al., ACM CCS, 2011. [17] S. Skorobogatov, IEEE International Workshop on Hardware-Oriented Security and Trust, 2009.

Motivations

- **Motivation**
  - Powerful abilities and knowledge are typically assumed for attacker
    - master system structure[18]
    - control data and measurements are corrupted[19]
    - communication link is altered[20]

  Passive design form, analysis simplicity but unrealistic for attacker
    - control-communication is protected with strong encryption[21]
    - system structure is unknown beforehand and can dynamically change[22]
  - Physical attacks mainly focus on specific sensor, not generalized
- **What we investigate**
  - generalized and intelligent attacks with weak knowledge of MRNs

    ▷ Entrap a robot ▷ Sneak into the MRN
    - what other knowledge to learn? how to learn?
    - how to design attack strategies? how to optimize the performance?

[18] F. Pasqualetti et al., IEEE TAC, 2012. [19] R. Su et al. Automatica, 2015. [20] Z. Feng et al. Int. J. Robust Nonlinear Control, 2016. [21] M.S. Darup et al., IEEE Control Syst. Lett., 2018. [22] M. Khalili et al., Automatica, 2018.

Contribution

- **Main contributions of this work**
  - We reveal the learnability of the interaction rules in MRN
    - weak prior knowledge, without system dynamics or internal access
    - partial observation and bounded moving abilities
  - We design intelligent physical attacks against MRNs
    - obstacle-disguising attack: fool a victim into preset trap
    - sneak attack: replace a target robot in the MRN
  - We analyze and optimize the attack performance
    - the feasibility criterion is provided
    - the bound of attack cost is proved

MRN Modeling

- **Goal**: The MRN $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ runs to goal z_g with pre-defined shape
  - directed network structure
    - ▷ interaction weight $a_{ij} > 0$ indicates j sends information to i
    - ▷ in-neighbor $\mathcal{N}_i^{in} = \{j \in \mathcal{V} : a_{ij} > 0\}$ out-neighbor $\mathcal{N}_i^{out} = \{j \in \mathcal{V} : a_{ji} > 0\}$
  - consensus-based formation control

$$\dot{z}_i = \sum_{j \in \mathcal{N}_i^{in}} a_{ij}(z_j - z_i - h_{ij}), \quad \dot{z}(t) = -Lz(t) + Lh$$

▷ z_i: state of robot $i \in \mathcal{V}$ ▷ $\{h_{ij}\}$: shape configuration ▷ L: Laplacian matrix of $\mathcal{G}$

  - obstacle-avoidance mechanism g

$$\dot{z}_i = g(z_{ob} - z_i, z_{i*} - z_i, v_{ob}, v_i).$$

▷ z_{i*}: the desired state of i ▷ z_{ob} and v_{ob}: the state and velocity of the obstacle

Attacker Modeling

- **Goal**: Observe $\mathcal{G}$ and learn the interaction rules, then launch attacks
  - Discrete dynamics

$$z^{k+1} = (I - \varepsilon_T L)z^k + \varepsilon_T u^k = W z^k + \varepsilon_T u^k$$

▷ ε_T - the sampling period ▷ formation input $u = Lh + [0 \cdots 0 \; c]^\mathsf{T}$

Note: W equivalently represents the internal interaction structure as L
MRN division $\mathcal{V} = \mathcal{V}_F \cup \mathcal{V}_{F'}$

$$\begin{bmatrix} z_F^{k+1} \\ z_{F'}^{k+1} \end{bmatrix} = \begin{bmatrix} W_{FF} & W_{FF'} \\ W_{F'F} & W_{F'F'} \end{bmatrix} \begin{bmatrix} z_F^k \\ z_{F'}^k \end{bmatrix} + \varepsilon_T \begin{bmatrix} u_F^k \\ u_{F'}^k \end{bmatrix}$$

▷ F - observable part ▷ F' - unobservable part

  - Under partial observation over $\mathcal{V}_F \subseteq \mathcal{V}$

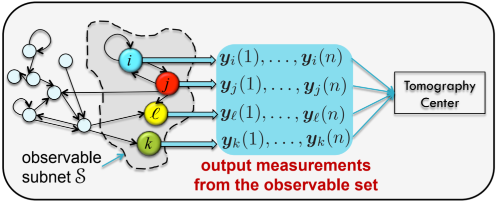
$$\tilde{z}_F^{k+1} = W_{FF}\tilde{z}_F^k + \varepsilon_T \hat{u}_F^k + \xi_F^k + W_{FF'}\tilde{z}_{F'}^k \Rightarrow \text{influenced by unobservable part}$$

▷ $\tilde{\cdot}$ indicates observations ▷ ξ^k is i.i.d zero-mean Gaussian observation noise

  - Bounded moving ability $\|u_a(k)\|_2 \leq \mu$

Key Ideas

- **Inspirations:** formation control is fundamentally adopted to keep a pre-defined geometric shape in applications of MRN
  - In shape forming and maintaining, internal interaction structure determines the convergence speed and stability
  - In obstacle/collision avoidance, external interaction mechanism steers robots to adapt the environment obstacles



Source: A. Santos, et al., IEEE Transactions on Information Theory, 2019.

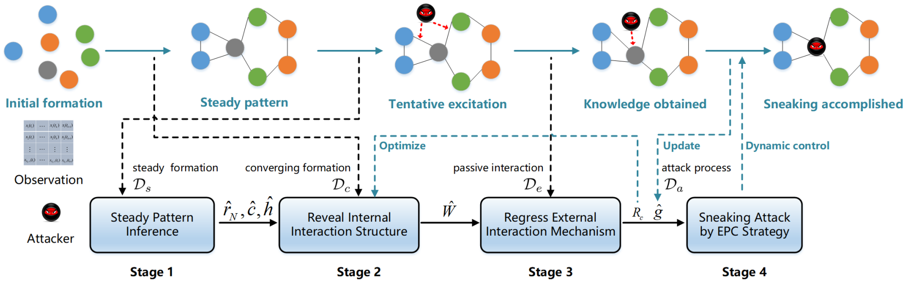
- **Insight:** the state evolution of MRN reveals the interaction rules

excite the robot and observe the reaction

Attack Formulation

▶ Overview

Characterize the whole process as four stages $\Rightarrow$ record dataset



▶ Process description

- shape forming (observe) $\Rightarrow \mathcal{D}_c$
- formation maintenance (observe) $\Rightarrow \mathcal{D}_s$
- tentatively trial (excite) $\Rightarrow \mathcal{D}_e$
- entrap/sneak (attack) $\Rightarrow \mathcal{D}_a$

} Infer knowledge from datasets
Design attacks using knowledge

Steady Pattern Identification

▶ Steady Pattern Identifiability

- **Theorem:**[state separability] Suppose $\mathcal{G}$ has a spanning tree, under $u = Lh + [0 \cdots 0\ c]^{\mathsf{T}}$, we have

$$\lim_{t \to \infty} \|z(t) - ct \cdot \mathbf{1} - s\|_2 = 0,$$

▷ c - leadership velocity ▷ s - offset vector and $(s - s^{[i]}\mathbf{1})$ is equivalent to Lh

Note

- the convergence is guaranteed by the spanning tree structure
- the state can be divided into: common speed and specified shape
- providing the feasibility to infer the steady pattern

- How to obtain the steady pattern parameters?

Steady Pattern Identification

▶ Calculation procedures

- Define 2nd-order state difference accumulation

$$\Delta S_i^{k_0:k_0+l} = \sum_{k=k_0+1}^{k_0+l-1} \|\Delta z_i^{k+1} - \Delta z_i^k\|_2 \;\Leftarrow\; \text{time window } [k_0, k_0+l]$$

  - Step 1: find the ϵ-convergence time of the steady pattern

$$k^* = \inf \left\{ k_0 : \left(\sum_{i \in \mathcal{V}_F} \Delta \tilde{S}_i^{k_0:k_0+l} \right) \leq \epsilon \right\}$$

  - Step 2: compute the steady velocity

$$\hat{c}(k^*, l) = \arg\min_c \sum_{k=k^*}^{k^*+l} \left\| \tilde{z}_F^k - (c\varepsilon_T k + b_0)\mathbf{1} \right\|_2^2 \quad \triangleright \mathbf{1} \text{ - all-one vector}$$

  - Step 3: derive the formation shape configuration

$$\hat{h} = \hat{s} - \hat{s}_j \mathbf{1}, \text{ where } \hat{s} = \sum_{k=k^*+1}^{k^*+l} (\tilde{z}_F^k - \hat{c}\varepsilon_T k \cdot \mathbf{1})/l$$

- steady pattern determined $\Rightarrow$ converging process also determined

Internal Interaction Structure Approximation

▶ Structure inference under partial observation

- Recalling observations over $\mathcal{V}_F \subseteq \mathcal{V}$

$$z_F^{k+1} = W_{FF} z_F^k + \varepsilon_T \hat{u}_F^k + \xi_F^k + W_{FF'} z_{F'}^k$$ ⇒influenced by unobservable part

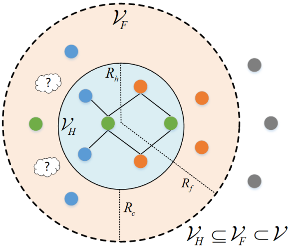
- Information shortage
  - inevitably incur large error to infer W_{FF} directly

- Transform inference objective
  - narrow down inference set $\mathcal{V}_H \subseteq \mathcal{V}_F$
  - range determination

$$R_f > R_c, \quad R_h = R_f - R_c.$$

▷ R_c - interaction range of the robots

▷ R_f - radius of $\mathcal{V}_F$ ▷ R_h - radius of $\mathcal{V}_H$



Internal Interaction Structure Approximation

▶ Approximation modeling

- **Theorem:**[structure approximation] Using linear state space model, observations in $\mathcal{D}_c$ satisfy

$$y_H^{k+1} = W_{HF} y_F^k,$$

▷ $\begin{cases} y_H^k = \tilde{z}_H^k - \hat{h}_H - \varepsilon_T \hat{c} \mathbb{I}_H \\ y_F^k = [(\tilde{z}_H^k - \hat{h}_H)^\mathsf{T}, (\tilde{z}_{H'}^k)^\mathsf{T}]^\mathsf{T} \end{cases}$ ▷ $\mathbb{I}_F^{[i]} = \begin{cases} 1, & \text{if } i \in \mathcal{V}_F \text{ is the leadership} \\ 0, & \text{otherwise.} \end{cases}$

  - linear model provides simplicity for the structure representation

- How to approximate W_{HF}?

  Corollary: If $|\mathcal{V}_F| + 1 \leq l \leq k^*$, the least square estimation of W_{HF} is

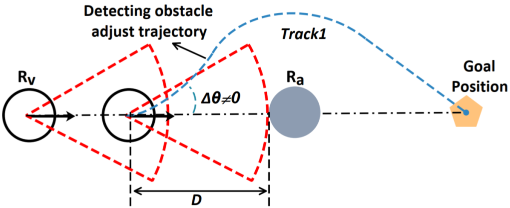
$$\phi(\mathcal{D}_c): \ \hat{W}_{HF} = \left((Y_F Y_F^\mathsf{T})^{-1} Y_F Y_H^\mathsf{T} \right)^\mathsf{T},$$

  ▷ $Y_H = [y_H^2, y_H^3, \cdots, y_H^l]$ ▷ $Y_F = [y_F^1, y_F^2, \cdots, y_F^{l-1}]$

  Note: converging time k^* and data amount before k^* determines the feasibility and accuracy

External Interaction Mechanism Regression

▶ **Key idea:** r_i will deviate its ideal trajectory once an obstacle detected



▷ r_i - robot i

▷ r_a - attack robot

▷ r_v - victim robot

- **Definition:** A node is directly controllable if one can control it to reach any given state z_c^* in finite steps by direct external excitations.
- **Theorem:** If g is known, and $(z_{i*} - z_i)$, $(z_a - z_i)$ and v_i are measurable, then r_i is directly controllable by r_a.
  - g determines the avoidance behavior $\Rightarrow$ causal relationship
  - given a input configuration, the output is unique $\Rightarrow$ regression feasibility
- How to approximate g? $\Leftarrow$ from effects to reveal the causes

External Interaction Mechanism Regression

- **Regression procedures**
  - Obtain input configuration
    - Based on $\hat{W}_{HF}$, the desired position of r_i is

$$\hat{z}_{i*}^{k+1} = \sum\nolimits_{j \in \mathcal{V}_F} \hat{a}_{ij} (\tilde{z}_j^k - \tilde{z}_i^k - h_F^{[j]} + h_F^{[i]}),$$

▷ $\hat{a}_{ij} = \hat{w}_{ij} / \varepsilon_T \ (i \neq j)$

    - z_i and v_i are measurable under fast-rate sampling
  - Tentatively excite the target robot and record its reaction

$$Q_{in}^k = [\tilde{z}_v^k - \tilde{z}_a^k, \tilde{z}_{v*}^k - \tilde{z}_v^k, \Delta\tilde{z}_v^k / \varepsilon_T, \Delta\tilde{z}_a^k / \varepsilon_T], \quad Q_{out}^k = \Delta\tilde{z}_v^{k+1}$$

Note: R_c and obstacle detection range $\mathcal{A}_d$ is also inferred by trial[23]

  - Construct $\mathcal{D}_e = \{\cup \{Q_{in}^k, Q_{out}^k\}\}$ and regress g

$$\hat{g} = \underset{g: Q_{in} \mapsto Q_{out}}{\arg\min} \sum\nolimits_{k=1}^{L'} \left\| Q_{out}^k - g(Q_{in}^k) \right\|_2$$

    - many mature learning methods are available, e.g., SVR.

[23] Y. Li, et al., IEEE ACC, 2019.

Attack 1: Entrap a Robot

▶ **Shortest-path strategy:** the path length from the position where r_v is initially attacked to preset trap is shortest ⇒ optimize direct attack cost

$$\mathbf{P_1}: \min_{H,\boldsymbol{u}_{a,0:H}} C_s(\boldsymbol{u}_{a,0:H}) = \sum_{k=0}^{H} \|\hat{z}_v(k+1) - z_v(k)\|_2$$

$$\begin{aligned} \text{s.t.} \quad & \|u_a(k)\|_2 \le \mu, && \Leftarrow \text{bounded velocity} \\ & \|z_v(H) - z_t\|_2 \le \delta, && \Leftarrow \text{driven into trap} \\ & \eta \le \|z_a(k) - z_v(k)\|_2, && \Leftarrow \text{not too close} \\ & p_a(k) \in \mathcal{A}_d(z_v(k)). && \Leftarrow \text{continuous excitation} \end{aligned}$$

} hard to solve analytically, using heuristic methods

- **Theorem:**[path length] By the shortest-path strategy, we have

$$(\pi/2 + \xi - \cos\xi) r_{\min} + d_{te}(\cos\xi - 1) \le C_s - C_s^* \le (\frac{7}{6}\pi - 1 - \sqrt{3}) r_{\max},$$

▷ $r_{\min}/r_{\max}$ - min/max reaction radius ▷ $d_{te} = \|z_t - z_v(0)\|_2$ ▷ $\xi = \arcsin(\frac{r_{\min}}{d_{te} - r_{\min}})$

  - sub-optimal but efficient
  - upper bound indicates worst case, hard to meet

Attack 1: Entrap a Robot

▶ **Hands-off strategy:** fool r_v into the trap with the maximum hands-off state ratio (sparsity) during the attack ⇒ optimize attack stealth

$$\begin{aligned}
\mathbf{P_2}: \quad & \min_{H, \boldsymbol{u}_{a,0:H}} \; C_h(\boldsymbol{u}_{a,0:H}) = \|\boldsymbol{u}_{a,0:H}\|_0 \\
\text{s.t.} \quad & \|u_a(t)\|_2 \leq \mu, \\
& \|z_v(H) - z_t\|_2 \leq \delta, \\
& \eta_1 \leq \|z_a(t) - z_v(t)\|_2 \leq \eta_2, \quad \Leftarrow \text{relax excitation constraint}
\end{aligned}$$

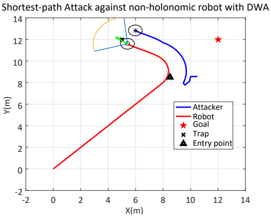
Hard to be solved analytically ⇒ using heuristic based methods

- **Theorem:**[active period] By the hands-off strategy, we have

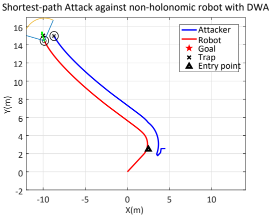
$$C_h(\boldsymbol{u}_{a,0:H})/H \leq 0.5.$$

  - largely reduce the activity of r_a during the process
  - feasibly counter some threshold-based anomaly detection techniques

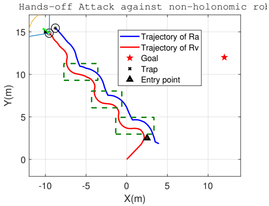
- **Examples**



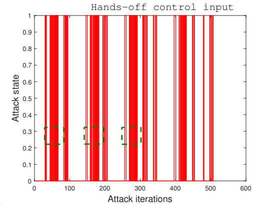
(l) Case 1 of S-attack



(m) Case 2 of S-attack



(n) Case of H-attack



(o) Input under H-attack

- **Sneak attack:** r_a sneaks into the MRN $\mathcal{V}$ by replacing $r_v \in \mathcal{V}$.
  - The state update of $i \in \mathcal{V}$ is influenced by its in-neighbors $\mathcal{N}_i^{in}$
    - whose impact is larger ?
  - **Definition:** A node is indirectly controllable if one can control another node to chainnedly make it reach any z_c^* in finite steps.
  - **Lemma:** Given desired state z_c^* and initial state z_i^0, r_i is indirectly controllable by r_j iff

$$\begin{cases} u_e u_c > 0, & \text{if } (z_c^* - z_i^0)u_c > 0, \\ |p_{1j} u_e| > |p_{1N} u_c|, & \text{if } (z_c^* - z_i^0)u_c < 0, \end{cases}$$

▷ $p_1 = [p_{11}, \cdots, p_{1N}]^\mathsf{T}$ is the left eigenvector for λ_1 of L.

- Sufficient and necessary condition, requiring network structure L
- Unavailable under partial observation

Attack 2: Sneak into MRN

▶ Attack feasibility

- **Theorem :** Given z_c^* and z^0, r_i is indirectly controllable by r_j when

$$\begin{cases} u_e u_c > 0, & \text{if } (z_c^* - z_i^0)u_c > 0, \\ |a_{ij} u_e| > |\bar{a}_{ij} u_c|, & \text{if } (z_c^* - z_i^0)u_c < 0. \end{cases}$$

▷ $\bar{a}_{ij} = \sum_{j' \in \{\mathcal{N}_i^{in} \setminus j\}} a_{ij'}$ ▷ u_e - excitation input of r_j ▷ u_c - leadership input

Note:

- sufficient condition, without relying on global network structure
- available under partial observation
- provide attack feasibility

- How to design the attack strategy?

Key idea: find the most valuable target robot r_v, steer it out of the interaction range of its neighbors and take over its control over $\mathcal{V}$.

Attack 2: Sneak into MRN

- **ECR strategy:** Evaluate-Cut-Restore
  - Evaluate phase: larger out-degree ⇒ broader impact on others
    smaller in-degree ⇒ less affected by others

$$\max_{r_i} \quad (|\mathcal{N}_i^{out}| + \|W_{HF}^{[:,i]}\|_1 - |\mathcal{N}_i^{in}| - \|W_{HF}^{[i,:]}\|_1)$$
$$\text{s.t.} \quad i \in \mathcal{V}_H,\ \mathcal{N}_i^{out}| \geq 1, |\mathcal{N}_i^{in}| \leq \alpha_1,$$

  - Cut phase: break the connections between r_v and its in-neighbors

$$\max_{u_a^k} \ \alpha_2 \|\hat{z}_v^{k+1}(u_a^k) - \hat{z}_{v*}^{k+1}\|_2 + \alpha_3 \sum_{j \in \mathcal{N}_v^{in}} \|\hat{z}_j^{k+1} - \hat{z}_v^{k+1} - \tilde{h}_{jv}\|_2$$

If r_v is not easily to approach, attack $r_j \in \mathcal{N}_v^{in}$ first (indirect controllability)

  - Restore phase: make r_a recognized by the out-neighbors of r_v, then restore the formation shape

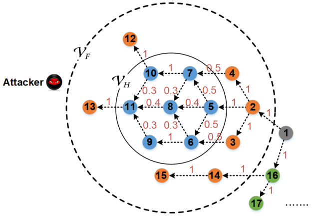
$$u_a^k = \arg\max_{u_a} \left\{ \left\| \hat{z}_v^{k+1}(u_a) - \hat{z}_{v*}^{k+1} \right\|_2 : z_a^{k+1} \in \mathcal{Z}_v^f \right\}.$$

▷ $\mathcal{Z}_v^f = \{z : \|z(t) - z_{j*}(t)\|_2 < \|z_j(t) - z_{j*}(t)\|_2, \forall j \in \mathcal{N}_i^{out}\}$

Performance Evaluation

▶ Simulation setting

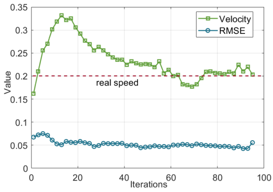
- MRN of 17 robots, two kinds of interaction structure



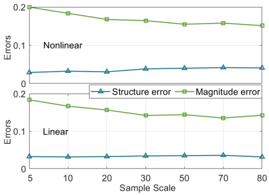
- $u_c = 0.2m/s$, $R_c = 7m$, $R_o = 2m$ and $R_s = 0.5m$
- Dynamic model
  - linear $\dot{z}(t) = -Lz(t) + Lh + u_0$ ▷ $u_0^N = u_c$
  - nonlinear $\dot{z}(t) = -Lz(t) + Lh + u_s(t)$ ▷ $\lim_{t\to\infty} u_s^N(t) = u_c$
- Metric of evaluation: structure (ε_1) and magnitude (ε_2) error

$$\varepsilon_1 = \frac{\|\text{sign}(\hat{W}_{HF}) - \text{sign}(W_{HF})\|_0}{|\mathcal{H}||\mathcal{F}|}, \quad \varepsilon_2 = \frac{\|\hat{W}_{HF} - W_{HF}\|_F}{\|W_{HF}\|_F}$$

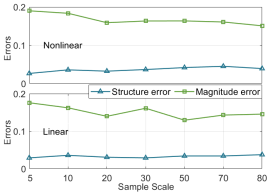
▶ **Stage 1:** identity the steady pattern



(c) velocity estimation



(d) approximation errors of structure 1



(e) approximation errors of structure 2

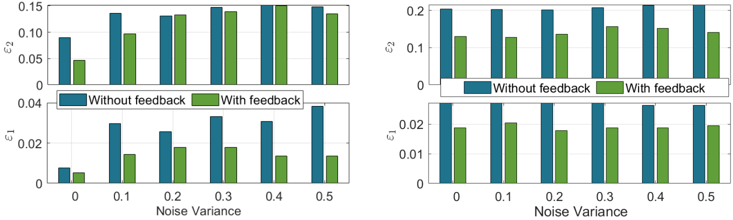
Figure 2 Results evaluation of Stage 1

- the velocity estimation remains stable when $\mathcal{V}$ reaches steady state
- accuracy of convergence time k^* mainly affects ε_2
- as the sample scale grow, ε_1 and ε_2 become stable

Performance Evaluation

▶ **Stage 2:** infer the internal interaction structure

Note feedback means using estimation of R_c as a constraint to infer W_{HF}



(a) under linear model

(b) under nonlinear model

Figure 3 The approximation result comparison of $\hat{W}_{HF}$

- ε_1 is small and generally stable under different noise
- the errors decrease significantly if feedback is adopted
- linear approximation works well in two situations in terms of ε_1

Performance Evaluation

- **Stage 3:** infer the external interaction mechanism

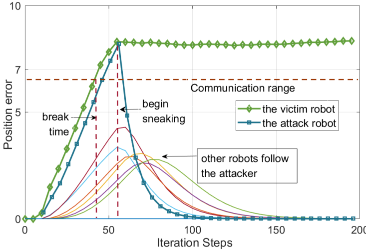
Table 2 Statistic results of obstacle-avoidance mechanism regression

	25 samples			50 samples		
Index	MDA	RMSE	MAE	MDA	RMSE	MAE
Training	0.880	0.253	0.154	0.913	0.217	0.113
Testing	0.933	0.601	0.404	0.933	0.581	0.300
	100 samples			200 samples		
Index	MDA	RMSE	MAE	MDA	RMSE	MAE
Training	0.910	0.333	0.146	0.923	0.426	0.206
Testing	0.956	0.541	0.291	0.967	0.496	0.264

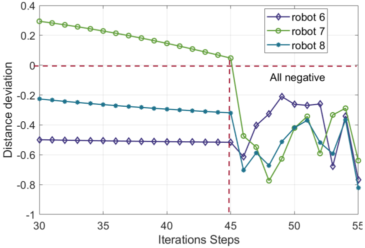
- $\mathrm{MDA} = \frac{1}{m}\sum_{i=1}^{m} \text{sign}(y_i - {y'}_i)$, $\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - {y'}_i)^2}$, $\mathrm{MAE} = \sum_{i=1}^{m}\frac{|y_i - {y'}_i|}{m}$
- more samples brings more accurate results but not significant improvement

Performance Evaluation

- **Stage 4:** ECR attack strategy



(a) The position errors between the real and the desired positions, and r_a takes the z_5^* as its desired position.



(b) The distance deviations $(\|z_a - \tilde{z}_j\|_2 - \|\tilde{z}_i - \tilde{z}_j\|_2)$, here $i = 5$, $j = 6, 7, 8$.

Figure 4 ECR strategy.

- r_v is gradually pulled out of the interaction range of its in-neighbors
  - break point: connection between $\mathcal{N}_v^{in}$ and r_v break
  - sneak point: r_a is recognized by $\mathcal{N}_v^{out}$
- the indirect controllability is verified

Conclusions

- Conclusions
  - reveal the learnability of the interaction rules in MRNs
  - design entrap-robot and sneak-into-MRN attack strategies
  - prove the conditions to launch the attacks
  - obtain performance bounds of the proposed attacks
- Open problems
  - explore advanced attacks with lower cost and higher rewards
  - design efficient detection methods to identify the potential threats
  - secure the interaction by leaking confusing states

Thank You !

Q&A