INSTITUTE VLATACOM

Adaptive Communication in Decentralized Multi-Agent Reinforcement Learning

Nemanja Ilić, Miljan Vučetić

Vlatacom Institute

Outline

- About Vlatacom Institute
- Interdisciplinary research activities at the Institute
- Overview of one research direction spanning multiple projects
  - Gentle introduction
  - Highlights of three related research projects
    - Automated penetration testing in cybersecurity
    - Distributed target tracking in visual surveillance systems
    - Intelligent traffic control in urban mobility networks

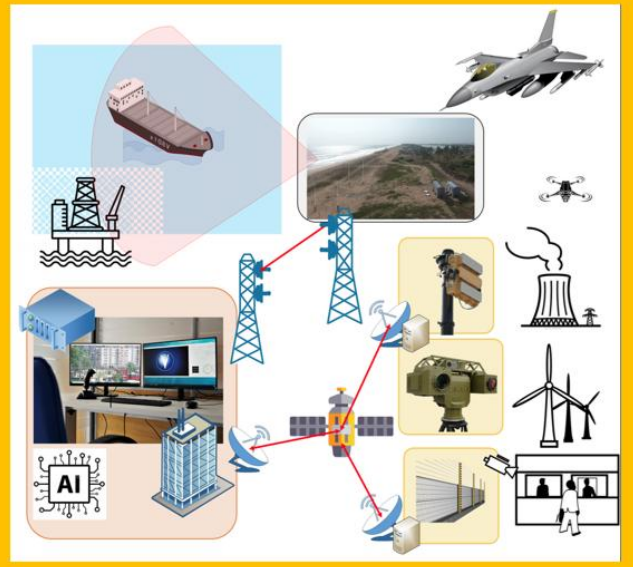
About Vlatacom

- 1997 Company founded in Belgrade, Serbia
- 2002 Infrastructure for eDocuments in Serbia
- 2004 First hardware product Vlatacom Document Reader – VDR
- 2005 Intellectual property sold to Motorola
- 2008 First Projects abroad
- 2010 Got license for production & trade in military equipment
- 2011 Accredited as R&D Center
- 2015 Accredited as R&D Institute (reaccredited 2019, 2023)
- 2023 170 employees among which
  - 30 PhD
  - 33 at PhD studies
  - More than 100 MSc and BSc
- 2023 Established UAE office Vlatacom Technology, Abu Dhabi
- More than 99% income from abroad in last 10 years

Vlatacom expertise

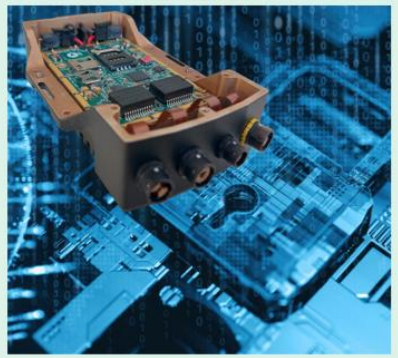
Electro-optics

Maritime security



System integration

Vlatacom Institute is focused on development of cutting-edge technologies and products



Crypto

Military

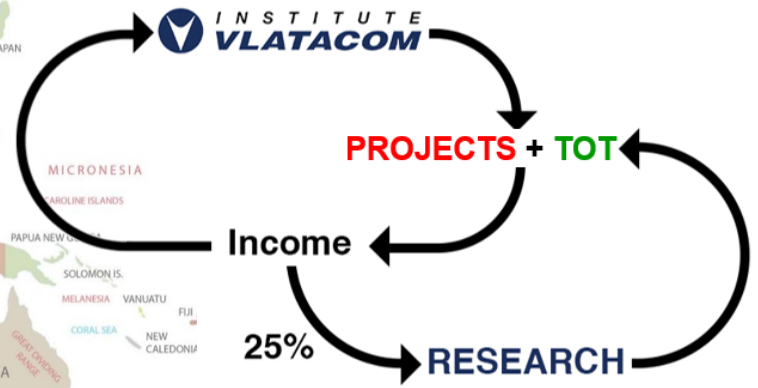
Worldwide activities, references & research

Africa:
- Algeria
- Angola
- Botswana
- Burkina Faso
- Central African Republic
- Democratic Republic Congo
- Egypt
- Equatorial Guinea
- Eritrea
- Ethiopia
- Gabon
- Gambia
- Ghana
- Guinea – Bissau
- Ivory Coast
- Kenya
- Malawi
- Mauritius
- Namibia
- Niger
- Nigeria
- Senegal
- Sierra Leone
- South Africa
- South Sudan
- Sudan
- Uganda
- Zanzibar

Middle East & Asia:
- Bangladesh
- China
- Iraq
- Israel
- Japan
- Jordan
- Lebanon
- Maldives
- Mongolia
- Philippines
- UAE
- Uzbekistan

South America:
- Barbados
- Grenada
- St Kitts and Nevis
- Suriname
- Venezuela

Europe:
- Moldova



Scientific results

Scientific result	Quantity
Monographs	7
International journal papers	34
National journal papers	20
International conferences	129
Technical solutions	36
Doctoral dissertations	8
Total	**234**

Title	Number
University professors	14*
Senior research associate	6
Research associate	7
Research assistant	7
Junior research assistant	14
Principal technical associate	8
Senior technical associate	21
Technical associate	13

*Note: some researchers have both University and Institute title

- Total 234 scientific result achieved in 2019 – 2024
- According to Serbian Ministry of Education, Science and Technological development 76 employees have Institute title (scientific or technical)
- Additionally, there are more than 120 intellectual property items: new algorithms, application software, hardware solutions, new measurement methodologies, original work flow procedures, etc.

Topic ∈ Multi-domain research landscape

- Artificial Intelligence (AI)
  - Multi-agent systems, autonomous decision-making
- Machine Learning (ML)
  - Reinforcement learning, representation learning, learning under partial observability
- Multi-Agent Reinforcement Learning (MARL)
  - Coordination, exploration, decentralized policy learning
- Decentralized / Distributed Systems
  - Local communication, networked decision-making, distributed computation
- Communication & Networking
  - Adaptive message passing, bandwidth constraints, communication protocols for agents
- Control Theory & Distributed Control
  - Consensus, distributed optimization, stability under communication constraints
- Graph / Network Science
  - Agents connected via dynamic or static communication graphs
- Robotics / Cyber-Physical Systems
  - Swarm robotics, sensor networks, autonomous vehicles
- Application domains of Cybersecurity, Visual Surveillance Systems, Intelligent Transportation Systems

Title breakdown

- **Adaptive Communication**
  - Why adapt communication?
    - Limited bandwidth → cannot share all information all the time
    - Dynamic environments → neighbors' relevance changes over time
    - Scalability → reduces unnecessary messages in large networks
    - Robustness → avoids congestion, ensures critical info is prioritized
- **Decentralized**
- **Multi-Agent**
  - Many real-world problems involve multiple decision-makers
  - Enables cooperation and coordination to achieve global objectives
  - Handles complex, large-scale environments (traffic networks, sensor grids, swarms)
- **Reinforcement Learning**

Decentralized

- Why decentralization?
  - No central coordinator → robustness to failure
  - Scales to large networks (sensors, agents, intersections…)
  - Reduces communication overhead
  - Local decisions → faster real-time responses
  - Supports privacy and autonomy of individual nodes
- How decentralization works
  - Each node/agent uses local observations
  - Exchanges information only with neighbors
  - Iterative updates (e.g., consensus, gossip, message passing)
  - Global objectives emerge from local cooperation
  - Can be applied in: traffic control, IoT, swarm robotics, multi-agent RL

Reinforcement learning

- One of the three main branches of machine learning
  - Alongside supervised and unsupervised learning
- Agent interacts with an environment over time
- Learns to choose actions that yield higher rewards
- Uses observations to improve its understanding and behavior

Project 1

ELSEVIER

Artificial Intelligence

Volume 326, January 2024, 104032

Distributed web hacking by adaptive consensus-based reinforcement learning

Nemanja Ilić [a b], Dejan Dašić [a c], Miljan Vučetić [a c], Aleksej Makarov [a], Ranko Petrović [a]

Show more

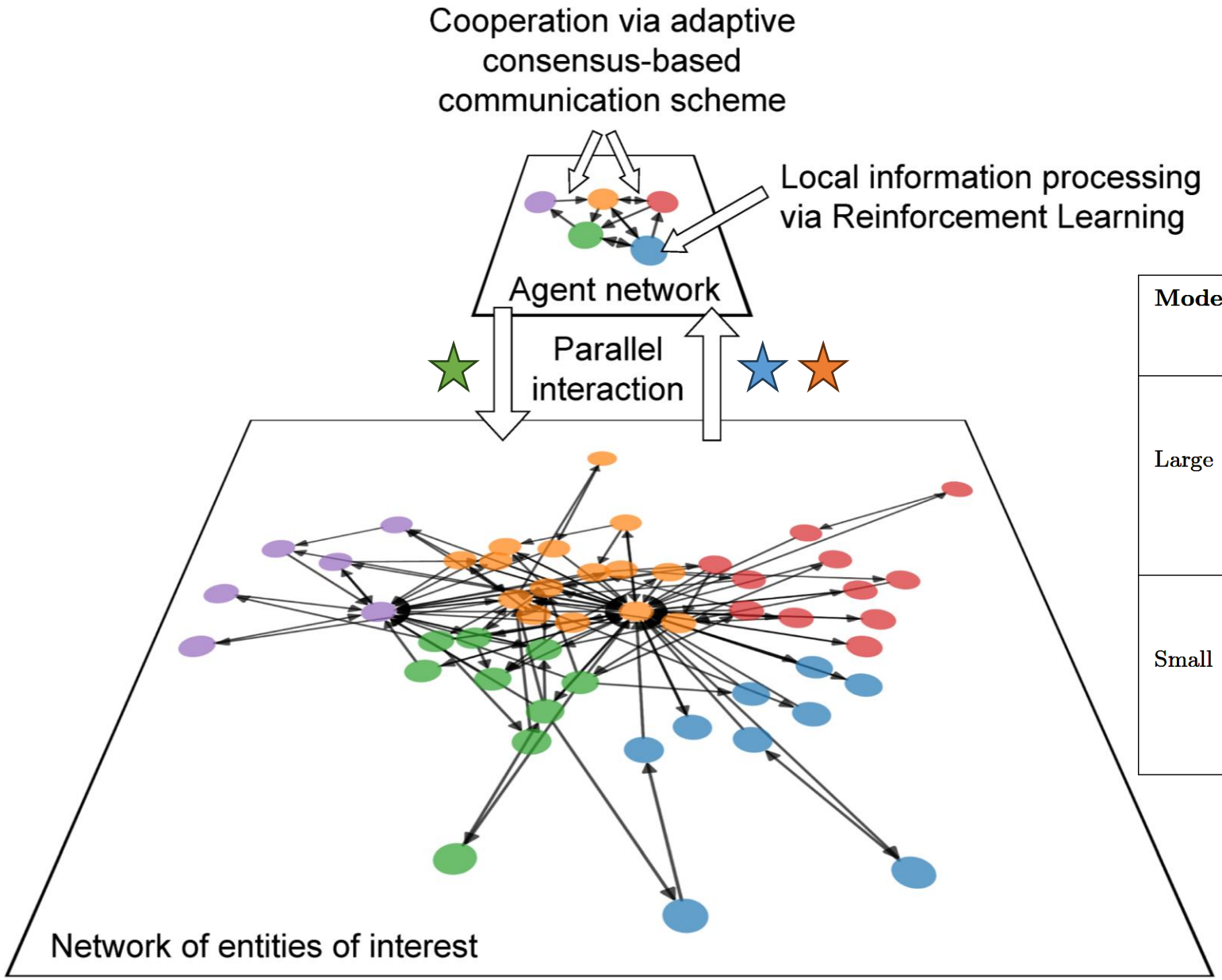
+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.artint.2023.104032

Get rights and content

Automated penetration testing

- More than 80% of all malware attacks are performed by autonomous bots - the focus is on the use of such bots, i.e., intelligent agents, in ethical hacking
- Agent interaction with the system is modeled via the Reinforcement Learning (RL) framework
- Time plays vital role in hacking tasks ⇒ dispatch multiple cooperative agents ⇒ this would hopefully allow for faster solutions
- The agents interact with multiple copies of the environment in parallel, trying to reach the target by:
  1. processing local information (learning from interaction experience - focused on agent policy optimization)
  2. communicating with each other (aimed at providing all agents with viable network-wide information)



Model	Action	Observation	Reward
Large	*Read (node i)*	List of linked nodes	-1
	Search (node i)	False (flag not in node i)	-1
		True (flag in node i) *Ends episode	100
	None (node i)	None	-1
Small	*Read (current node)*	List of linked nodes	-1
	Search (current node)	False (flag not in node i)	-1
		True (flag in node i) *Ends episode	100
	Switch (to node i)	None	-1

Context

Aspect	**Related literature**	**Our approach**
how multiple agents operate in parallel towards common goal	agents contribute to and utilize global shared learning results	agents are empowered with global learning results in a completely decentralized manner
how are learning results of agents that possess relevant information prioritized	through a global prioritized entity	through the design of the inter-agent communication scheme
how is communication protocol applied to the agents' local processing results	as a part of the RL formalism or modeled within the Markov games framework	in an ad-hoc manner

Communication scheme

- Focus is on consensus algorithms - they have demonstrated effectiveness in facilitating efficient coordination among multiple agents
- Each agent is assumed to communicate with only a subset of other agents (neighbors) - by appropriately designing these communications, the desired global results can be obtained after multiple propagations

Local processing

- The existing web agent model is extended to the distributed RL context and a modified version is proposed that enables scalability to large networks
- The considered problem of web hacking is specified by the “Capture the Flag” (CTF) formulation (flag represents an abstraction of the relevant information)
  - The task is to design a web agent that would successfully search a network (an abstraction of a simple website) composed of a set of nodes (web pages) and a set of links (web links) for a node that contains the flag

Convergence speed optimization

- Bounds for the parameter underlying the design of the consensus communication matrices which ensure convergence are obtained
- The optimal parameter which achieves the fastest convergence is also obtained

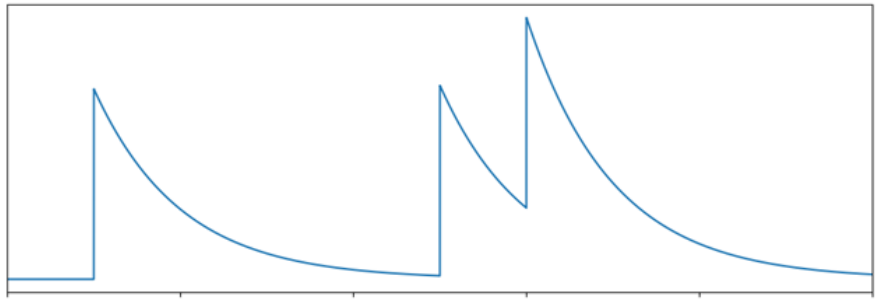
Local processing

- The agents have a common task of finding the optimal policy for solving the set CTF problem and use the Q-learning algorithm:

$$Q'_{i,t}(s,a) = Q_{i,t}(s,a) + \alpha \left(R_{i,t+1} + \gamma \max_{a} Q_{i,t}(S_{i,t+1}, a) - Q_{i,t}(S_{i,t}, A_{i,t}) \right) \cdot \mathbf{1}_{(s,a)=(S_{i,t},A_{i,t})}$$

- The idea - enable agents help each other by exchanging information & incorporate some merit of "quality" of the exchanged information based on

$$E'_{i,t}(s,a) = \gamma\lambda E_{i,t}(s,a) + \mathbf{1}_{(s,a)=(S_{i,t},A_{i,t})}$$



Communication scheme

- Using the above, the goal is to design such a communication strategy that would asymptotically (w.r.t. the number of consensus steps L) give:

$$\lim_{L\to\infty} \bar{Q}_{i,t}^{[L]}(s,a) = \frac{\sum_{j=1}^{M} E'_{j,t}(s,a) Q'_{j,t}(s,a)}{\sum_{j=1}^{M} E'_{j,t}(s,a)}$$

- We propose an algorithm with the following network-wide form (C represents the so-called consensus matrix):

$$\bar{Q}_t^{[L]} = \frac{(C^L \otimes I_M) \cdot (E'_t \odot Q'_t)}{(C^L \otimes I_M) \cdot E'_t}$$

The algorithm

Algorithm Adaptive consensus-based distributed Q-learning algorithm
(at time t, for agent i and state/action pair (s, a))

Input: $Q_{i,t}(s,a)$, $E_{i,t}(s,a)$

Output: $Q_{i,t+1}(s,a)$, $E_{i,t+1}(s,a)$

Update $Q'_{i,t}(s,a)$ and $E'_{i,t}(s,a)$ using (1) and (2)

$\Gamma_{i,t}^{[0]}(s,a) = E'_{i,t}(s,a)Q'_{i,t}(s,a)$

$\Sigma_{i,t}^{[0]}(s,a) = E'_{i,t}(s,a)$

for $l = 1$ to L **do**

Send $\Gamma_{i,t}^{[l-1]}(s,a)$, $\Sigma_{i,t}^{[l-1]}(s,a)$ to all out-neighbors

Receive $\Gamma_{j,t}^{[l-1]}(s,a)$, $\Sigma_{j,t}^{[l-1]}(s,a)$ from all in-neighbors

$\Gamma_{i,t}^{[l]}(s,a) = \sum_{j\in\mathcal{J}_i} C(i,j)\Gamma_{j,t}^{[l-1]}(s,a)$

$\Sigma_{i,t}^{[l]}(s,a) = \sum_{j\in\mathcal{J}_i} C(i,j)\Sigma_{j,t}^{[l-1]}(s,a)$

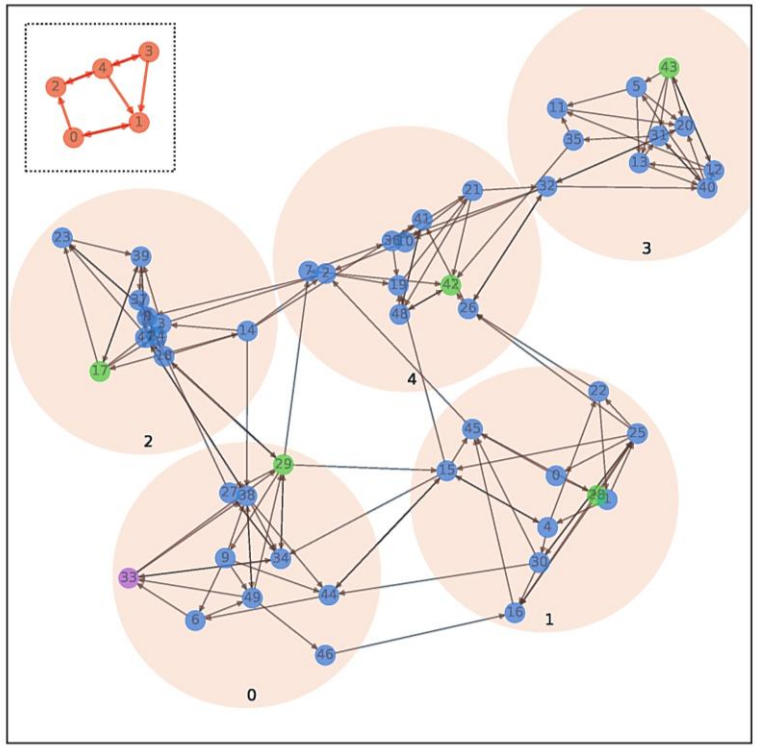
end for

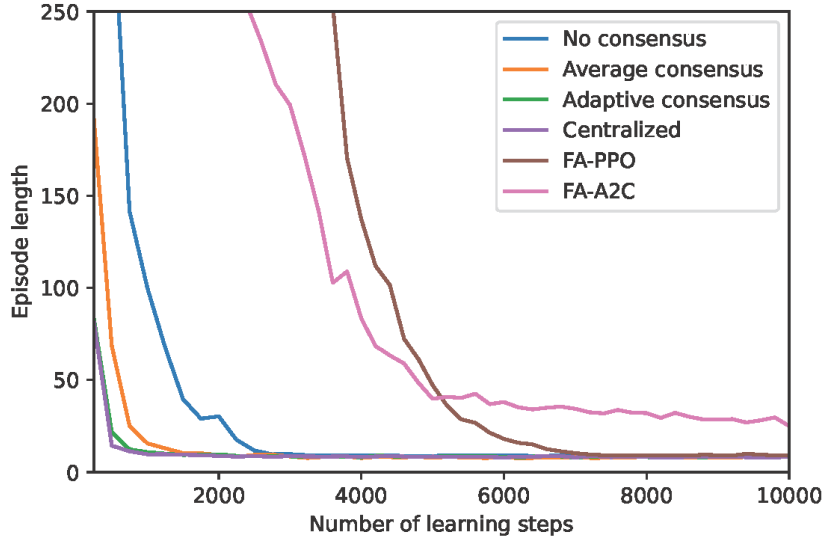
$\bar{Q}_{i,t}^{[L]}(s,a) = \Gamma_{i,t}^{[L]}(s,a)/\Sigma_{i,t}^{[L]}(s,a)$

Predict $Q_{i,t+1}(s,a)$ and $E_{i,t+1}(s,a)$

Experimental results

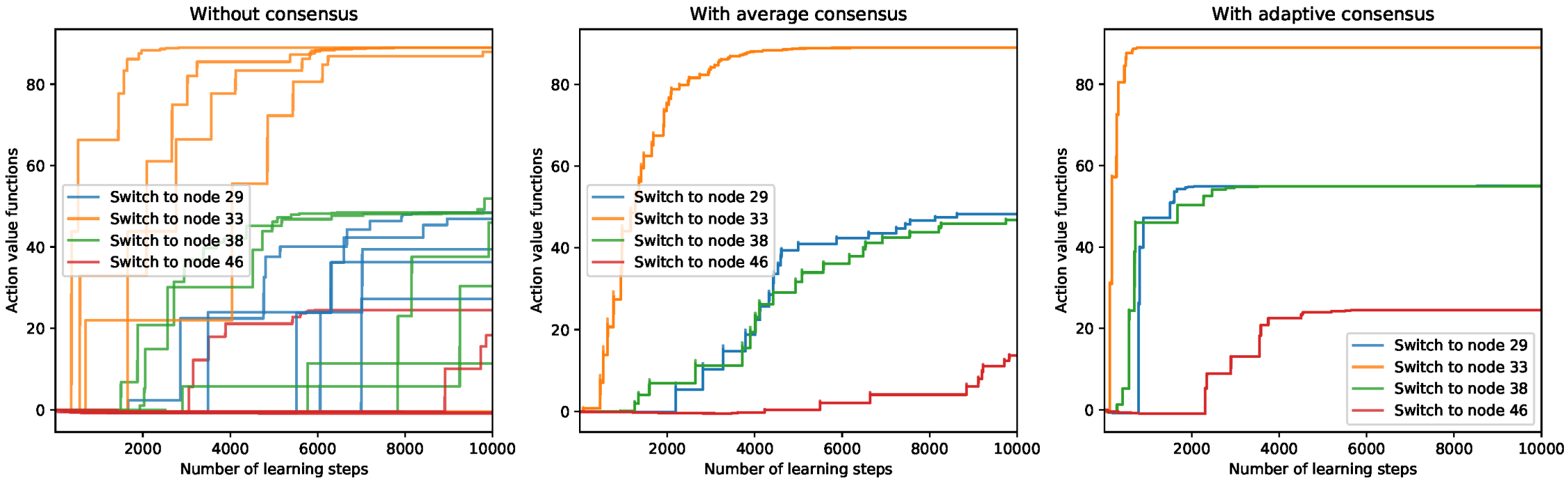
- The proposed adaptive scheme significantly accelerates the learning process compared to the average consensus (very close to the centralized scheme)
- The function approximation algorithms (A2C and PPO) require more learning steps to approach optimal solutions (less sample efficient)



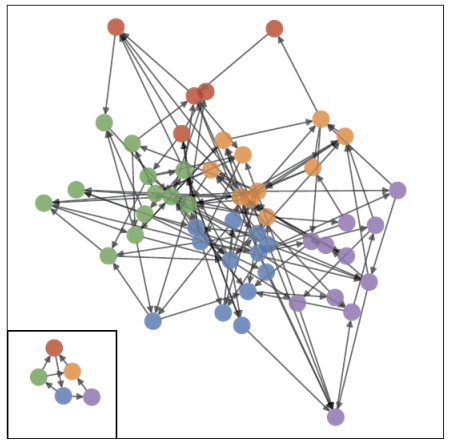


Experimental results

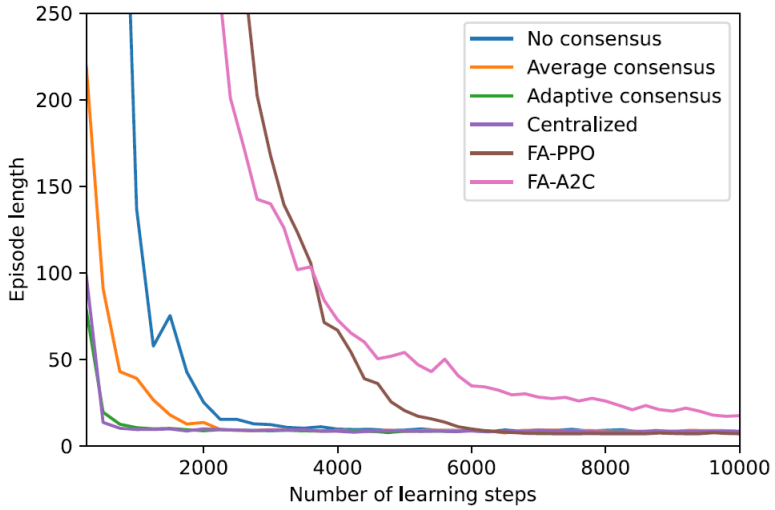
- The obtained network-wide results are close to those of the “fastest” individual agents in the non-cooperative case

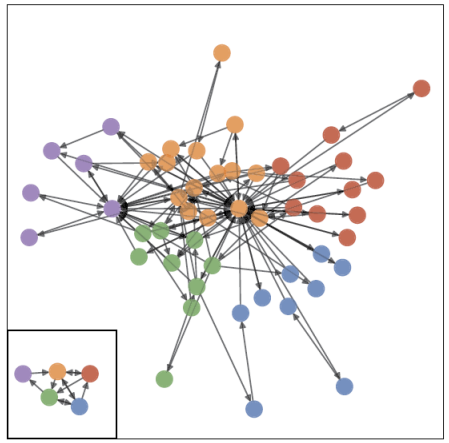


Experimental results

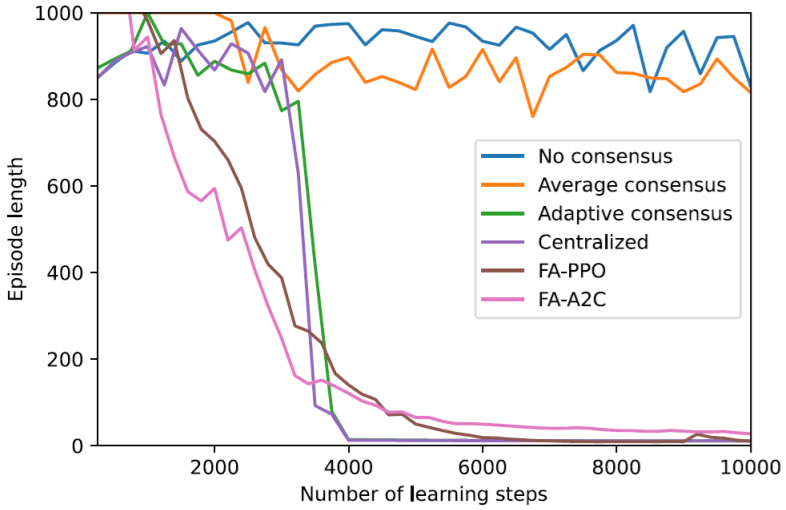


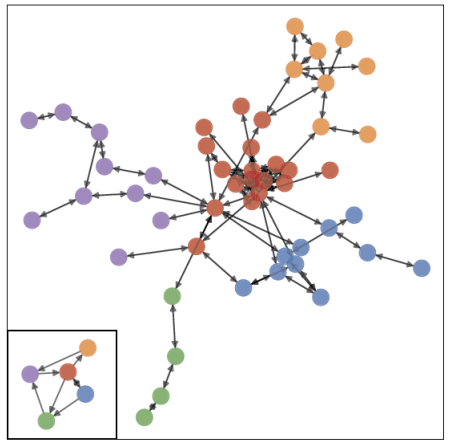
Erdős–Rényi graph



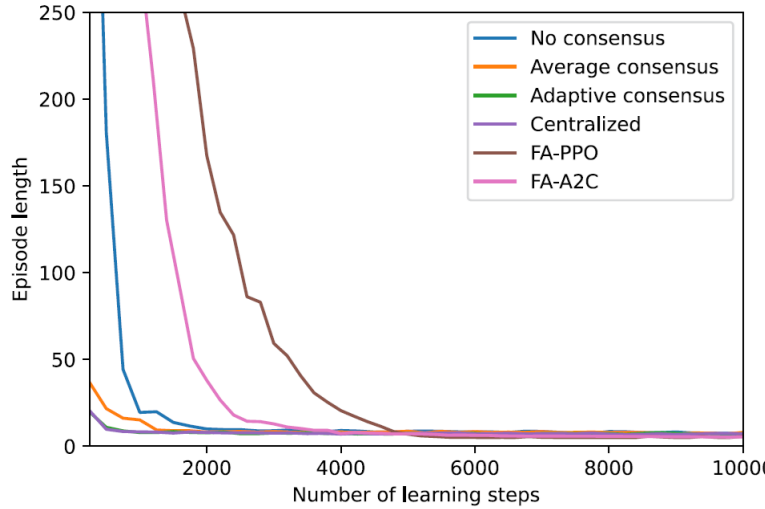


Scale-free graph





Wiskonsin web graph



Project 2

Adaptive Asynchronous Gossip Algorithms for Consensus in Heterogeneous Sensor Networks

Nemanja Ilić ; Miljan Vučetić ; Aleksej Makarov ; Ranko Petrović ; Marija Punt

Published in: IEEE Internet of Things Journal (Volume: 12 , Issue: 13, 01 July 2025)

Page(s): 25516 - 25532

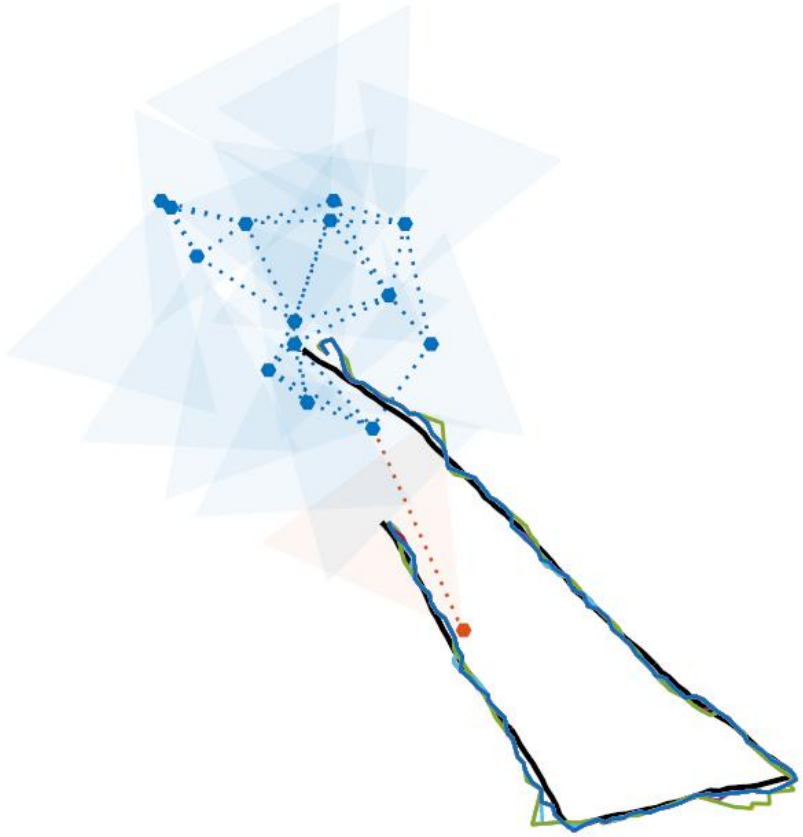
DOI: 10.1109/JIOT.2025.3559242

Date of Publication: 10 April 2025

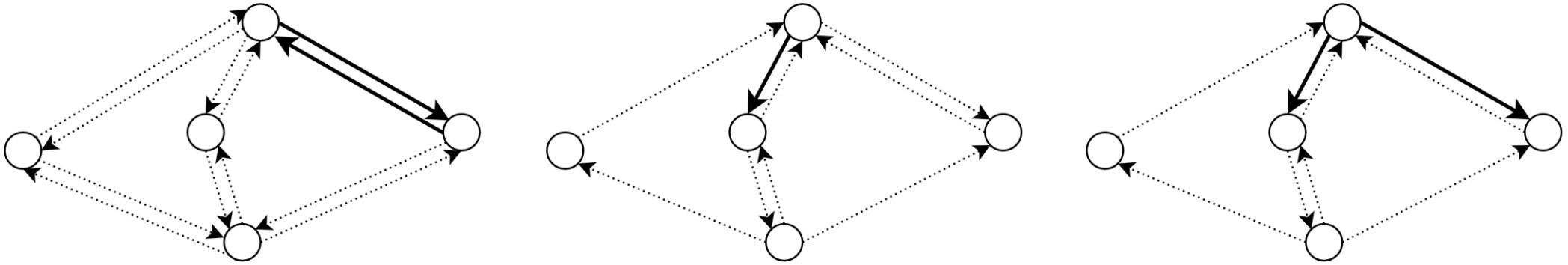
Publisher: IEEE

Context

- IoT and sensor systems increasingly rely on heterogeneous networks
- Target tracking using static cameras and drones is challenging
  - Drone mobility introduces sparse information sources and a dynamic communication graph, complicating global agreement
- Asynchronous communication is better suited for such settings
- Existing asynchronous gossip schemes are inefficient for disseminating sparse information; adaptive methods are needed to accelerate agreement with minimal communication



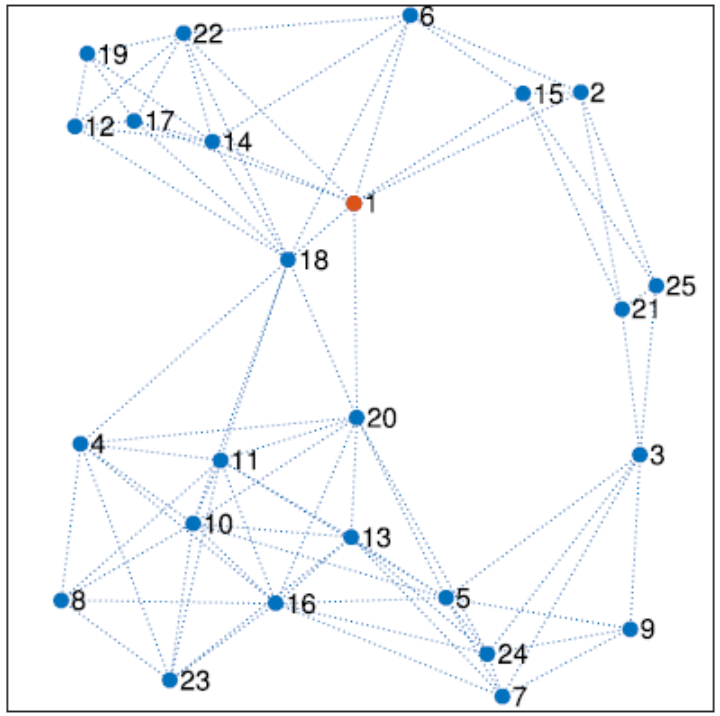
Two-way, one-way & broadcast gossip



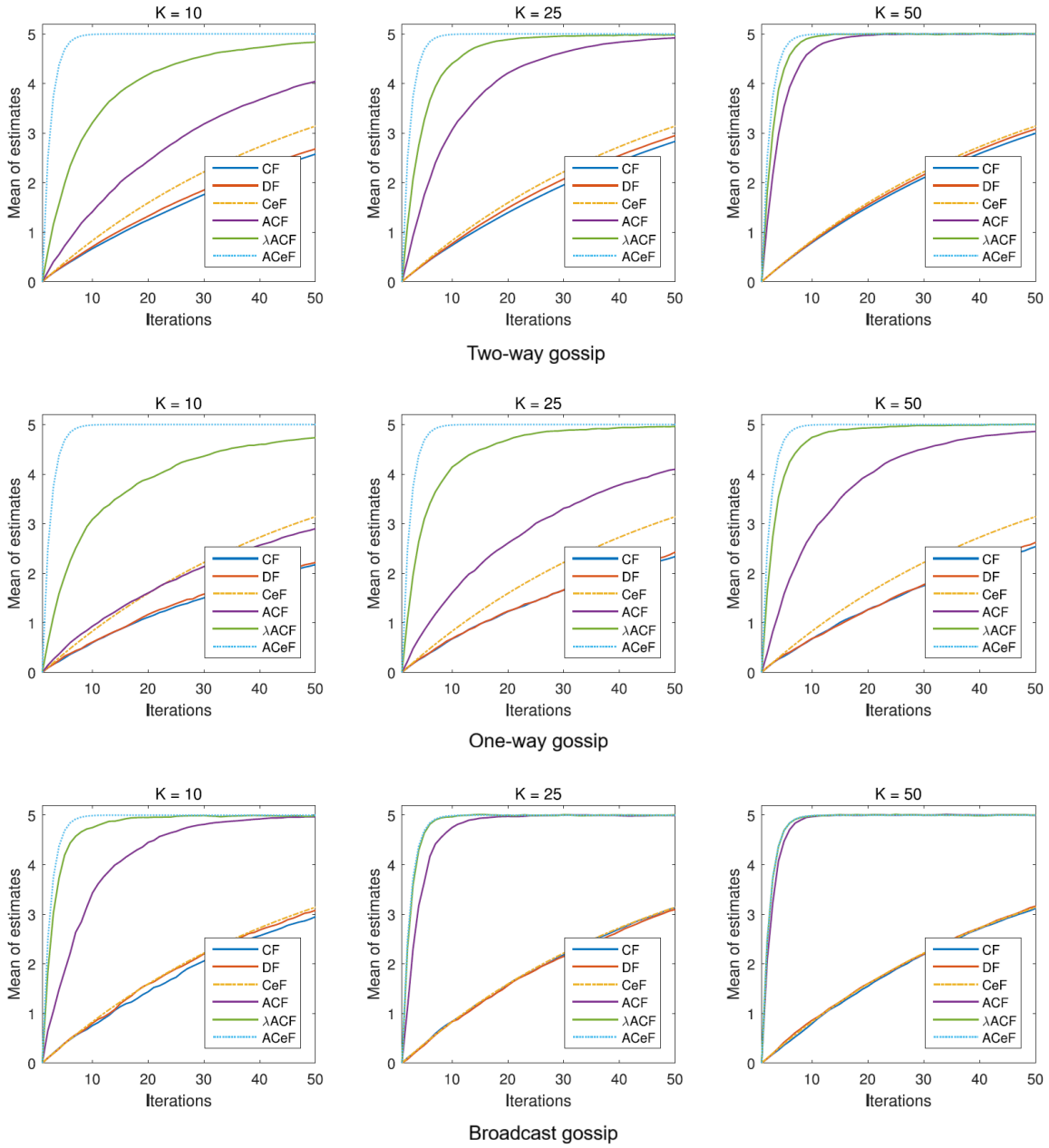
- Without adaptation
  - All nodes – same rate
- With adaptation
  - Dynamic rate

Results

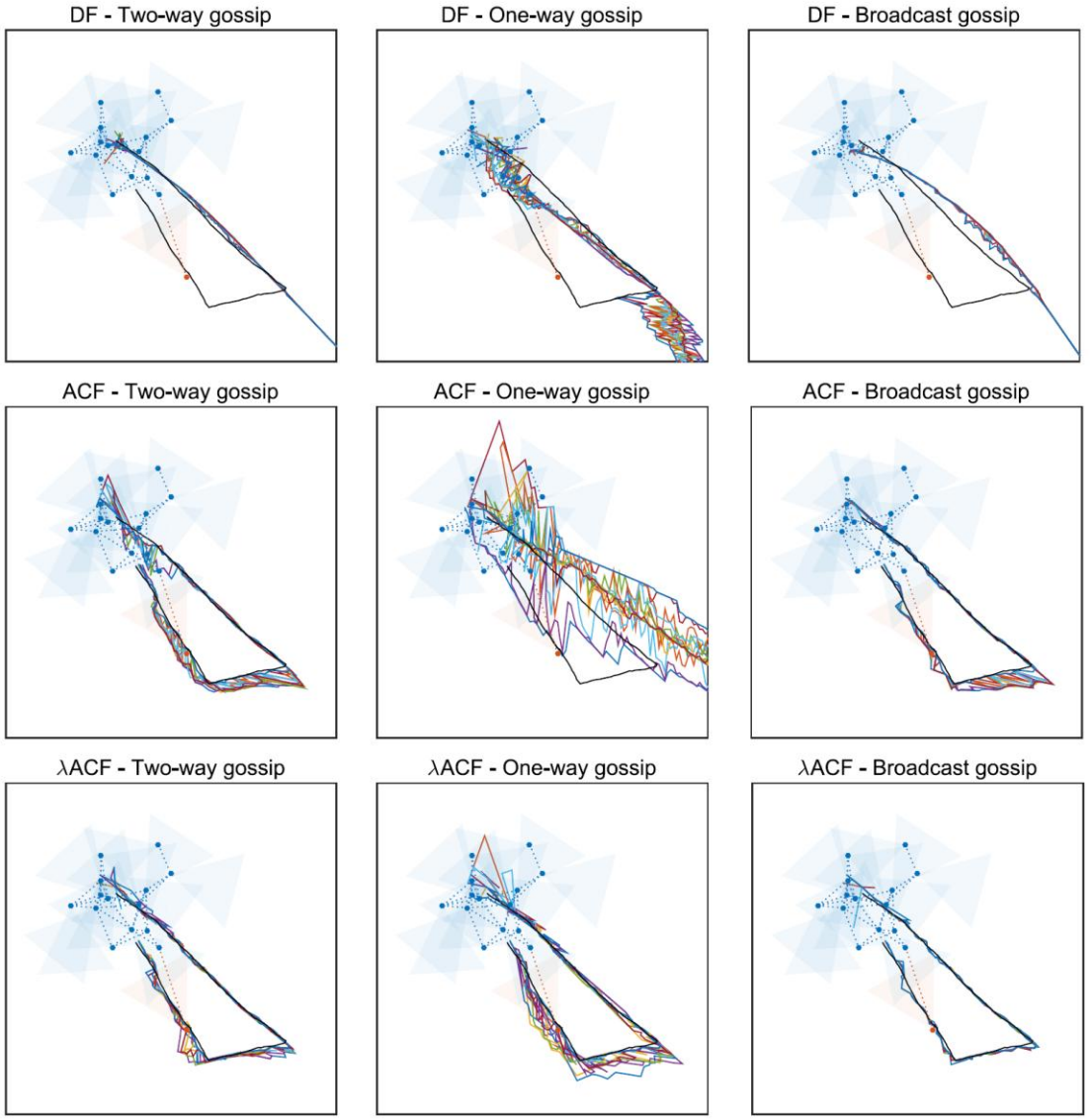
- Distributed parameter estimation



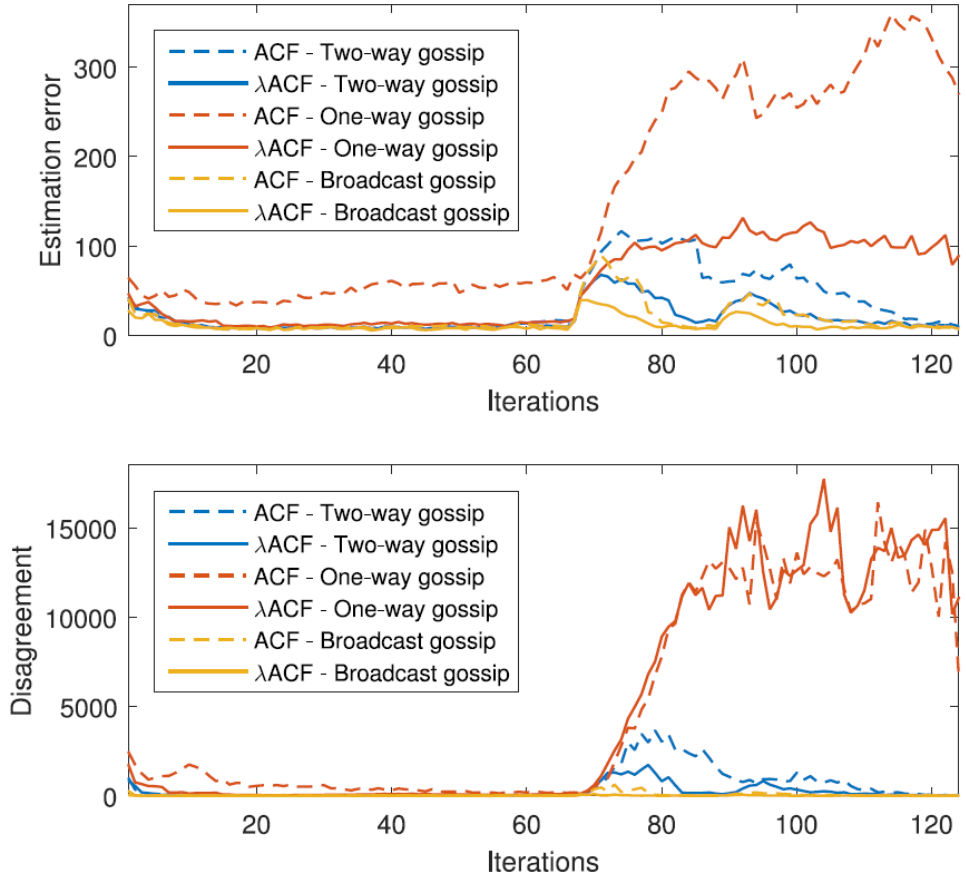
The network & Performance metrics



Results – target tracking



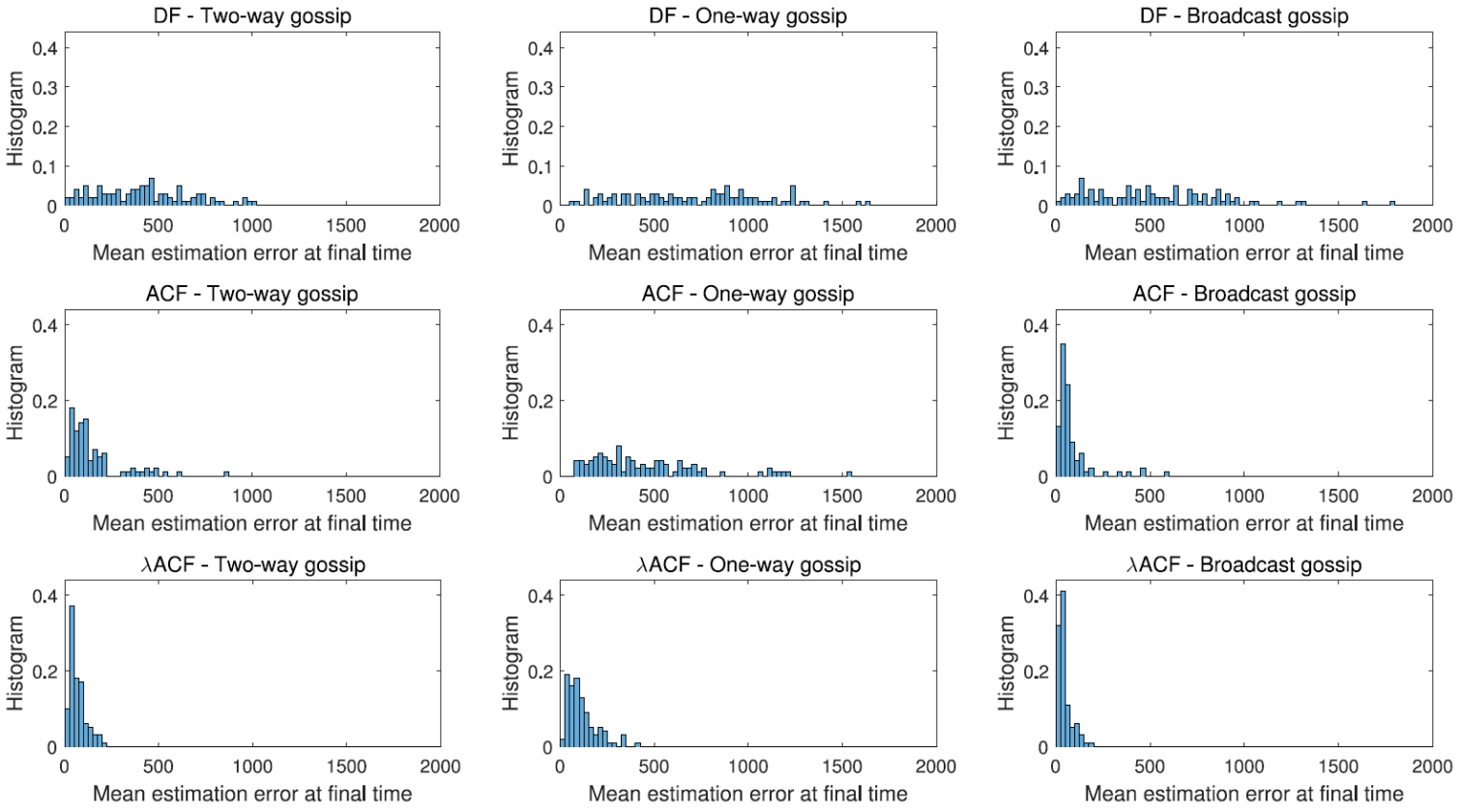
Representative simulation runs



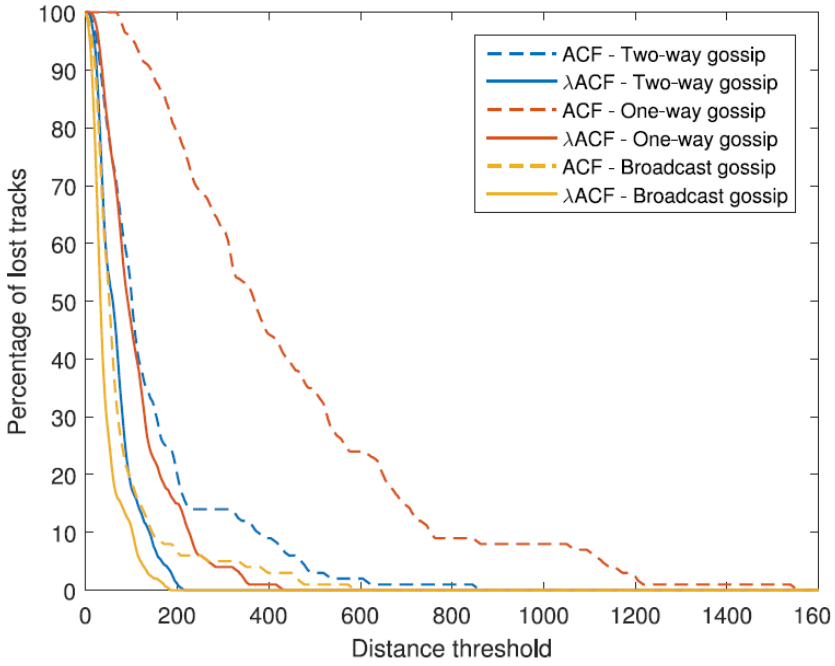
Average performance metrics

INSTITUTE VLATACOM

Results – target tracking (2)



Performance indicators



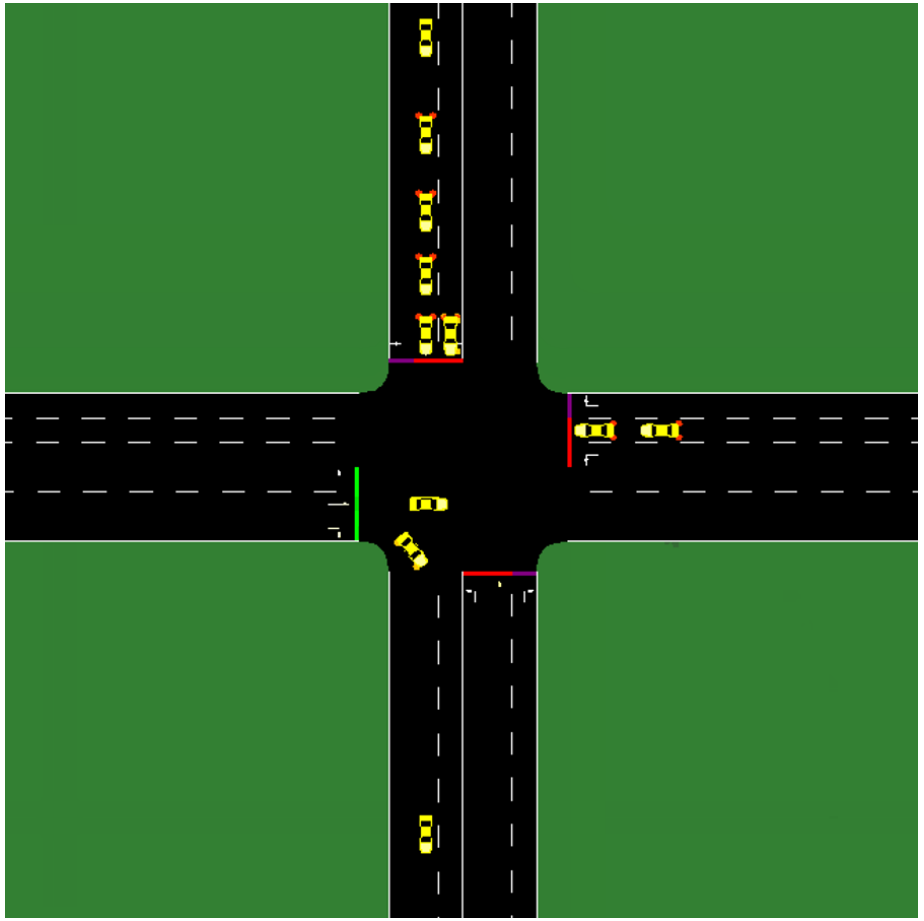
Average performance metrics

Project 3 – Intelligent traffic control

- Traffic networks are dynamic and hard to model explicitly
- Reinforcement Learning (RL) is often used
- RL learns adaptive signal timing from interaction with the environment
- Agents = intersections; actions = signal phases
- Goal: reduce delay, queues, stops, emissions

Environments

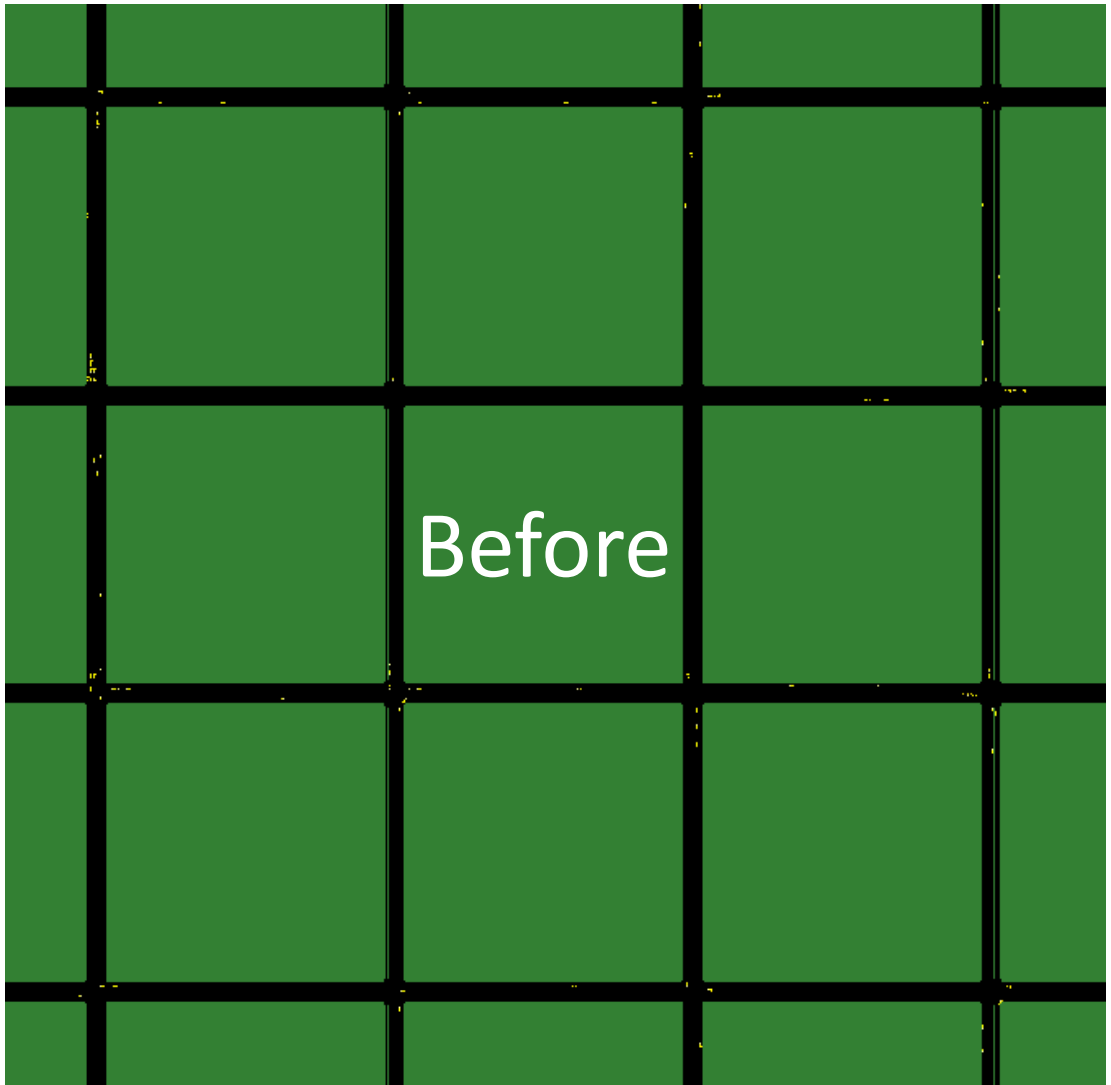
- Microscopic simulators
  - SUMO, CityFlow — detailed vehicle dynamics, realistic flow
- Mesoscopic/macroscopic simulators
  - MATSim, Aimsun — scalable for large networks
- Gym-like wrappers
  - SUMO-RL — standard RL API for quick experimentation
- Network variability
  - grids, corridors, real-city maps
- Control tasks
  - fixed-time optimization, adaptive signal control, corridor coordination



Multi-Agent RL in Traffic

- Each intersection is a learning agent
- Coordination matters to avoid local optima
- Rewards often based on waiting time or queue length
- Current trends:
  - Lightweight communication (consensus/gossip) for scalability
  - CTDE (e.g., QMIX/VDN): centralized training, decentralized execution
    - Global state used only during training

Results



Results

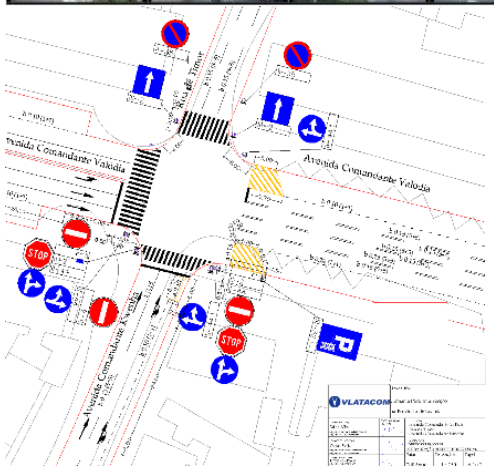
```
============================================
Running Random Agent
============================================
Agents: 16 | Actions: 8 | Max Steps: 360
============================================

  Step |    Wait | Stop | Speed
--------------------------------------------
     0 |       0 |    1 |   8.3
    30 |     651 |   19 |   6.1
    60 |     651 |   26 |   6.4
    90 |     670 |   29 |   6.8
   120 |    1480 |   41 |   6.1
   150 |    1731 |   44 |   6.6
   180 |    2994 |   87 |   5.9
   210 |    2218 |   82 |   7.0
   240 |    3451 |   64 |   5.7
   270 |    1449 |   38 |   7.2
   300 |     581 |   21 |   8.1
   330 |    1553 |   30 |   4.1

============================================
```

```
============================================
Running QMIX Agent
============================================
Agents: 16 | Actions: 8 | Max Steps: 360
============================================

  Step |    Wait | Stop | Speed
--------------------------------------------
     0 |       0 |    1 |   7.2
    30 |      67 |    6 |   9.1
    60 |      79 |   10 |   8.9
    90 |      51 |    7 |   9.5
   120 |     127 |   11 |   9.3
   150 |     114 |   12 |   8.9
   180 |     299 |   19 |   9.8
   210 |     384 |   31 |   8.5
   240 |      35 |    9 |  10.0
   270 |     132 |   15 |   8.9
   300 |     117 |    7 |  10.7
   330 |     144 |    6 |   9.1

============================================
```

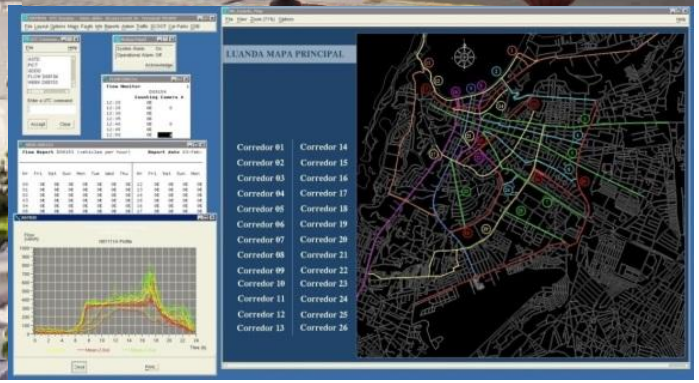
Performance metrics

```
================================================================================
COMPARISON RESULTS
================================================================================

Metric              |      Random |     Trained |  Improvement
--------------------|-------------|-------------|-------------
Avg Wait Time (s)   |     1435.01 |      132.28 | ↓       90.8%
Avg Stopped         |       41.44 |       11.70 | ↓       71.8%
Avg Speed (m/s)     |        6.55 |        9.52 | ↑       45.5%
================================================================================
```

Safe City Luanda, Tashkent



- Traffic monitoring and control projects have evolved to smart city projects with extensive application of AI systems



Q & A

Nemanja Ilić
Miljan Vučetić
VLATACOM INSTITUTE d.o.o.
5 Milutina Milankovića Blvd.
11070 Belgrade
Serbia

www.vlatacom.com
nemanja.ilic@vlatacom.com
miljan.vucetic@vlatacom.com

Phone: +381 [0] 11 3771100