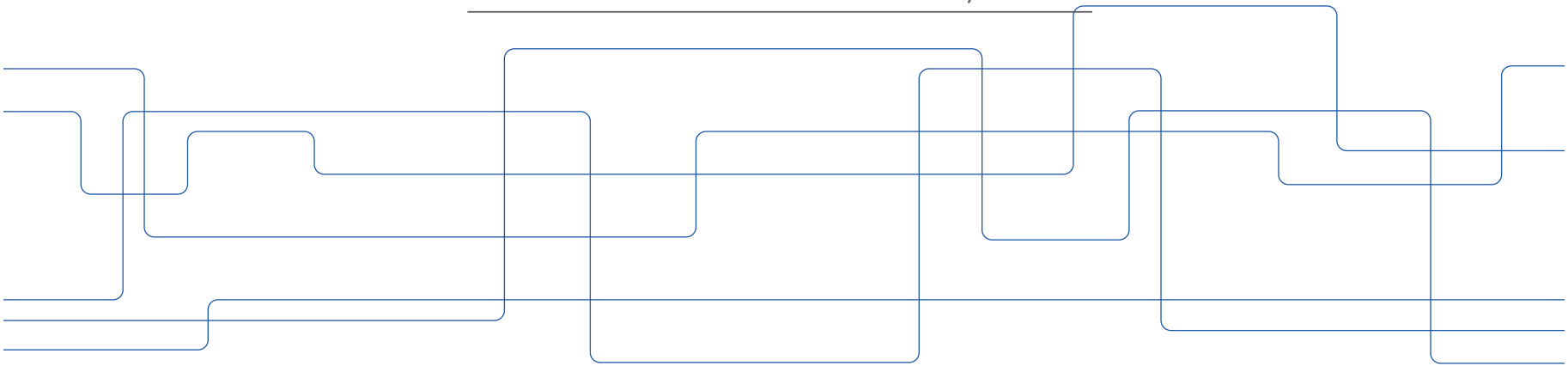
KTH ROYAL INSTITUTE OF TECHNOLOGY

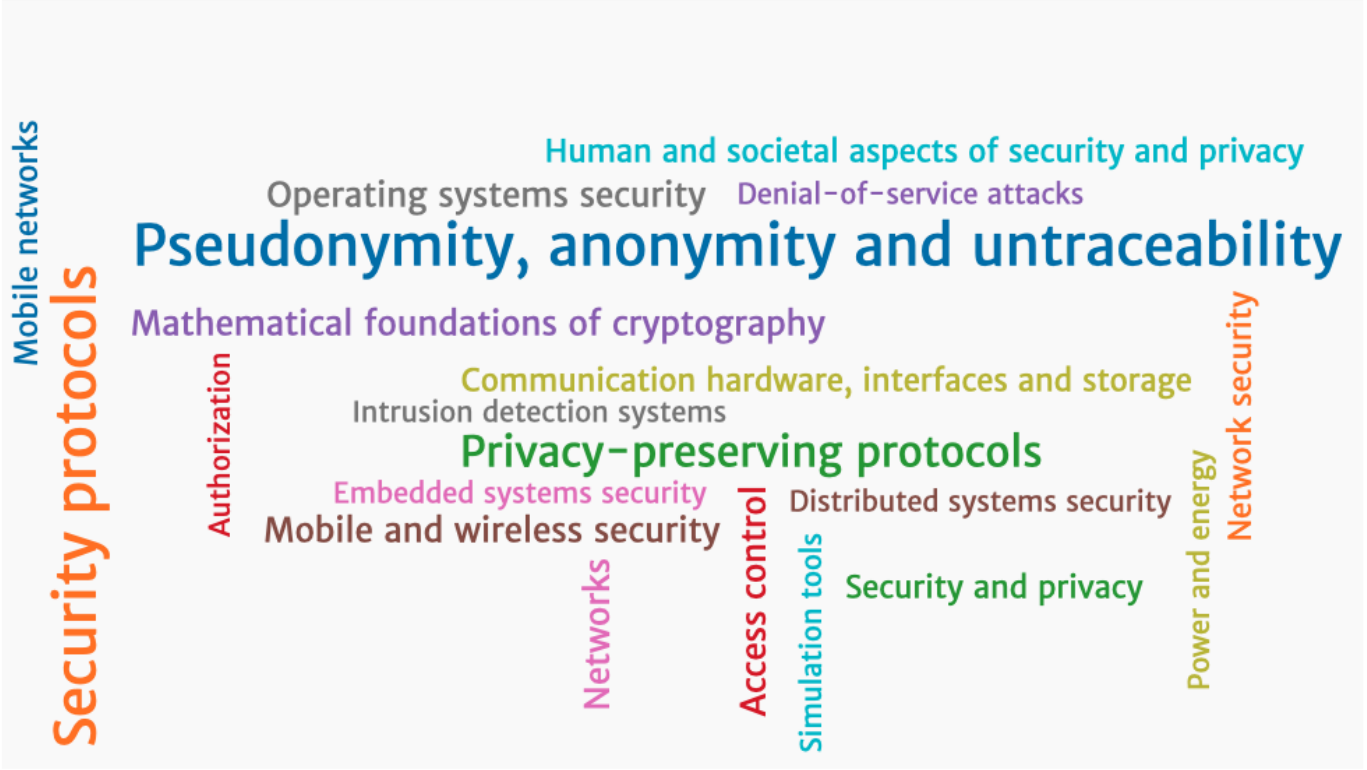
Networked Systems Security

Panos Papadimitratos

www.eecs.kth.se/nss



Systems Security Research @ KTH Workshop, April 24, 2026

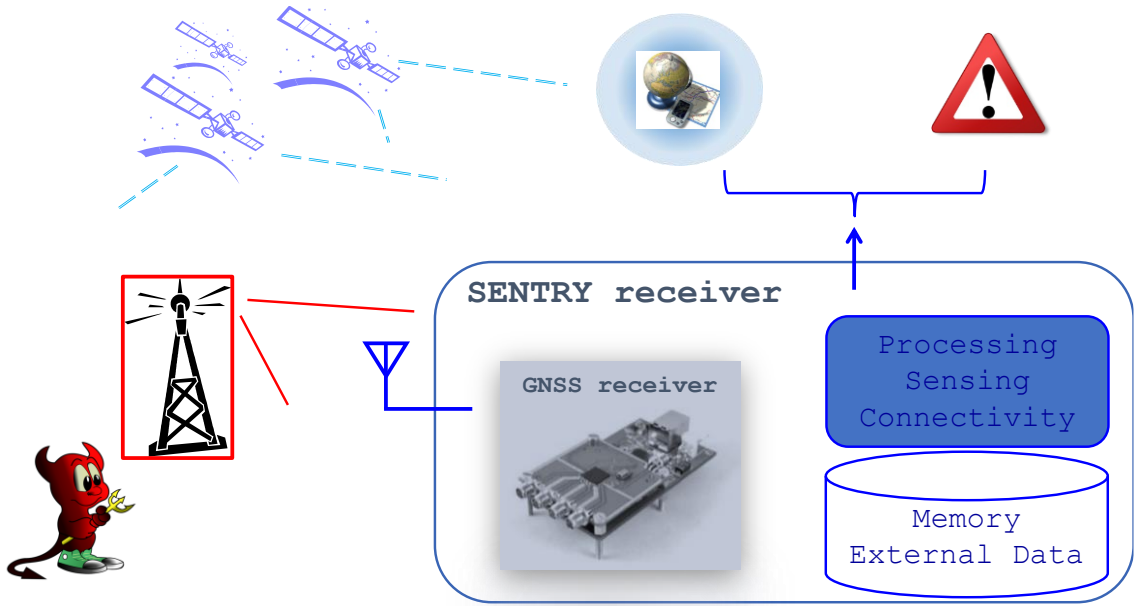


What we are after is to make current and upcoming networked systems trustworthy, shielding them and their users from attacks and abuse. Our research agenda includes a gamut of security and privacy problems, with emphasis on wireless and mobile systems. We have a strong systems character, implementing and evaluating our solutions. At the same time, we pay close attention to theoretical methods, including, notably, formal protocol analysis and information-theoretic results.

Secure location and time

Secure Global Navigation Satellite System (GNSS) receivers

- Challenge
  - Attackers can easily ‘overwrite’ legitimate Global Navigation Satellite Systems
  - Victim receivers with false position and/or time
- Solution
  - Detect the attack
  - Compute a trustworthy position and clock correction

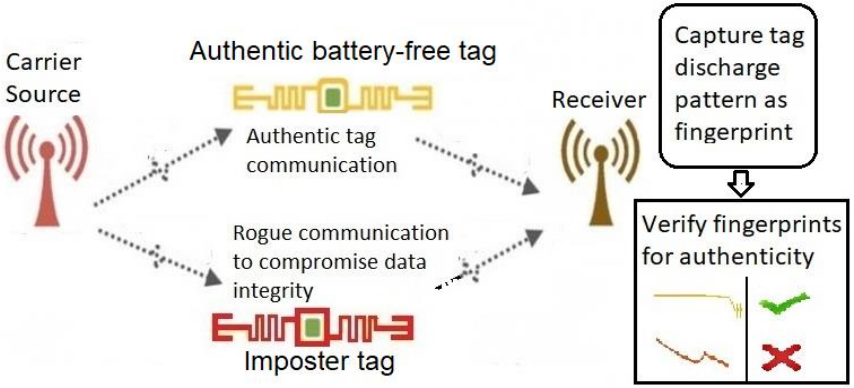


Projects with
- **European Space Agency (ESA)**
- **FOI**

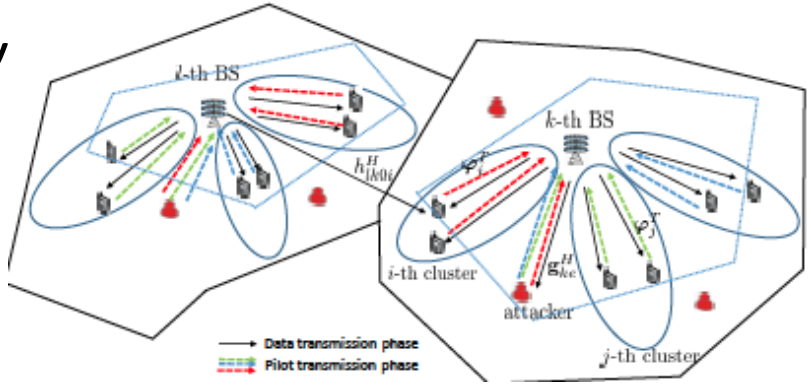
[USPO 2012, ICL-GNSS, IS-GNSS 15, IEEE Aero 2019, **ION ITM 2019**, **ION GNSS+ 2019, ION/IEEE PLANS 2020, ION GNSS+ 2020**, IEEE OJCOM 2021, **ACM WiSec 2021, ION ITM 2022, IEEE Aero 2022, …., IEEE TAES 2025, IEEE T-IFS 2025, IEEE/ION PLANS 2025 x3, ION Navigation 2026**]

Device & physical layer security

Battery-less; MIMO NOMA; Scaling laws

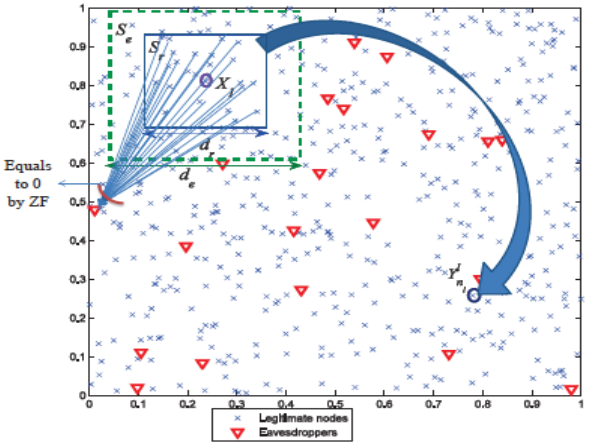


Monitoring oscillator discharge patterns: security for battery-free backscatter tags



Massive MIMO-NOMA Systems Secrecy in the Presence of Active Eavesdroppers

Scaling laws: Secrecy capacity: Secure communication for free as the network grows



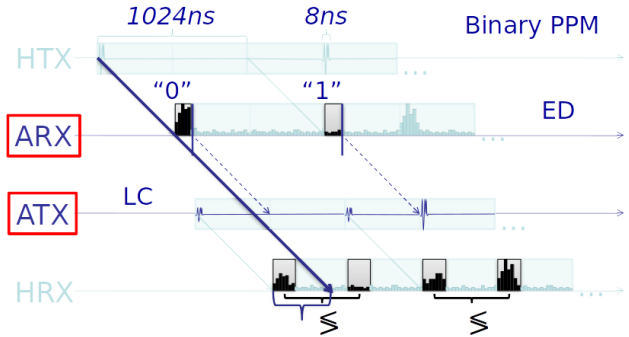
[**IEEE TWC 2026, IEEE INFOCOM 2014,** Allerton 2015, IEEE ISIT 2016, **IEEE Trans. on Information Theory 2017,** IEEE ICCCN 2021, **ACM HotNets 2021**]

Secure and privacy-preserving protocols

Cyber-physical system security

Standardization & harmonization (IEEE 1609.2, ETSI, C2C-CC)

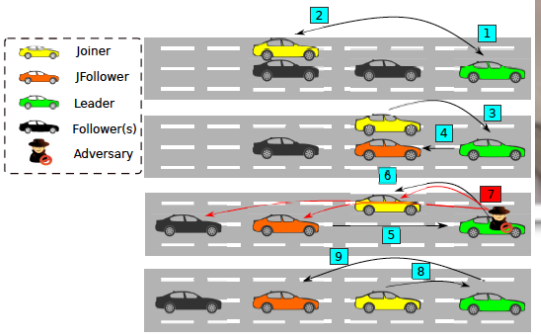
- Secure and privacy-preserving vehicle to everything protocols
- Secure driving automation – platooning
- Secure Neighbor Discovery
- Thwarting clogging denial of service attacks

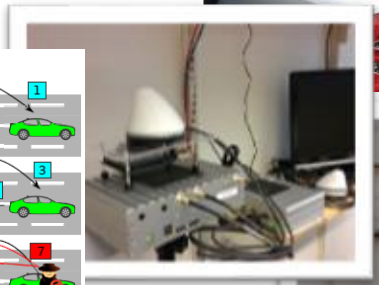


Relay time-shift ***450ns = 512ns – 62ns***
= half symbol duration – early detection time

PRESERVE
preparing secure v2x communication systems
Final event, 2015

CAR 2 CAR
COMMUNICATION CONSORTIUM
First demo, 2008





[ACM SAC 2025, Computer Security 2025, ACM WiSec 2022 & 2019, IEEE WiMob 2025, IEEE TMC 2021, IEEE IoT-J 2021]

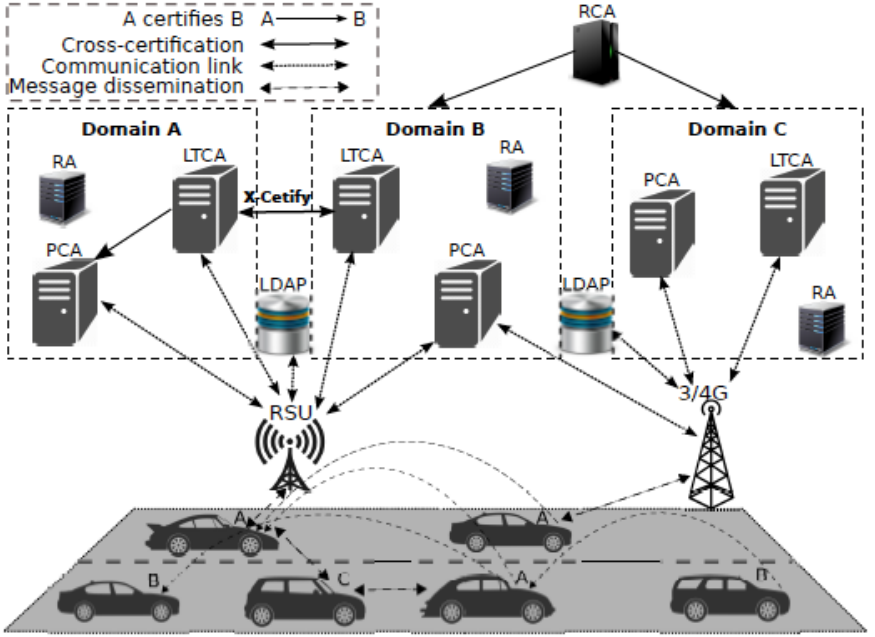
Trust management

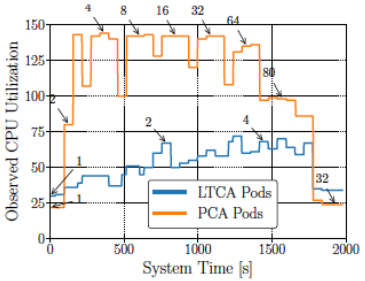
Scale up credential management

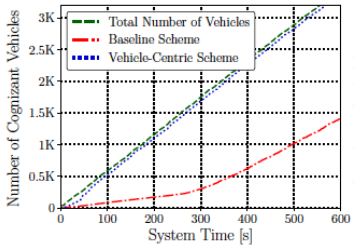
- Credential provisioning, revocation, resolution
- Balancing act: security, privacy, and efficiency
- Scalable cloud-based systems

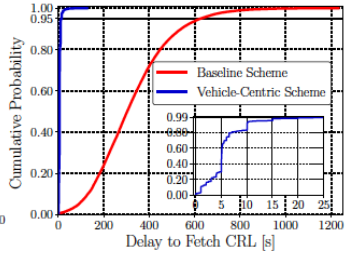
Internet Society/IRTF Applied Networking Research Prize 2018

[**IEEE TCC 2023, IEEE TMC 2021, ACM WiSec 2019, IEEE ITS 2018, ACM WiSec 2018**]





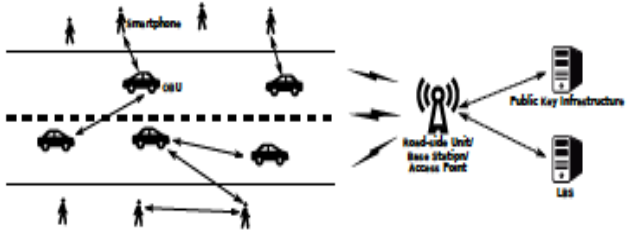


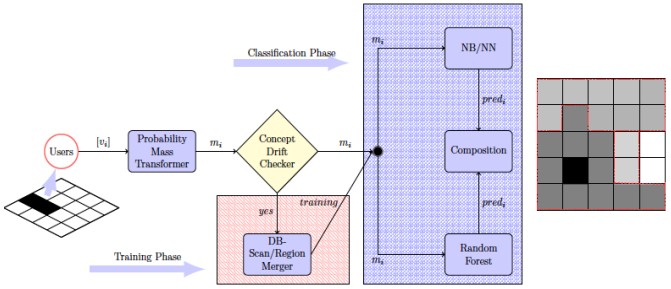


Mobile Security and Privacy

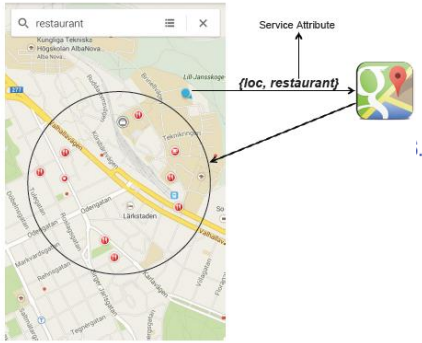
Location Based Services, Mobile Crowd-Sensing, Mobile Platforms

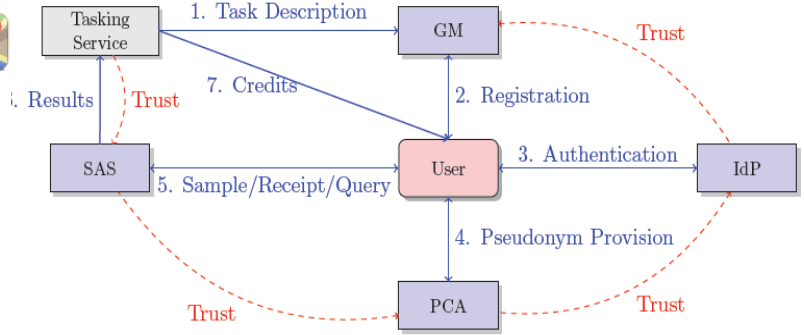
- Protect users from the system (privacy)
- Protect the system from the users (security)
- Assess data trustworthiness
- Privacy-preserving and secure queriers
- Wi-Fi Privacy





- **Attacks on Google Maps (2021 and 2025)**



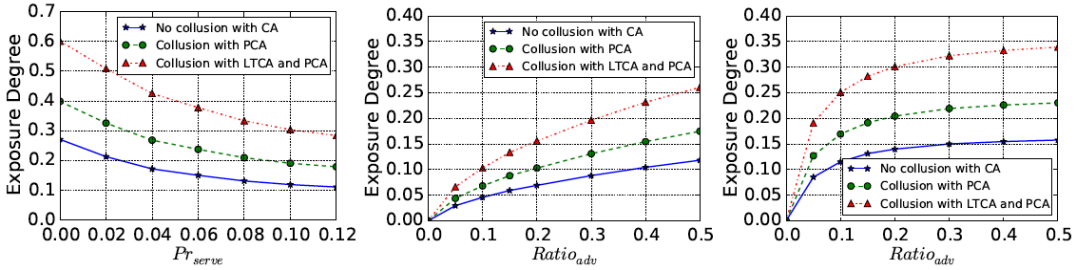


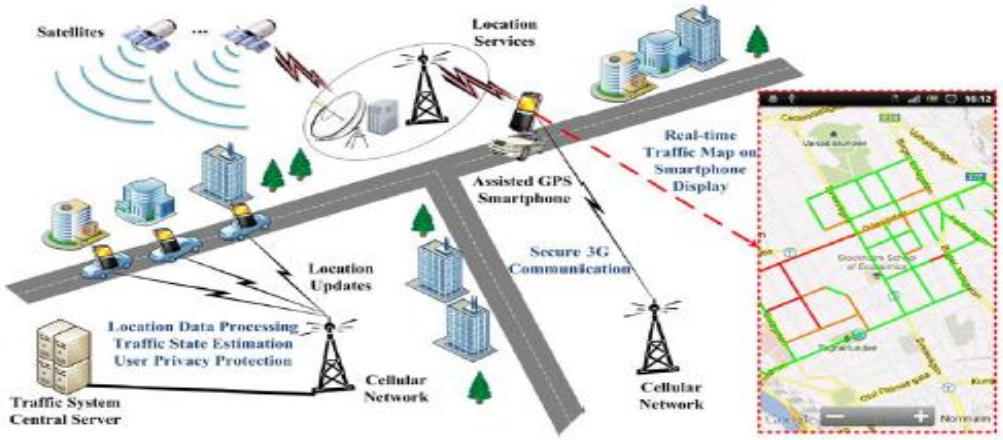
[ACM TOPS 2019, ACM WiSec 2022, IEEE TrustSense 2025, ACM CCS W 2024, ACM WiSec 2024]

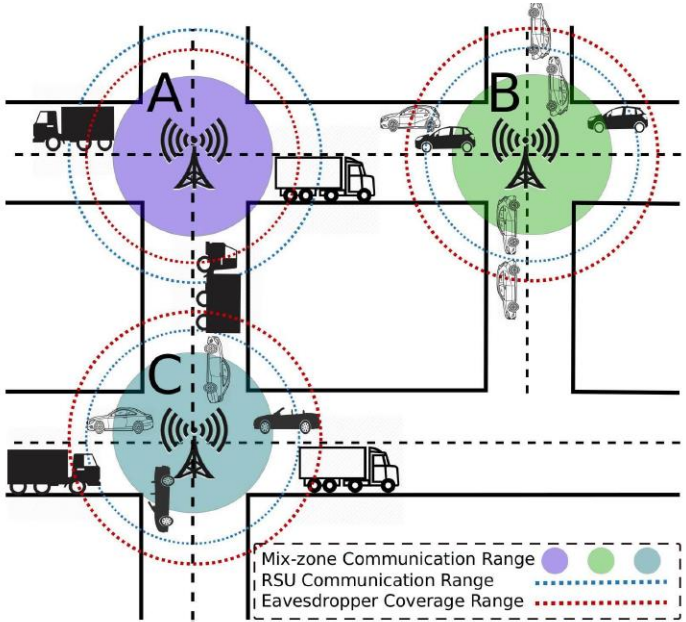
Privacy

Reduce user exposure

- Protection against
  - Attackers
  - Curious service providers







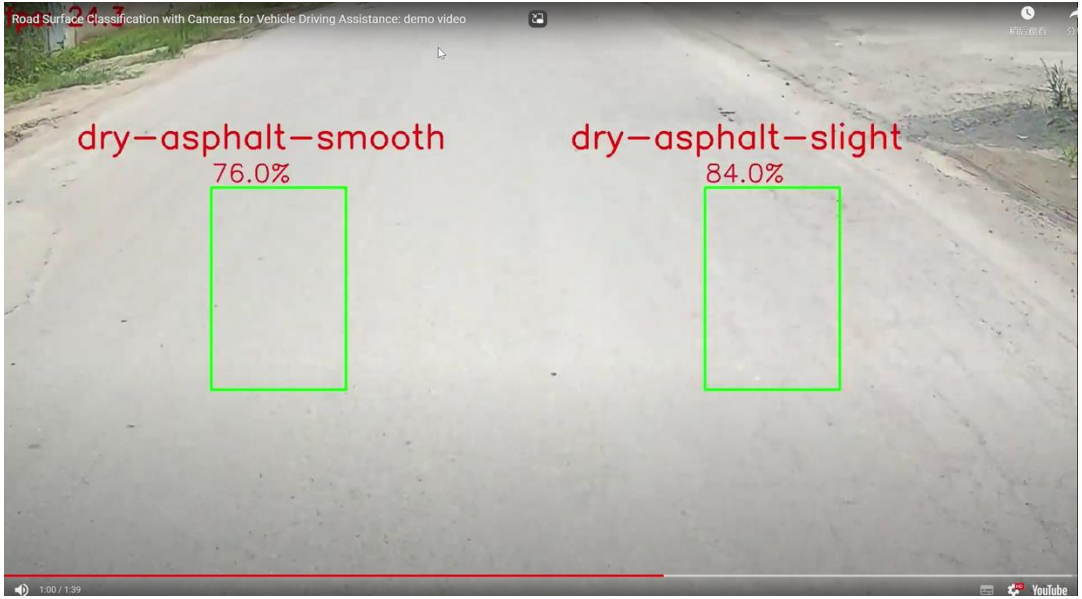
[IEEE TDSC 2014, ACM TOPS 2019, IEEE IoT 2021]

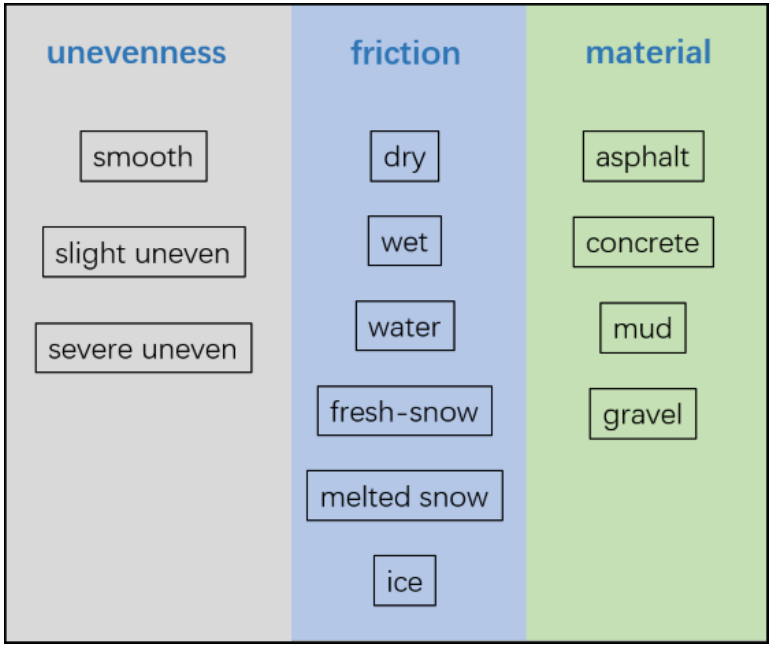
ACM WiSec Best Paper Award 2014

[IEEE T-ITS 2015, IEEE IoT 2016]

Safeguarding Federated Learning-based Road Condition Classification

Road Condition Classification (RCC)





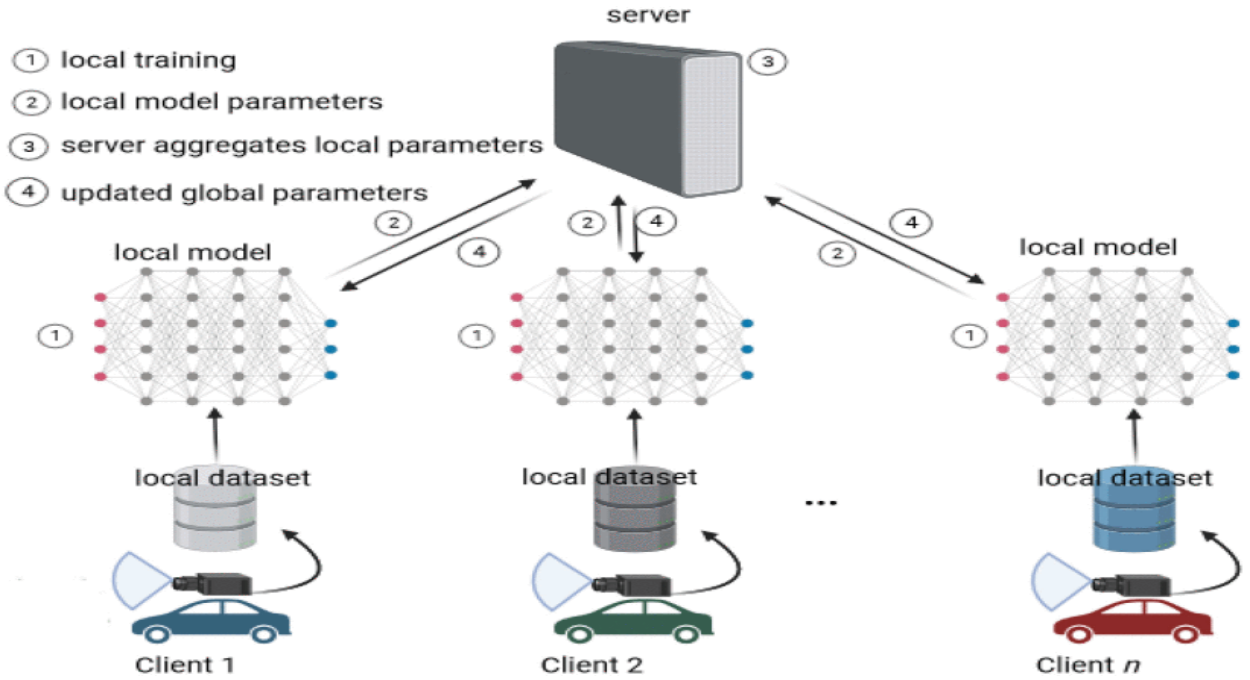
[IEEE CNS 2025, ACM SAC 2026]

Safeguarding Federated Learning-based Road Condition Classification

Using ML for RCC or Federated Learning (FL)

Impractical in the long term

- Heterogeneity
- Imbalanced resource consumption
- Privacy regulations [1][2][3]

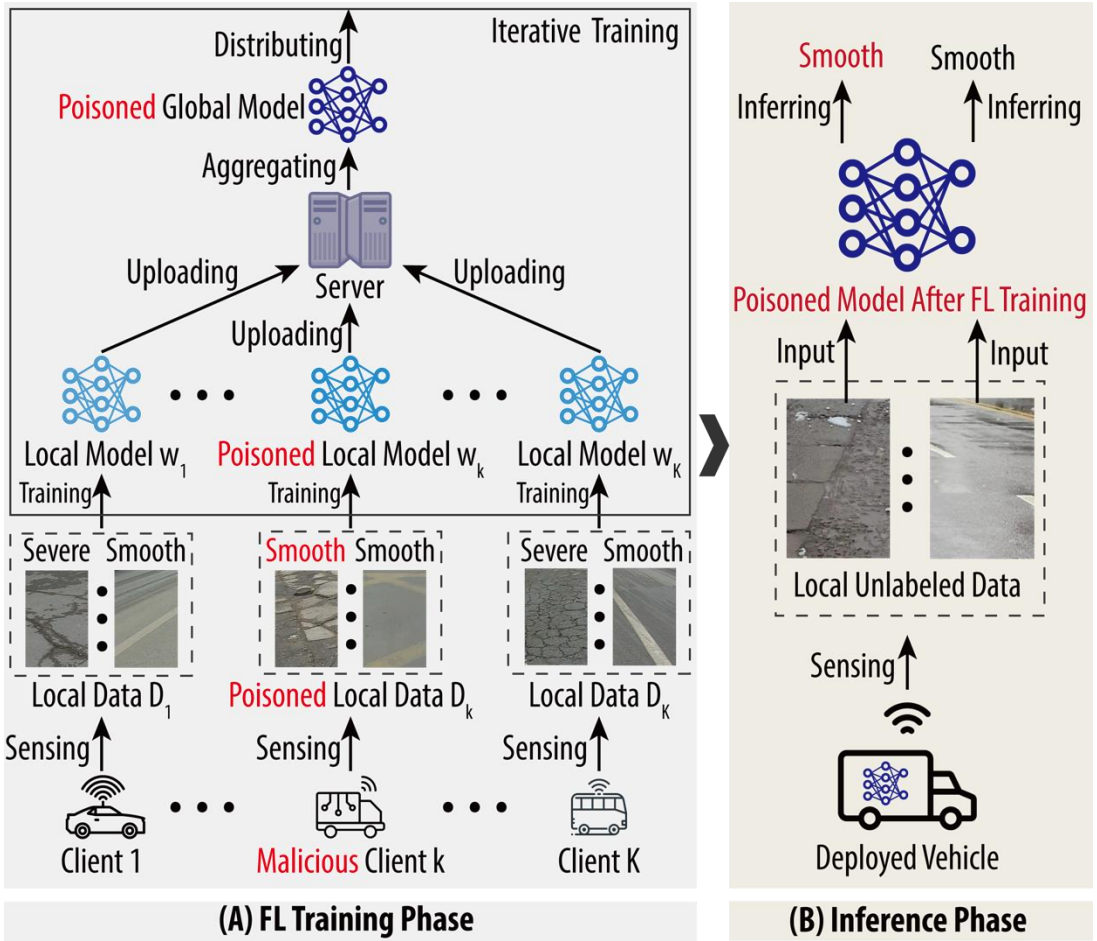


[1] https://gdpr-info.eu
[2] https://www.oag.ca.gov/privacy/ccpa
[3] http://en.npc.gov.cn.cdurl.cn/2021-12/29/c_694559.htm

[**IEEE CNS 2025, ACM SAC 2026**]

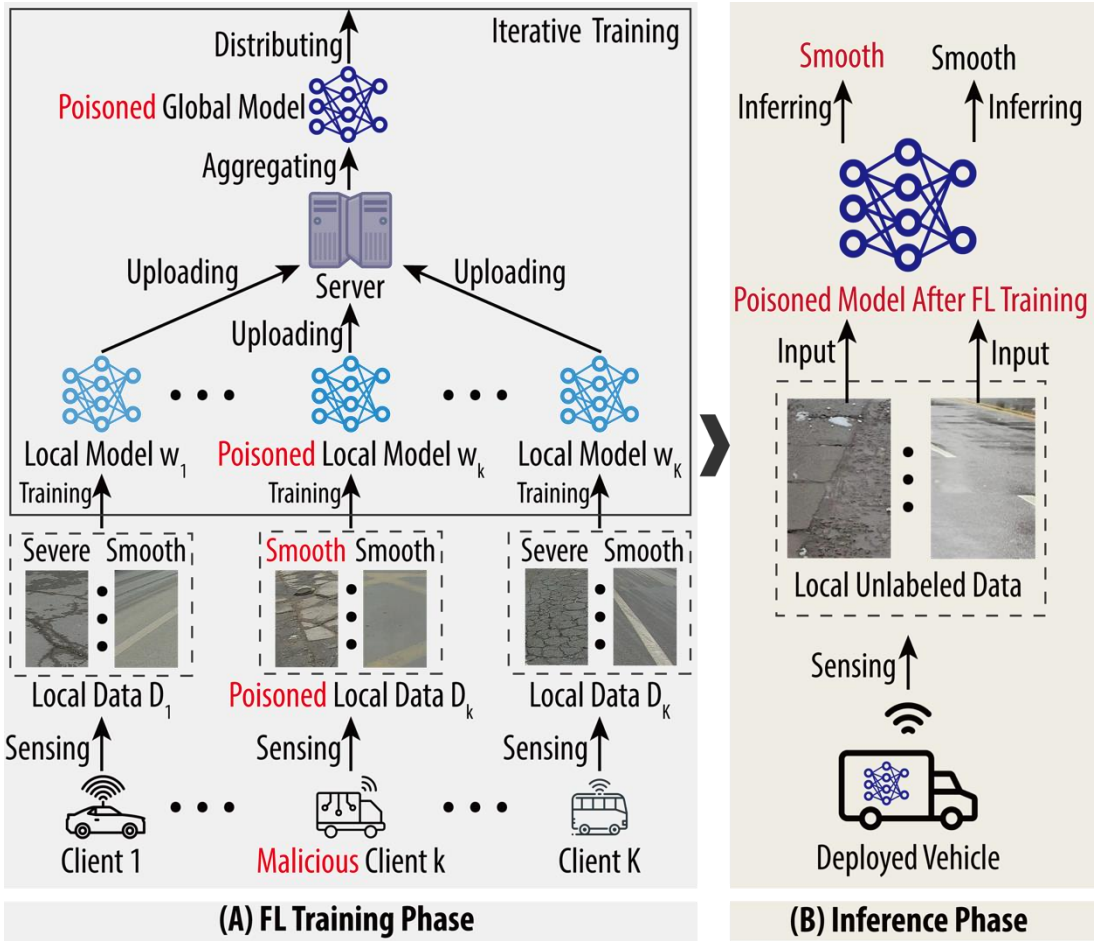
Attacking FL

Targeted Label Flipping Attacks (TLFAs) on FL-RCC



Attacking FL

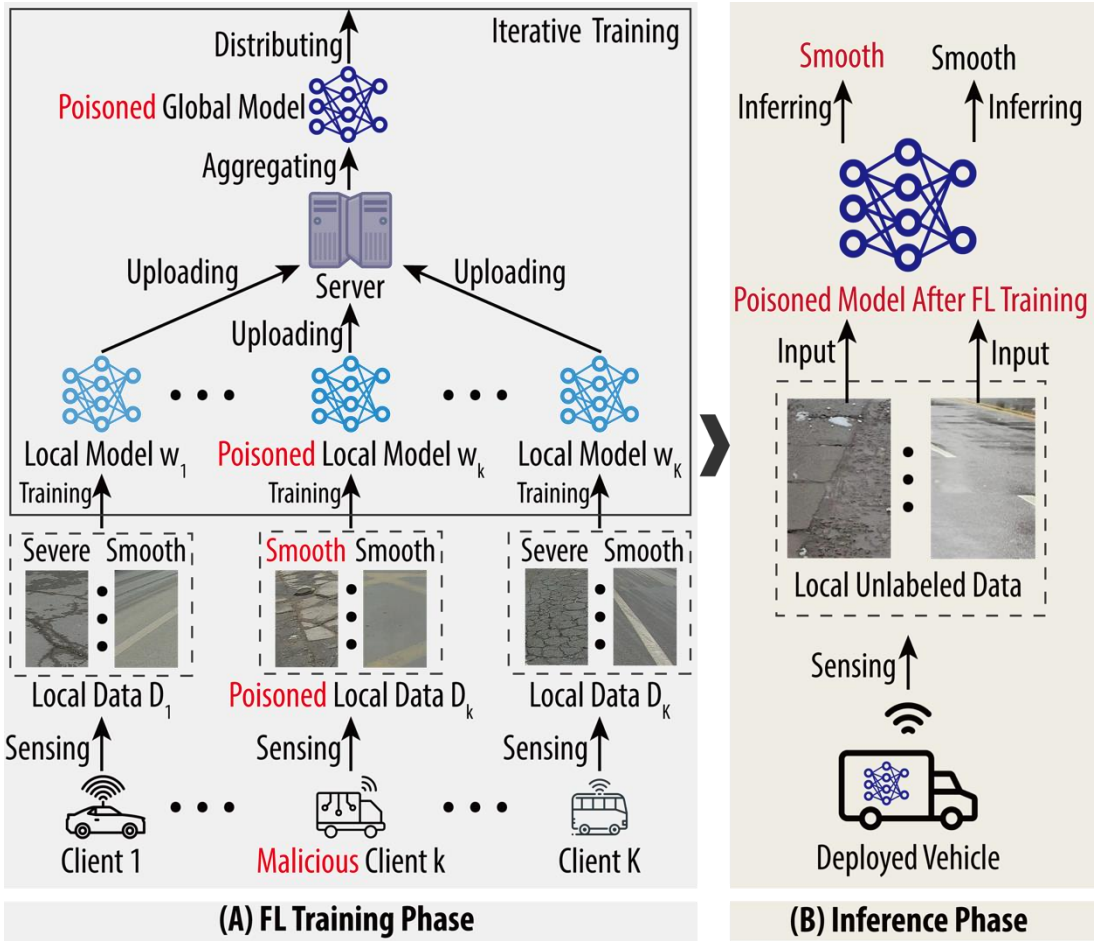
Targeted Label Flipping Attacks (TLFAs) on FL-RCC (cont'd)



- Malicious clients (i.e., **vehicles**) deliberately mislabel data from the source class (e.g., **severe-uneven** in the figure) to a target class (e.g., **smooth** in the figure) through label-flipping, which threatens **transportation safety** because the safety risk is underestimated.

Attacking FL

Targeted Label Flipping Attacks (TLFAs) on FL-RCC (cont'd)

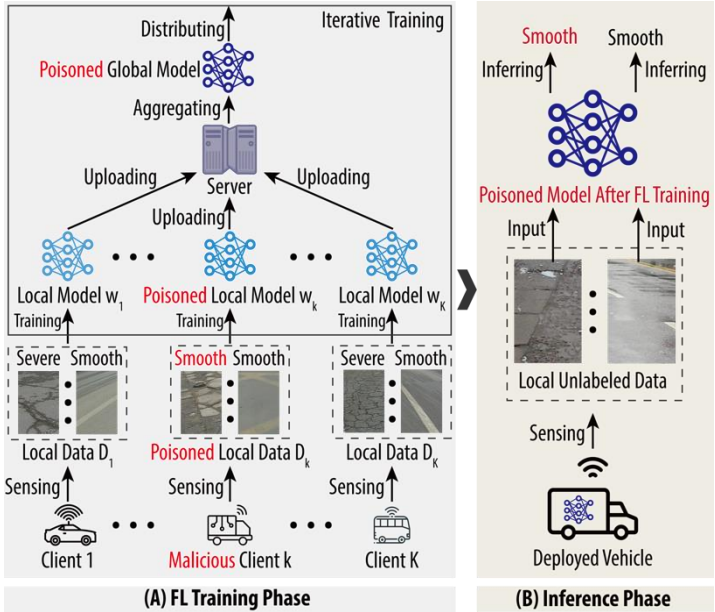


- If the flipping direction is the **opposite**, the attack goal is **traffic efficiency** rather than transportation safety.

Safeguarding FL-RCC

Key ideas and contributions

1) TLFAs have a **significant impact** on existing FL-RCC systems.
2) New label-distance-based **metric** to precisely quantify the transportation safety risks posed by TLFAs.
3) FLARE, a new **defensive mechanism** leveraging neuron-wise analysis to mitigate TLFA effects
4) Improved, extended defense mechanism, DEFEND



www.eecs.kth.se/nss

Safeguarding FL-RCC

FLARE protocol evaluation (cont'd)

Three Evaluation Metrics:

- **Global Accuracy (GAcc)**: The ratio of all correct predictions to the total number of testing instances.
- **Source recall (SRec)**: The proportion of the correct predictions for the source class to the total number of samples in the source class.
- **Attack Success Rate (ASR)**: The ratio of samples with the source label misclassified into the target class

$$GAcc = \frac{\sum_{i=1}^{L} n_{i,i}}{\sum_{i=1}^{L} \sum_{j=1}^{L} \times n_{i,j}}$$

$$SRec = \frac{n_{sr,sr}}{\sum_{i=1}^{L} n_{sr,i}}$$

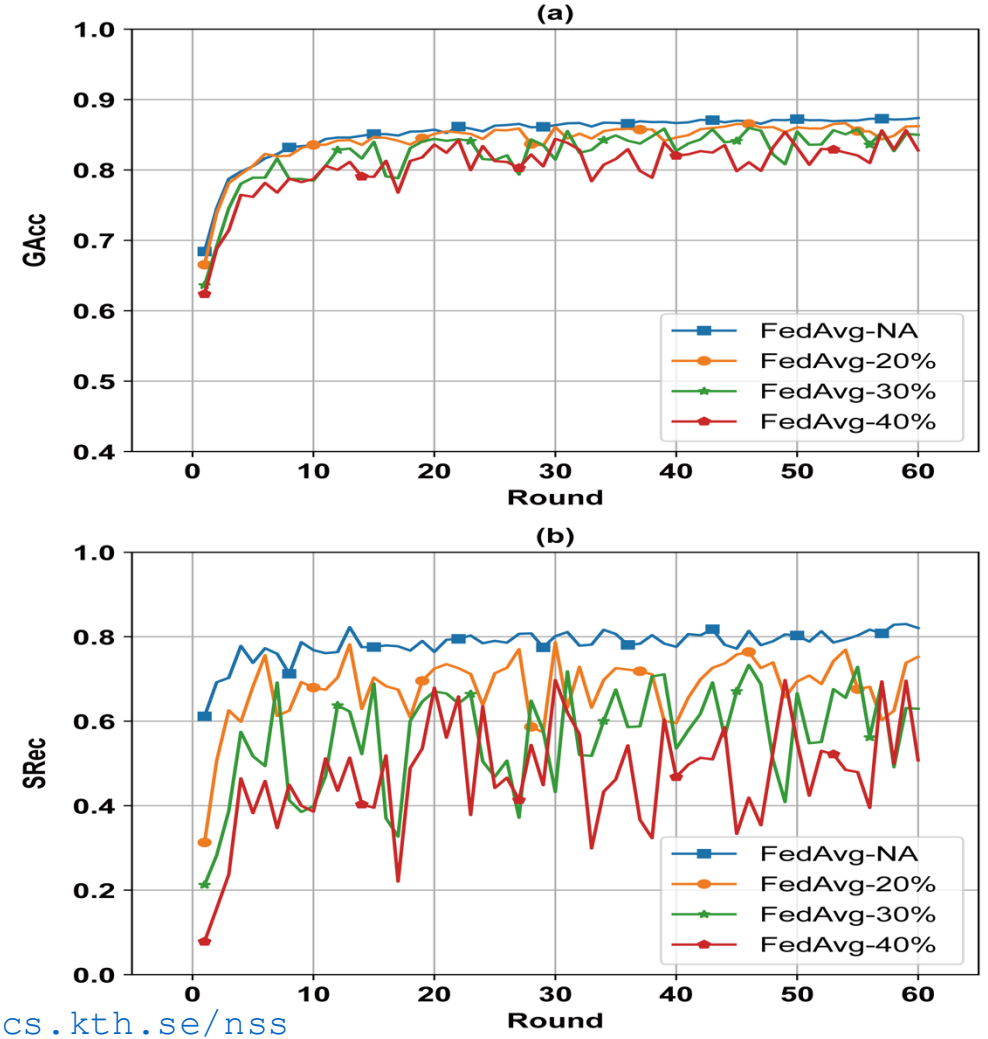
$$ASR = \frac{n_{sr,tr}}{\sum_{i=1}^{L} n_{sr,i}}$$

Safeguarding FL-RCC

FLARE protocol evaluation (cont'd)

Influence Analysis of TLFA on FL-RCC

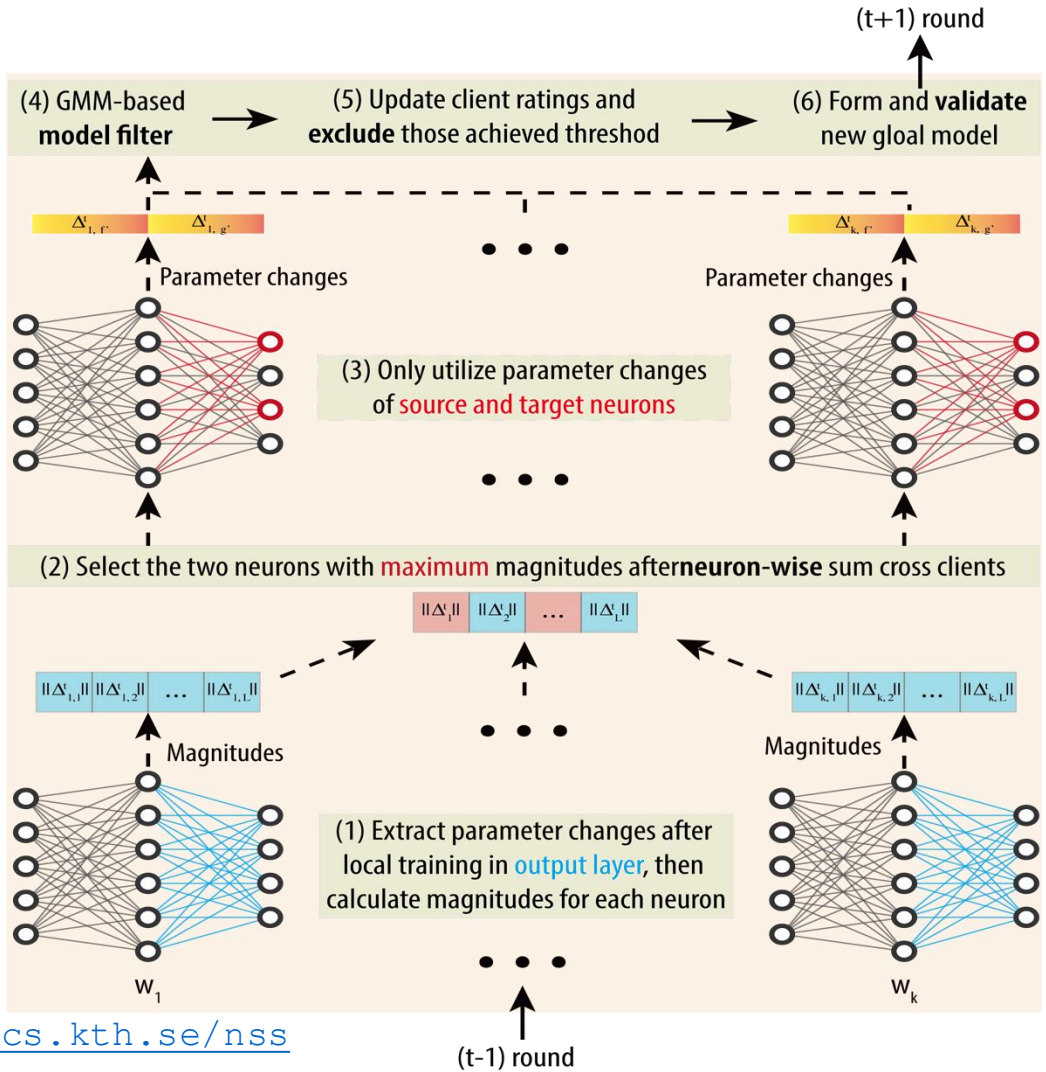
- TLFA primarily influences ASR and SRec.
- TLFA has a more significant impact on Unevenness.
- A higher poisoned rate leads to worse performance.



Safeguarding FL-RCC

DEFEND protocol

- **DEFEND**[1]: Poisoned model Detection and malicious client Exclusion mechanism for FEderated learning-based road coNDition classification



Safeguarding FL-RCC

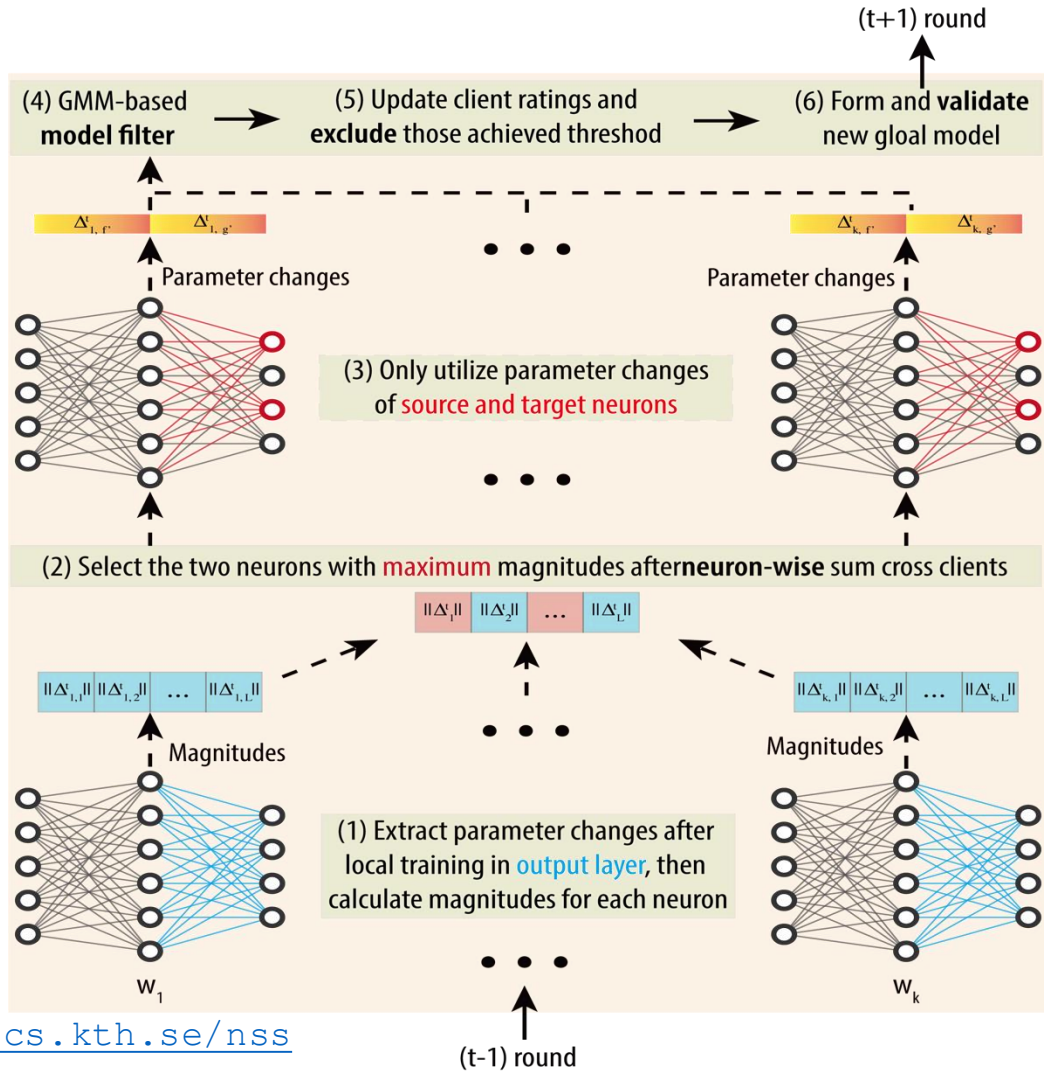
DEFEND protocol (cont'd)

- **DEFEND**[1]:
  - Calculate **L_2-norm**-based magnitudes rather than the direct changes.

$$\|\Delta_{k,l}^t\|_2 = \sqrt{\sum_{i=1}^{d_l} \left(\Delta_{k,l,i}^t\right)^2}, \qquad k \in C_t,\ l \in L.$$

$$\|\Delta_l^t\|_2 = \sum_{k \in C_t} \|\Delta_{k,l}^t\|_2, \qquad l \in L.$$

  - Use **GMM** (Gaussian Mixture Model) for clustering rather than HDBSCAN.



Safeguarding FL-RCC

DEFEND protocol (cont'd)

- **Client rating** strategy rather than a count-based strategy.
  - ✓ A client rating is **decremented** if its model detection result is poisoned in a round and

$$r_k(t) = \begin{cases} \max\{\, r_k(t-1) - \gamma,\ r^{\min} \,\}, & \text{if } c_k \text{ input bad,} \\ \min\{\, r_k(t-1) + \beta,\ r^{\max} \,\}, & \text{if } c_k \text{ input good.} \end{cases}$$

- **Validation** strategy via SRec and ASR values during training.
  - ✓ If these metric values achieve predefined performance change **thresholds**, the current global model is still considered poisoned even after the filter and is discarded.

Algorithm 1 Protocol of DEFEND

Initialize black list $\mathbb{B} = \emptyset$, $SRec^{old} = 0$, $ASR^{old} = 1$, and rating value $r_k(0) = \delta(r^{max} - r^{min})$ for each client v_k
for each round $t \in [1, T]$ **do**
  $\mathbb{C}^t \leftarrow$ randomly select M clients from $\mathbb{C} - \mathbb{B}$
  The server sends ω^t to all clients in $\mathbb{C}^t$
  for each client $c_k \in \mathbb{C}^t$ in parallel **do**
    Update local model ω_k^t
    Send ω_k^t back to the server
  end for
  The server receives ω_k^t from $\mathbb{C}^t$
  for each ω_k^t the server **do**
    $\Delta_{k,L}^t = \{\omega_{k,l}^t - \omega_l^t | l \in L\}$ ▷ Output layer changes
    Calculate magnitudes $||\Delta_{k,l}^t||$ for $l \in L$ ▷ ℓ_2-norm
  end for
  $\{||\Delta_{l_1}^t||, ..., ||\Delta_{l_E}^t||\} \leftarrow$ neuron-wise magnitudes
  $f', g' \leftarrow$ Top-2($\{||\Delta_{l_1}^t||, ..., ||\Delta_{l_E}^t||\}$) ▷ $f' < g'$
  $U^t \leftarrow \{\Delta_{k,l}^t | c_k \in \mathbb{C}^t, l \in \{g', f'\}\}$
  $\mathbb{C}_{out}^t = \text{GMM}(U^t)$
  $\omega^{t+1} = Aggeregate\{\omega_k^t | c_k \notin \mathbb{C}_{out}^t\}$
  $SRec^{new}, ASR^{new} \leftarrow$ Validate (ω^{t+1})
  $\Delta SRec = SRec^{new} - SRec^{old}$, $\Delta ASR = ASR^{new} - ASR^{old}$
  if $\Delta SRec < SRec^{thr}$ or $\Delta ASR > ASR^{thr}$ **then**
    $\omega^{t+1} = \omega^t$
  end if
  $SRec^{old} = SRec^{new}$, $ASR_k^{old} = ASR_k^{new}$
  for $c_k \in \mathbb{C}^t$ **do**
    if $c_k \in \mathbb{C}_{out}^t$ **then**
      $r_k(t) = max\{r_k(t-1) - \gamma, r_{min}\}$
      if $r_k(t) \leq r^{min}$ *And* $c_k \notin \mathbb{B}$ **then**
        Add c_k in $\mathbb{B}$
      end if
    else
      $r_k(t) = min\{r_k(t-1) + \beta, r^{max}\}$
    end if
  end for
end for
return ω^{T+1}

Safeguarding FL-RCC

DEFEND protocol (cont'd)

Experimental Setting: similar to FLARE

- participant rate: 20% (20 out of 100)
- Poisoned rate: 30% (30 out of 100)

Performance Analysis

- **2.81%, 24.99%, and 15.78% improvement** against best baselines for DEFEND regarding GAcc, SRec, and ASR, respectively.
- Average ASR and SRec values of DEFEND are 5.22% and 79.79%, respectively, even slightly better than FedAvg-NA (5.48% and 77.18%, respectively).

DEFEND can achieve the same performance as attack-free situations when under attack!

Model	Method	RCC @ Friction		
		GAcc ↑	SRec ↑	ASR ↓
ResNet-18 [8]	FedAvg-NA‡ [22]	85.26	72.28	5.84
	FedAvg [22]	72.83	44.88	30.85
	Krum [4]	81.44	48.89	22.92
	TMean [40]	80.78	45.84	18.48
	Median [40]	81.56	48.44	22.52
	FoolsGold [7]	81.64	50.64	15.56
	FLAME [23]	63.32	51.20	21.28
	FLARE [18]	82.80†	61.72	14.64
	DEFEND (Ours)	**84.93***	**74.20**	**2.84**
EfficientNet-B1 [32]	FedAvg-NA	86.08	75.68	3.40
	FedAvg	80.48	38.12	32.96
	Krum	64.49	4.92	69.32
	TMean	81.46	46.52	27.04
	Median	82.54	53.88	22.16
	FoolsGold	83.64	59.80	16.76
	FLAME	82.56	52.36	20.96
	FLARE	83.46	62.04	16.08
	DEFEND (Ours)	**84.43**	**80.40**	**5.40**
Deit-Tiny [34]	FedAvg-NA	86.54	77.32	3.44
	FedAvg	77.89	25.52	48.40
	Krum	67.15	17.48	38.04
	TMean	64.03	34.88	40.84
	Median	78.91	33.64	45.72
	FoolsGold	82.61	53.60	25.08
	FLAME	57.54	45.60	25.84
	FLARE	83.02	58.84	17.80
	DEFEND (Ours)	**84.58**	**80.76**	**4.84**

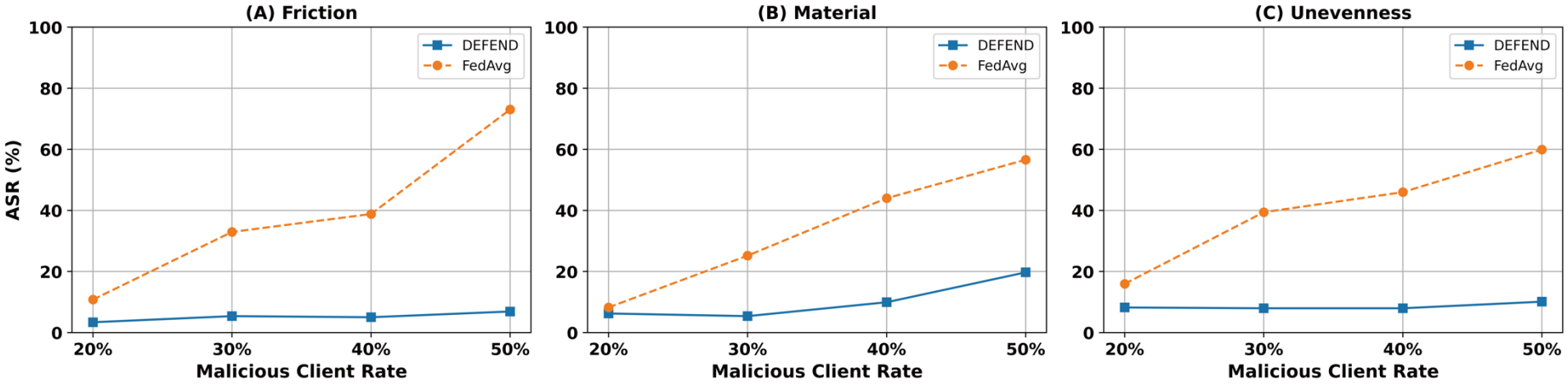
* **Bold numbers are the best performances in a group.**

† Numbers with underline are the second-best values in a group.

‡ NA denotes No Attack. Others without this symbol are all under TLFAs.

Safeguarding FL-RCC

DEFEND protocol (cont'd)

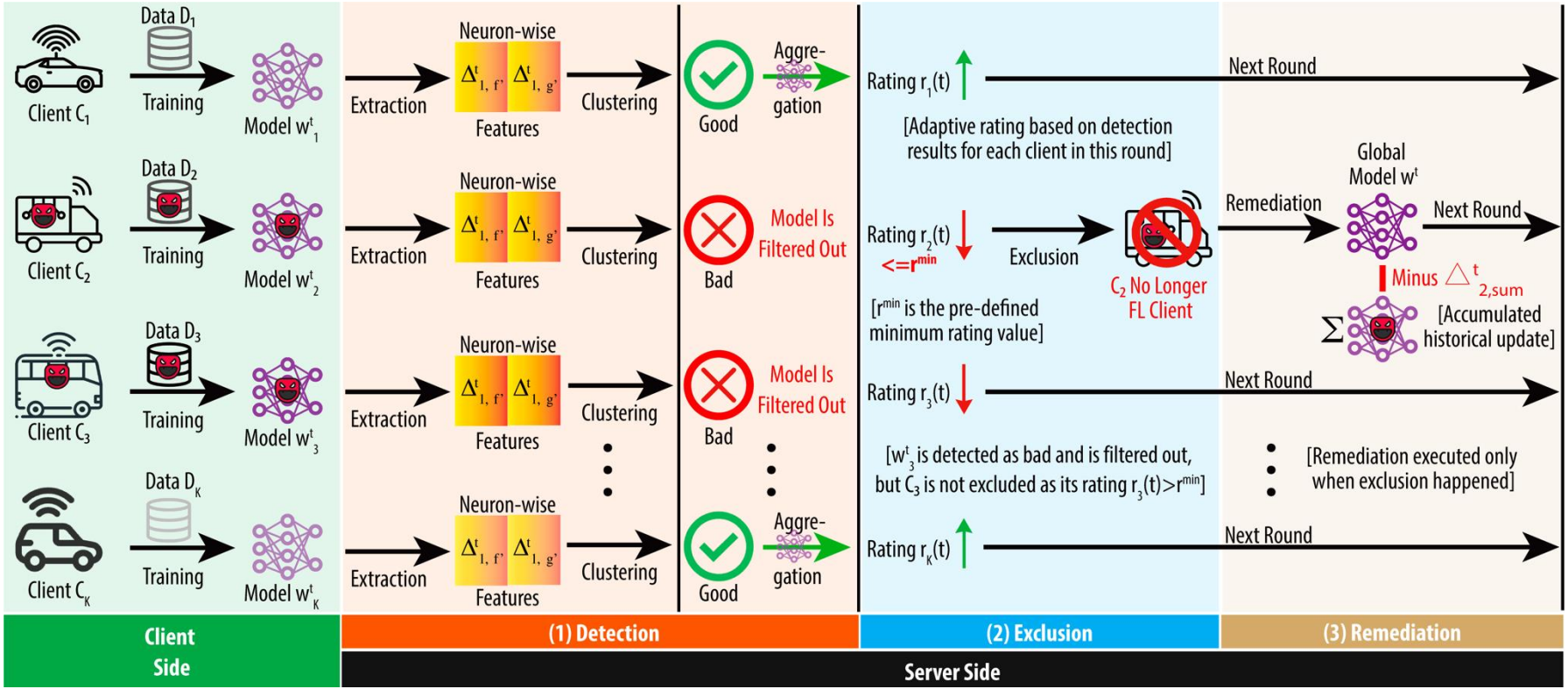


Performance Analysis

- When malicious client rates increase from **20% to 50%,** DEFEND still keeps ASR values very **low**
- FedAvg performances deteriorate significantly

Safeguarding FL-RCC

FedTrident (cont'd)



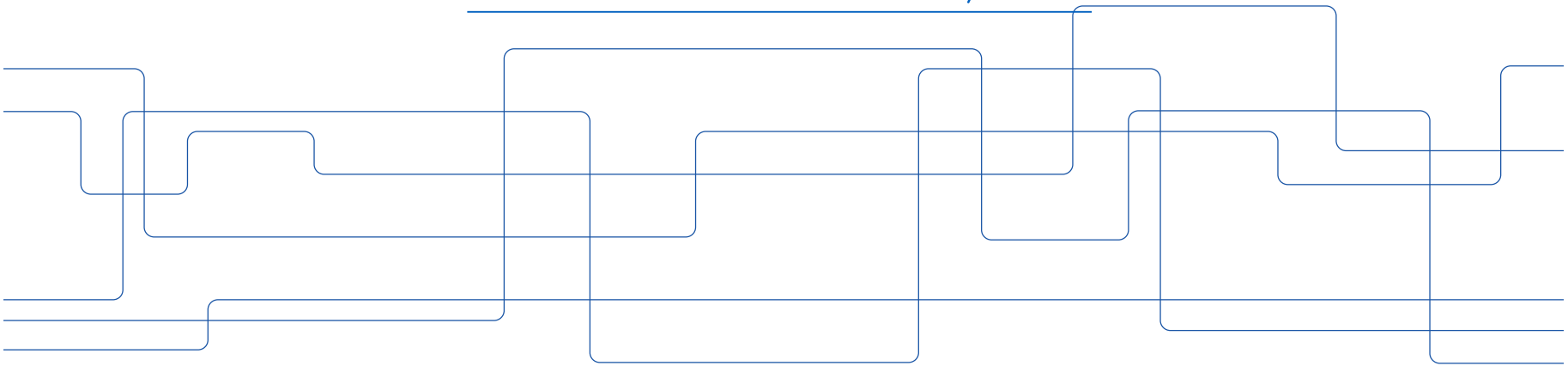
- Latest work – FedTrient – further improves resilience: https://arxiv.org/pdf/2603.19101

KTH ROYAL INSTITUTE OF TECHNOLOGY

Networked Systems Security

Panos Papadimitratos

www.eecs.kth.se/nss



Systems Security Research @ KTH Workshop, April 24, 2026