(Online) ISSN 2744-1741
Original Research

Defense and Security Studies
Vol. 7, No. 2, 2026, pp.260-284
https://doi.org/10.37868/dss.v7.id349

Resilient network design for cognitive radio networks under spectrum sensing data falsification attacks

Sarah Ali Abdullah[1*]
[1] University of Information Technology and Communications, College of Engineering, Baghdad, Iraq

*Corresponding author E-mailsarahaabdullah@uoitc.edu.iq

Received: Sep. 9, 2026
Revised: Oct. 12, 2026
Accepted: Oct. 13, 2026
Online: Oct. 14, 2026

Abstract

Cognitive radio networks (CRNs) rely on cooperative spectrum sensing (CSS), but fusing reports from distributed secondary users (SUs) exposes the network to Spectrum Sensing Data Falsification (SSDF) attacks. Most countermeasures evaluate SSDF defenses at the sensing layer, leaving open whether the network continues to meet its service requirements under attack. This paper presents SC-Joint, an integration framework that treats SSDF defense as an admission constraint within a cross-layer problem. A Beta–Bernoulli reputation layer admits SUs, and the cooperating set and sensing time are then jointly selected subject to QoS-SLA constraints on detection probability, throughput and network lifetime. The reputation is oracle-free: each SU is scored by its agreement with the leave-one-out majority of the other SUs, never against ground-truth labels. Under OR fusion, a single always-yes attacker saturates the global false-alarm probability and drives throughput to zero, and under nested injection the undefended baseline loses feasibility as attackers are added (Propositions 1–2). We derive an admission condition (Proposition 3) that predicts the falsification probability below which stealthy attackers are admitted ($p^* \approx 0.44$) and the PU activity above which honest high-SNR SUs are excluded ($P(H_1) \approx 0.42$). MATLAB simulations of a 12-SU network indicate that undefended OR fusion loses all feasibility once one attacker is admitted, whereas SC-Joint sustains 79.7–100% feasible allocation at about 2.1 Mb/s, and 97–100% under network stress, matching oracle exclusion in paired tests. The ablation suggests that the gain stems primarily from trust-based admission, whereas sensing-time optimization mainly secures lifetime. The mechanism is, however, limited to OR fusion, $P(H_1) < 0.45$ and non-colluding attackers; stealthy attackers with $p < p^*$ are admitted.

© The Author 2026.
Published by ARDA.

Keywords: Cognitive Radio Networks; Beta–Bernoulli Reputation; SSDF attacks; Cross-Layer Optimization; Network Feasibility.

1. Introduction

The limited radio spectrum remains under pressure from the rapid growth of wireless connectivity driven by the Internet of Things (IoT), automation networks, and emerging sixth-generation (6G) services. Mitola and Maguire introduced cognitive radio (CR) as a way to make software radios more personal [1], and Haykin later

This work is licensed under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/) that allows others to share and adapt the material for any purpose (even commercially), in any medium with an acknowledgement of the work's authorship and initial publication in this journal.

recast it as a brain-empowered wireless system that is aware of its environment [2]. In CR, SUs are unlicensed secondary users who can access licensed bands if PUs (primary users) leaves them idle, but they have to give the frequency or power off immediately PUs become active [3]. Hence everything hinges on spectrum sensing. Energy detection is the common method because it boils down the decision to "the sum of the squares of statistically independent Gaussian variates"[4]. But usually, SU alone cannot sense in fading reliably or if hidden it is terminal. Cooperative spectrum sensing (CSS) comes for the rescue: by combining several SUs decisions CR network not only can detect licensed users more accurately but also reduce the probability of interference. Data fusion constitutes the central element of CSS. For example, Chen et al. studied the Byzantine failure problem in data fusion [5]. The cause of such fault may be either faulty sensing device or even a Spectrum Sensing Data-Falsification (SSDF) attack. [6]; in any case the collector is fed with distorted sensing info and the outcome of fusion may be distorted [5]-[7]. OR-approach is quite popular for spatial diversity, also Chair-Varshney combiner is an example [8]. They remain bound by the sensing–throughput trade-off [9], [10], and in battery-limited SUs that trade-off is tied in turn to energy and lifetime [11], [12]. The aggregation that makes CSS strong is also where it is weakest. In an SSDF (or Byzantine) attack, compromised SUs sends fabricated reports to mislead the FC [13], [14]. OR fusion is permissive, so the damage can be out of proportion to the number of attackers. A handful of "always-yes" attackers is enough to raise the global false-alarm probability, push SUs off idle channels and drive throughput down [5], [14]. The problem persists with very large attacker populations [15], and it also affects learning-based sensing, whose detectors can themselves be manipulated by an adversary [16].

Most remedies in the literature belong to one of two groups. One group tries to detect and remove malicious reports, using outlier filtering, clustering or learned detectors [17] [18]. In other methods, the trustworthiness of a network participant is determined not only by their past behaviors but also through mechanisms such as blockchain or adaptive threshold which in some cases support the establishment of a node's credibility through such a reputation system [15], [19], [20], [21], [22]. On the other hand, reinforcement-learning-based schemes are able to fine-tune their defenses while running the program [23]. The two groups differ in mechanism but are usually judged in the same way. Performance is reported at the sensing layer, as a detection rate or a classification accuracy, and many of the methods need labeled data or a ground-truth oracle [24].

This leaves an important question open, since detecting an attacker is not equivalent to keeping the network operational. What matters to a deployed CRN is whether feasible allocations still exist, whether the throughput SLA is met, and whether the lifetime requirement is satisfied; these outcomes are governed by network-layer variables (which SUs are admitted, the cooperating-set size, sensing time, and resource allocation). Yet cross-layer resource studies optimize these while assuming honest reporting [25], [26], [27], [28], whereas security studies optimize detection while largely ignoring its allocation consequences; physical-layer-security and game-theoretic formulations likewise operate below the QoS level [29], [30], [31]. To our knowledge, few frameworks treat SSDF resilience as a constraint on network QoS, and fewer still do so without an oracle.

To address this gap, we treat SSDF defense as an admission constraint inside a network-design problem rather than as a detection problem alone. Figure 1 contrasts detection-level evaluation with the network-level evaluation adopted here. The proposed SC-Joint framework combines an oracle-free trust layer with joint optimization of the cooperating-set size K and the sensing time τ under detection, throughput and lifetime constraints. The trust layer scores each SU by a Beta–Bernoulli reputation computed from its agreement with the leave-one-out majority of the other SUs. SUs whose reputation falls below a threshold are excluded before fusion, and the configuration is then re-optimized over the admitted SUs, without ground-truth labels.

1.1 Contributions of this paper

The main contributions of this paper are as follows:

- ***SSDF defense as an admission constraint in a cross-layer problem.*** We integrate established components into a single network-level design. These are reputation-based trust [13], [22] and joint

optimization of the cooperating-set size K and the sensing time τ [9], [10]. SSDF defense enters as an admission constraint on the cooperating set. The optimizer then maximizes throughput subject to QoS-SLA constraints on detection probability, throughput and network lifetime. The evaluation target is therefore whether the network continues to meet its SLA under attack, rather than attacker detection alone.

- ***Oracle-free reputation admission.*** Each SU is scored by its agreement with the leave-one-out majority of the other SUs, without ground-truth PU labels (Section 4). At the evaluated operating point this score reduces to a report-rate threshold, and we report this equivalence explicitly (Section 7.13). The component ablation (Table 5) indicates that trust-based admission is the decisive component; sensing-time optimization mainly secures network lifetime.
- ***Analytical characterization.*** Propositions 1–2 state two design observations under OR fusion. A single always-yes SU drives throughput to zero, and under nested injection the undefended baseline loses feasibility as attackers are added. Proposition 3 gives an admission condition that bounds the probability of mis-admission. Its two closed-form thresholds, the stealth threshold $p^* \approx 0.44$ and the honest-exclusion threshold $P(H_1) \approx 0.42$, agree closely with the simulations.
- ***Evaluation with an explicit operating envelope.*** SC-Joint is compared with NaïveOR, a binary-report adaptation of Outlier/MAD [32], a report-rate threshold and a decayed reputation, under always-yes, always-no and probabilistic-flip attackers. All comparisons are paired on common random numbers and backed by robustness checks, and the code is public. We also state the operating envelope explicitly. Within this envelope, the mechanism appears to resist non-colluding SSDF under OR fusion for $P(H_1) < 0.45$. Stealthy attackers with $p < p^*$ get through, and this is what creates the detection valley in Section 7.9. Always-no attackers also get through, but they do almost no harm. Duty-cycled, colluding and majority-forming adversaries lie outside the envelope.

The rest of the paper is structured as follows. Related work is reviewed in Section 2, and the system and attack models are given in Section 3. Sections 4–5 develop SC-Joint and its analysis. Sections 6–9 present the simulation setup, results, discussion, and limitations with future work, and Section 10 concludes.

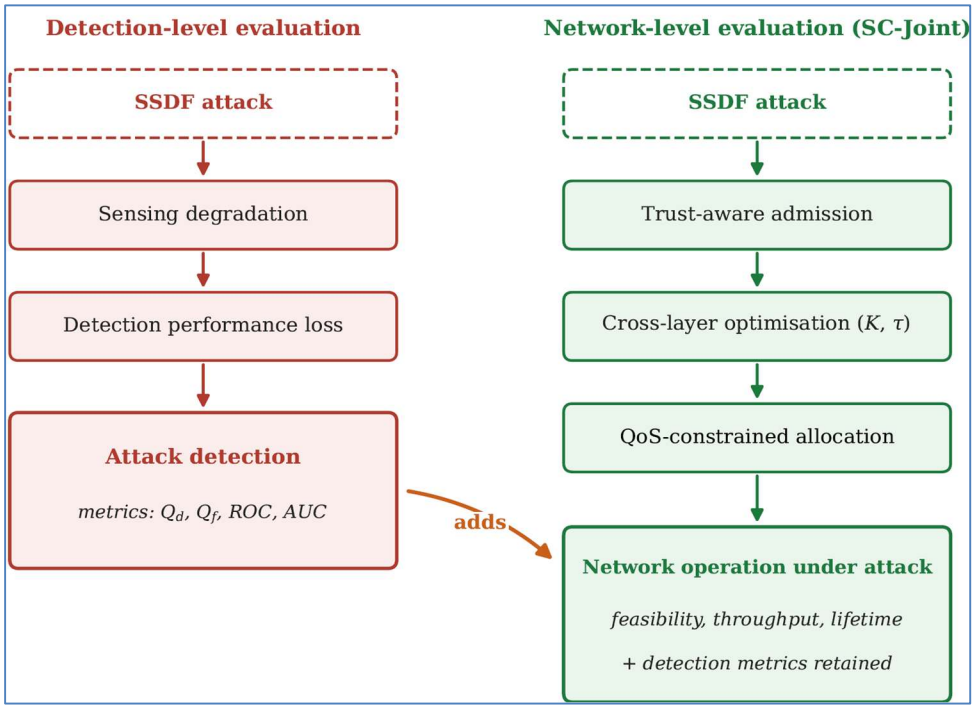

Figure 1. Detection-level versus network-level evaluation of SSDF defenses. SC-Joint adds trust-aware admission and SLA-constrained cross-layer optimization, and evaluates resilience by feasibility, throughput and lifetime in addition to detection metrics.

2. Related work

A natural defense against SSDF is to discard falsified reports and fuse only those from reliable SUs. We review prior work in five strands, noting for each what it optimizes and where it stops short of a network-level, security-constrained formulation.

2.1 Cooperative spectrum sensing and the sensing–throughput trade-off

Energy detection [4] remains the most common local-sensing technique owing to its low complexity and its independence from prior knowledge of the PU signal. Cooperative diversity across SUs mitigates fading and the hidden-terminal effect [7]. Hard-decision fusion under the Chair–Varshney optimal rule [8] and its OR/AND/majority special cases provide a practical FC architecture. Liang et al. [9] formalized the sensing–throughput trade-off, showing that the sensing time must be chosen to balance protection of the PU against the residual time available for transmission, and Peh et al. [10] extended this to the optimization of cooperative sensing. Our cross-layer optimizer adopts this trade-off but extends the decision space to include the admitted cooperating set under security constraints.

2.2 SSDF / Byzantine attacks and defense

Byzantine falsification of 1-bit local decisions sent to a fusion center has been analyzed in [33], [34], which characterize the attacker fraction that blinds the fusion center and the corresponding optimal attack strategies; these works do not impose network-level QoS constraints. Chen et al. [5] proposed the weighted sequential probability ratio test (WSPRT) and compared it with a variety of data fusion techniques under various network operating conditions; their results indicate that WSPRT was the most robust against the Byzantine failure problem among the techniques considered. Parhizgar et al. [17] cluster SUs and classify each cluster as reliable or unreliable using a neural network and conventional classifiers under both deterministic and stochastic attack strategies. Li et al. [18] combine isolation forests with spectral clustering to identify malicious sensor nodes without labels and to adapt the defense to varying attack intensities. Xu et al. [15] target massive SSDF with a joint sensing-and-access framework in which each SU reconciles its cooperative and independent decisions before acting. Paul et al. [35] jointly handle cooperative sensing and device-to-device transmission in energy-harvesting CRNs, using reputation-based clustering and ensemble learning to isolate malicious SUs. Wu et al. [36] formulate a random Byzantine attack model and apply a sliding-window reputation within a sequential-sensing scheme, and Dixit et al. [23] propose a deep-reinforcement-learning routing paradigm to mitigate combined Byzantine faults and SSDF in next-generation CRNs. Classical robust-fusion work derives weights or likelihood ratios that down-weight suspect nodes [5], [13], and MAD-based outlier rejection discards reports that deviate strongly from the cohort [32]. These works mainly measure success at the sensing layer, and most keep the cooperating set and resource allocation fixed once attackers are flagged. SC-Joint instead combines established admission and allocation components in one formulation, in which admission constrains the joint choice of cooperating set and sensing time under detection, throughput and lifetime requirements.

2.3 Trust and reputation management

Reputation-based CSS assigns each SU a trust score that summarizes its past reliability [22], [37]. Tiwari et al. [19] use a reputation score directly as a classifier that separates genuine from malicious SUs, Mahvash et al. [38] decentralize trust with a blockchain-inspired proof-of-trust mechanism, Xie et al. [20] adapt the trust threshold through reinforcement learning, Chen et al. [39] adopt a “slow-growth, fast-recovery” update with regional management and majority voting, Dewangan et al. [21] couple passive reputation with a hybrid-optimized deep model on a blockchain, and Giri et al. [40] train an extreme-learning machine to classify legitimate versus malicious SUs. Beta-distributed reputation, updated through Bernoulli agreement/disagreement events, is attractive because it is lightweight, online, and probabilistically interpretable. A common methodological trap is to update trust with ground-truth labels. Doing so assumes the very detection ability that the system is supposed to provide. The mechanism proposed here is strictly oracle-free: trust is

updated only from agreement with the leave-one-out majority of the other SUs, keeping the contribution at the network layer rather than the sensing layer.

2.4 Cross-layer optimization and resource allocation

Biswas et al. [25] formulate buffer-aware user selection and resource allocation for an opportunistic CRN as a mixed-integer nonlinear program, jointly choosing the cooperating set, sensing thresholds, and resource parameters; this is the closest structural precedent for our admission-and-allocation decision variable. Sheela et al. [26] coordinate the physical, MAC, and network layers with reinforcement learning, Chen et al. [41] propose a latency-aware four-layer joint design with closed-form Pareto weighting, Latif et al. [27] pursue Pareto-optimal energy- and spectral-efficiency allocation in CR-IoT, and Gupta et al. [28] and Zhang et al. [42] apply reinforcement and graph-reinforcement learning to distributed resource and routing allocation. These works demonstrate the value of joint cross-layer decisions, but none embeds an SSDF trust constraint into the admission decision, so an attacker can continue to perturb the operating point even when the allocation is optimal.

2.5 Gap and positioning

Existing studies have addressed individual aspects of secure cognitive radio networking. SSDF mitigation has been investigated through machine learning, statistical detection, and Byzantine-defense mechanisms [5], [13], [14], [16], [17], [18], [23], [35], [37]. Trust- and reputation-based sensing frameworks have been explored in [19], [20], [21], [22], [37], [38], [39]. Cross-layer optimization of cognitive radio networks is the subject of [25], [26], [41], and QoS-aware resource allocation that of [25], [27], [28], [42]. For the most part, though, these lines of work have developed separately. We are not aware of any prior work that brings SSDF-resilient trust management, admission of cooperating sensing nodes, sensing-time optimization and QoS-constrained operation together in one framework. SC-Joint is designed for that intersection. Its aim goes beyond detecting attacks: it seeks to keep the network operating feasibly and efficiently while it is under attack. Table 1 compares representative works on the five capabilities. None of the surveyed approaches has all five; SC-Joint does.

Table 1. Related work and research gap (✓ = capability present)

Work	SSDF	Trust mode	Layer opt.	QoS	Sec.- alloc.
SSDF detection [5], [13], [14]	✓	✗	✗	✗	✗
ML/Byzantine defense [15], [17], [18]	✓	✗	✗	✗	✗
Trust/reputation [19], [20], [21], [22], [37], [38], [39]	✓	✓	✗	✗	✗
Cross-layer CRN [26], [41]	✗	✗	✓	✓	✗
Buffer-aware allocation [25]	✗	✗	✓	✓	✗
QoS resource allocation [27], [28], [42]	✗	✗	✗	✓	✗
Proposed (SC-Joint)	✓	✓	✓	✓	✓

2.6 Security, adversarial robustness, and resilience

Another body of work looks at security and resilience from the learning side. Liu et al. [16] show that adversarial deep learning and data poisoning can degrade learned spectrum sensing, Lin et al. [29] enhance physical-layer security with a GAN-powered deep-reinforcement-learning policy, Taherpour et al. [30] provide Byzantine-resilient sensing through robust Bayesian belief updates, Tatar et al. [31] secure integrated sensing and communication via a Stackelberg game, Ferrag et al. [43] survey edge-learning vulnerabilities for 6G-enabled IoT, and Ezhilarasi et al. [44] survey CRN attack mitigation via machine learning and blockchain. These works frame resilience as robust learning or game-theoretic defense, often at the physical or learning layer. We instead

express resilience as a measurable network-layer property, namely the preservation of throughput and admission feasibility under an increasing attacker fraction, and we prove and quantify it directly.

3. System model

We consider a centralized CRN with N = 12 secondary users indexed by $\mathcal{N} = \{1, \ldots, N\}$, a single fusion center (FC), and one PU occupying a licensed channel. Time is slotted into frames of duration T, each partitioned into a sensing sub-slot of duration τ and a transmission sub-slot of duration T − τ, following the sensing–throughput trade-off. Each SU performs local energy detection and reports a binary decision to the FC, which combines the reports under the OR rule. Under an SSDF attack, compromised SUs inject falsified (always-yes) decisions into the fused result, as illustrated in Figure 2, inflating the global false-alarm probability and degrading network-layer operation. The trust-aware architecture that counters this attack is developed in Section 4 (Figure 3).

3.1 Network and sensing model

Each SU performs energy detection over the sensing window [4]. The local binary hypotheses are H_0 (PU absent) and H_1 (PU present): aligned with the real-valued energy-detector normalization actually implemented (variance 2u), with the threshold set under a CFAR criterion. The number of samples is $u = \tau f_s$ and the noise power is normalized to $\sigma^2 = 1$.

$$P_{f,i} = Q\left(\frac{\lambda_i - u}{\sqrt{2u}}\right), \qquad u = \tau f_s \tag{1}$$

$$P_{d,i} = Q\left(\frac{\lambda_i - u\,(1+\gamma_i)}{\sqrt{2u\,(1+2\gamma_i)}}\right) \tag{2}$$

where the CFAR threshold is $\lambda_i = u + \sqrt{2u}\,Q^{-1}\left(P_f^{loc}\right)$, $Q(\cdot)$ is the Gaussian tail function, and γ_i is the received PU SNR at SU_i.

The real-valued energy-detector normalization (variance 2u) is used throughout, with the threshold set under a CFAR criterion.

3.2 Cooperative fusion and global operating point

Let $\mathcal{S} \subseteq \mathcal{N}$ denote the admitted (trusted) cooperating set with $K = |\mathcal{S}|$. Under the OR rule, the FC declares the channel busy if any admitted SU reports a 1. For independent honest reports, the global detection and false-alarm probabilities are:

$$Q_d = 1 - \prod_{i\in\mathcal{S}}(1 - P_{d,i}), \quad Q_f = 1 - \prod_{i\in\mathcal{S}}(1 - P_{f,i}). \tag{3}$$

OR rule is the baseline because it is commonly used and its monotonic behavior makes the effect of a falsified always-yes reports easily analyzable (Section 5).

3.3 Throughput model and cross-layer coupling

Following the sensing–throughput framework of Liang et al. [9], the expected secondary throughput per frame, when the network transmits only on a sensed-idle channel, is

$$R(\mathcal{S}, \tau) = \frac{T - \tau}{T}\,[\,P(H_0)(1 - Q_f)C_0\,] \tag{4}$$

where $C_0 = \log_2(1 + \gamma_s) = 3.46$ Mb/s is the secondary link capacity on a truly idle channel (γ_s = 10 dB), and $P(H_0) = 0.70$, $P(H_1) = 0.30$, are the occupancy priors, with R expressed in Mb/s for W = 1 MHz. The residual capacity during a missed detection is neglected: under the PU-protection constraint $Q_d \geq 0.90$ such events are rare and the secondary rate then is negligible, so we set $C_1 = 0$ and throughput is governed entirely by the H_0 term, precisely the term an always-yes attacker destroys by forcing $Q_f = 1$.

Lifetime model. Sensing costs Ps = 110 mW during the sensing sub-slot, plus E_{rep} = 25 μJ per report. Network lifetime is governed by the first SU to deplete:

$$T_{\text{life}}(\mathcal{S}, \tau) = \frac{\min_{i \in \mathcal{S}} \mathcal{E}_i}{Ps \cdot \tau + E_{rep}} \quad [frames] \tag{5}$$

T_{life} is measured in sensing frames; with T = 100 ms, T_{req} = 1500 frames correspond to 150 s. The total energy per round reported in Figs. 5–6 is $E_{round} = K \cdot (Ps \cdot \tau + E_{rep}) + E_{tx}$.

3.4 SSDF attack model

We model malicious SUs as Byzantine nodes that report falsified decisions $\sim d_i^{(t)}$ instead of their true local decision $\sim d_i^{(t)}$. Let $\mathcal{M} \subseteq \mathcal{N}$ be the malicious subset with attacker fraction $\beta = |\mathcal{M}|/N$. Five adversaries are used, each matched to the quantity the corresponding experiment measures (Figure 2):

- **Always-yes (constant).** $\sim d_i^{(t)} = 1$ regardless of observation. This maximally inflates Q_f under OR fusion and is the relevant adversary for throughput and feasibility, since it attacks the network-layer rate directly (Proposition 1). Used in Scenarios 1–6.
- **Probabilistic flip (stealthy).** The SU flips its true decision, $\sim d_i^{(t)} = 1 - d_i^{(t)}$, with falsification probability p. This adversary is required for detection/AUC analysis: a constant attacker cannot move a rank-based AUC (Mann–Whitney statistic), whereas a probabilistic flip degrades separability. We use it in Scenarios 7–9, with p = 0.5 in Scenario 7, p = 0.9 in Scenario 8 and a sweep over p in Scenario 9.
- **On-off (duty-cycled).** Per frame, the SU attacks with probability δ = 0.5 (reporting $\breve{u}_i = 1$) and otherwise reports truthfully; phases are *i.i.d.* across frames and attackers. It checks whether falsifying only some of the time dilutes the reputation evidence that has accumulated.
- **Adaptive (consensus-mimicking).** The SU reports the observed leave-one-out majority with probability $1 - f$ and its complement with probability $f = 0.20$, maximizing reputation by conforming and deviating only occasionally.
- **Two-phase attacker.** The SU reports truthfully for the first 70% of frames, accumulating reputation, then switches to always-yes for the remainder. This adversary targets reputations that weight all frames equally.

The last three are used only in the reputation-robustness study (Table 6, M = 60 topologies per model–β pair, β extended to 0.42 to probe the majority boundary). Attackers are assumed to know the OR fusion rule and $P(H_0) = 0.70$, $P(H_1) = 0.30$ (idle-dominant occupancy) channel statistics, but not the reputation state of other nodes; unless stated otherwise, they act independently.

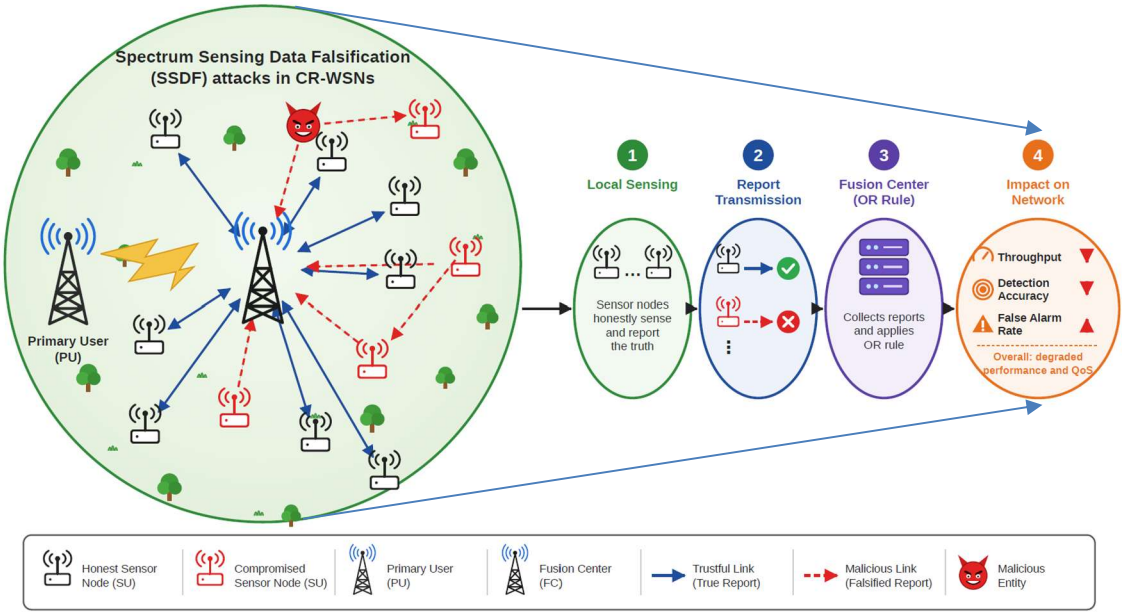

Figure 2. SSDF attack model: honest SUs report 1-bit decisions while malicious SUs inject falsified reports (always-yes, or probabilistic flip with probability p) into OR fusion, biasing global decision and inflating Q_f.

4. SC-Joint: security-constrained cross-layer framework for resilient operation

Fig. 3 shows the two stages of SC-Joint. In the first, an oracle-free trust-admission layer gives each SU a Beta–Bernoulli reputation and admits the SUs with $r_i \geq r_{th}$. In the second, a security-constrained cross-layer optimizer chooses the cooperating-set size K and the sensing time τ. The optimizer may pick the cooperating set only from the pool $\mathcal{A}$ produced by admission. What we want is a network that stays within its SLA during an attack; detecting the attackers is not, by itself, the goal. The trust layer excludes SUs whose reports disagree persistently with the consensus, and the optimizer then chooses K and τ over the admitted SUs to maximize throughput within the QoS-SLA region.

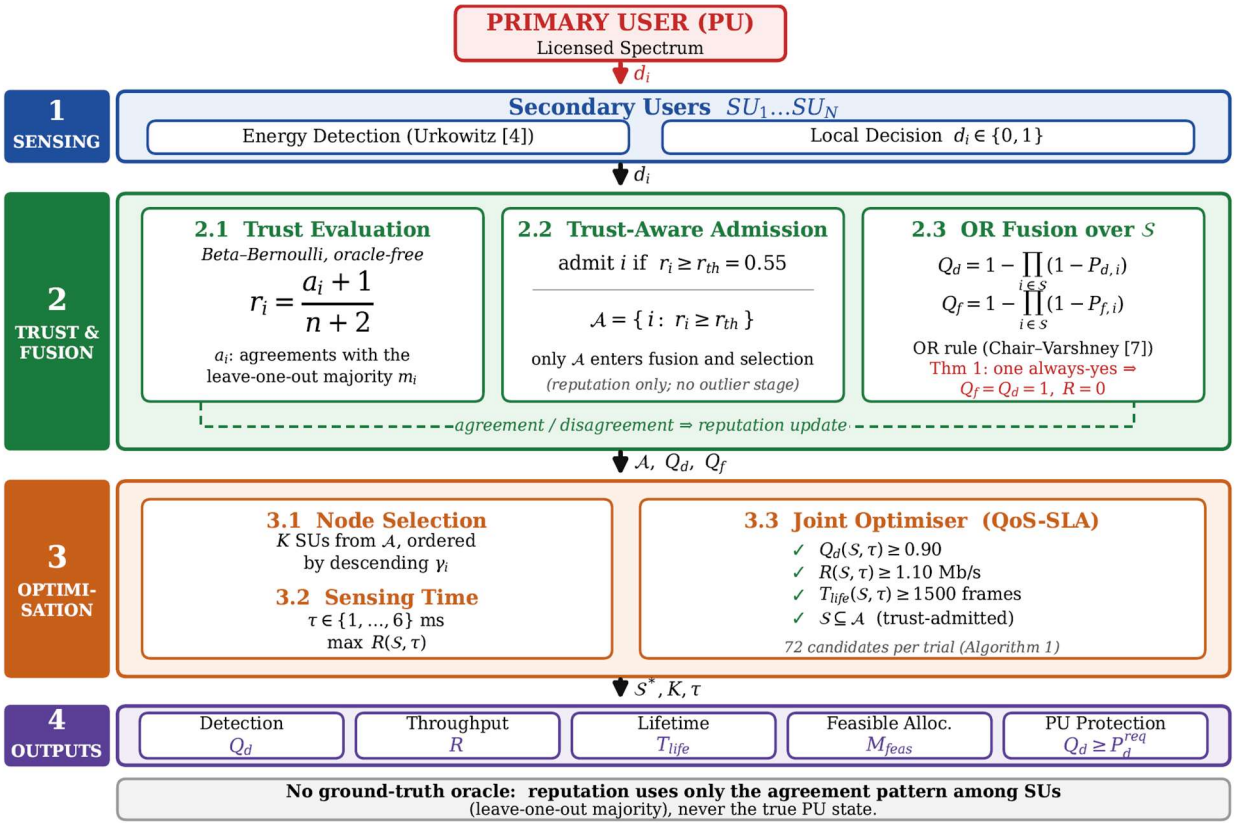

Figure 3. SC-Joint architecture for SSDF-resilient cognitive radio networks

4.1 Data-driven Beta–Bernoulli reputation

After each frame the FC compares SU_i's report $d_i^{(t)}$ against the **leave-one-out majority** m_i of the other SUs, not the fused decision D, which is itself contaminated by unfiltered reports.

For frame *t,* the leave-one-out majority seen by SU_i is

$$m_i^{(t)} = \mathbb{1}\left[\sum_{j \neq i} d_j^{(t)} > \frac{N-1}{2}\right]. \qquad (6)$$

At the evaluated operating point $P(m = 1 \mid H_0) \approx 0.001$ and $P(m = 1 \mid H_1) \approx 0.23$, so $r_i \approx 1 - \rho_i$ and admission reduces to a report-rate threshold $\rho_i \leq 1 - r_{th} = 0.45$; a report-rate-threshold baseline admits the identical set in every topology (Section 7.13).

Over n frames, let $a_i = \sum_{t=1}^{n} \mathbb{1}\left[d_i^{(t)} = m_i^{(t)}\right]$ be the agreement count. Under an uninformative Beta(1,1) prior, $a_i = (n - a_i) + 1$, so the posterior-mean reputation and the admission rule are [45]

$$r_i = \frac{a_i + 1}{n + 2} \in [0,1], \qquad i \in \mathcal{S} \Leftrightarrow r_i \geq r_{th} (= 0.55). \qquad (7)$$

Only the pattern of agreement among SUs enters the update; the true PU state never does. The mechanism is oracle-free in this sense, since it needs no labels. A reputation that looked at ground truth could not be built in a real deployment, and it would pull the problem back into a sensing-detection framing. What the reputation measures. All frames are weighted equally (no forgetting factor). More importantly, at the Table-2 operating point (τ = 4 ms, P_f^{loc}= 0.06) per-node detection probabilities span ≈0.07 at −33.7 dB to ≈0.99 at −14.2 dB, so most SUs are weakly informative and the leave-one-out majority is idle-dominated: P (m = 1 | H_0) ≈ 0.00 but P (m = 1 | H_1) ≈ 0.23.

What the reputation actually rewards, then, is agreement with an idle-dominated consensus, not correctness against ground truth. This has two effects, which Section 7.12 quantifies. An always-yes attacker stands out, because reporting 1 in every frame is far from that consensus. An honest high-SNR node, on the other hand, is penalized as a dissenter whenever it correctly reports the PU. Within the honest cohort, the SNR–reputation correlation is −0.70 (highest-SNR quartile 0.737 vs lowest 0.888), which is why the honest minimum (0.646) lies closer to r_{th}than the honest mean (0.80). Separation is thus conditional on idle-dominant occupancy; Section 7 reports its sensitivity to P(H_1).

4.2 Outlier/MAD baseline

The Outlier/MAD baseline [32] screens each SU by its report rate over a window of n sensing periods (n per scenario in Table 2):

$$\rho_i = \left(\frac{1}{n}\right)\Sigma_t d_{i(t)} \qquad (8a)$$

With $m = \text{medj}(\rho_j)$ and MAD = $\text{med}_i(|\ \rho_i - m\ |)$, the Outlier/MAD baseline admits:

$$\mathcal{O}\ = \{\, i : |\ \rho_i - m\ | \leq \kappa \cdot max(MAD, \epsilon)\}\ \ \epsilon = 10^{-9} \quad (8b)$$

Here, MAD is the median absolute deviation, a 50%-breakdown scale estimator [46].

4.3 Trust-aware admission and the cross-layer problem

SC-Joint admits SUs by reputation alone:

$$\text{A} = \{\text{i} \in \text{N}: r_i \geq r_{th}\,\},$$

The optimizer then selects the cooperating set $\mathcal{S} \subseteq \mathcal{A}$ and the sensing time τ:

$$\begin{aligned} & max_{\mathcal{S}\subseteq A,\tau}\ \ R(\mathcal{S},\tau) \\ s.t.\ \ & Q_d(\mathcal{S},\tau) \geq\ P_{d,min}\ =\ 0.90\,, \\ & R(\mathcal{S},\tau) \geq\ R_{min} = 1.10, \\ & T_{life}(\mathcal{S},\tau) \geq\ T_{req} = 1500s, \quad \tau\ \in \{1,\ldots.,6\}\, ms \end{aligned} \qquad (9)$$

$\mathcal{A}$ is the fixed output of the trust rule, and $\mathcal{S}$ is the decision variable drawn from it, so the formulation is not circular. Problem (9) is a mixed-integer nonlinear program, structurally analogous to the cross-layer user-selection problem of [25]. Instead of enumerating all $2^N \times 6 = 24{,}576$ configurations, we use the monotonicity of the OR rule. We order $\mathcal{A}$ by decreasing SNR and search nested prefixes $\mathcal{S}_K$ = {first K} jointly with τ, which gives 6 × N = 72 candidates per trial.

Optimality gap. Against exhaustive enumeration on 40 realizations, the nested search never disagreed on feasibility (0/40). The mean throughput gap was 0.421%: zero in 36 trials, with a worst case of 14.3% in one realization whose optimal set was not SNR-nested. The search is 289× faster (0.75 ms vs 218 ms per trial). We therefore describe it as feasibility-exact and throughput-near-optimal on the evaluated instances. N is not varied, so no scalability claim is made. Security enters only through $\mathcal{A}$ and through the effect of attacker reports on Qd, Qf and the feasible region. Reputation is computed over the same frames used for evaluation; a disjoint training window gives the same feasibility (88.0–89.7%, Section 7.13).

4.4 Compared schemes

SC-Joint is compared with five schemes and three illustrative allocation strategies:

- **NaïveOR:** OR fusion over all SUs with no trust layer; the undefended reference.
- **Outlier/MAD [32]:** admission by the MAD test of Section 4.2, followed by the same re-optimization of K and τ as SC-Joint, so the two differ only in admission.
- **Decayed Beta–Bernoulli reputation (DBR):** frame t is weighted by $\omega^{(n-t)}$, with an effective memory of ≈1/ (1− ω) frames. ω = 1 recovers SC-Joint, and we evaluate ω ∈ {0.99, 0.95} on identical topologies and seeds.
- **Throughput- and Energy-oriented variants:** the first maximizes R ignoring lifetime; the second minimizes energy subject to detection.
- **Fixed, Greedy and Random:** illustrative allocation strategies, not security comparators. Greedy gives the highest raw throughput (2.25 Mb/s), yet it breaks the detection SLA (Q_d = 0.66 < 0.90). We show them in Figures 5 and 6 only and leave them out of every resilience comparison.

The effects of trust admission, re-optimization and fusion are separated in the factorial ablation of Table 5.

5. Analytical results

We give three results in this section. Propositions 1–2 follow directly from the OR rule, and we present them as design observations. Proposition 3 is the main analytical result. Proposition 1 shows that a single always-yes SU saturates the global false-alarm probability and drives secondary throughput to zero under OR fusion. This is a network-layer failure that detection metrics alone do not reveal. Proposition 2 shows that, under nested injection on a fixed realization, the set of SLA-feasible configurations of the undefended baseline can only shrink as attackers are added. Together, they indicate why SSDF resilience must be handled at the network-design level: attackers must be kept out of the cooperating set before fusion. Proposition 3 then states when the trust-admission rule of Eq. (7) achieves this. It bounds the probability of mis-admission and gives closed-form thresholds for attacker admission and honest-node exclusion.

5.1 Throughput collapse under Always-Yes SSDF (*Proposition 1*).

Proposition 1. *Consider OR fusion over an admitted set $\mathcal{S}$ containing at least one always-yes attacker. Then*

$$Q_f = 1, Q_d = 1, R(S, \tau) = 0. \quad (10)$$

Proof sketch. Under the OR rule (3), the fused decision is D = 1 whenever any admitted SU reports 1. An always-yes attacker a ∈ $\mathcal{S}$ reports $\tilde{u}_a = 1$ in every frame, so D = 1 in every frame, independently of the true hypothesis. Hence:

$$Q_f = Pr(D = 1 \mid H0) = 1, \qquad Q_d = Pr(D = 1 \mid H1) = 1$$

Substituting into the throughput model (4), both bracketed terms vanish, since $(1 - Q_f) = 0$ and $(1 - Q_d) = 0$. Therefore R ($\mathcal{S}$, τ) = 0.

Remark. Nothing is left of the throughput. Under both hypotheses the fusion center sees the channel as busy all the time, and the secondary network never gets to transmit. The coincident Q_d = 1 means nothing here: it comes from reporting 1 all the time rather than from actual sensing, and it brings no throughput. Figure 6(a) shows the same thing in simulation, where NaïveOR throughput drops to 0.00 Mb/s as soon as one attacker is admitted.

Proposition 1 shows why the damage belongs to the network layer. In fact, a single attacker not only deteriorates the detection metric but also completely. Furthermore, it isolates the secondary network from idle spectrum and renders its throughput zero. That is the main justification of this paper for identifying SSDF as a network-design

issue. The goal of network resilience is mostly to maintain network operation feasible, and only secondarily a better detection statistics.

5.2 Monotone feasibility degradation (*Proposition 2*)

Consider a fixed network realization and a universe of candidate configurations $\mathcal{C}$. For a nested sequence of always-yes attacker sets $\mathcal{M}_0 \subseteq \mathcal{M}_1 \subseteq \cdots \subseteq \mathcal{M}_k$under OR fusion without trust filtering, let $F(\mathcal{M}) \subseteq \mathcal{C}$denote the configurations that meet every SLA constraint (detection, throughput and lifetime) when the admitted attacker set is A.

Proposition 2. $F(\mathcal{M}_{k+1}) \subseteq F(\mathcal{M}_k)$, and therefore $|F(\mathcal{M}_{k+1})| \leq |F(\mathcal{M}_k)|$: the feasible-configuration count is monotonically non-increasing as attackers are added by nested injection.

Proof. Let $c \in F(\mathcal{M}_{k+1})$. If the cooperating set of $c$included even one attacker, that always-yes node would drive the global false-alarm probability $Q_f \to 1$(Proposition 1), collapsing throughput and violating $R \geq R_{\min}$, contradicting $c \in F(\mathcal{M}_{k+1})$. Hence every feasible configuration is attacker-free, and its feasibility depends only on its honest members, which are unaffected by the attacker set. Since such a $c$meets every SLA constraint whatever $\mathcal{M}$ is, it remains feasible under $\mathcal{M}k$ as well, giving $F(\mathcal{M}_{k+1}) \subseteq F(\mathcal{M}_k)$. Note that Proposition 2 is about the unprotected baseline.

There, a growing attacker fraction can only remove feasible operating points, so the feasibility rate falls monotonically; Section 7 confirms and quantifies this. SC-Joint removes the premise of the proposition. Its trust filter takes attackers out before fusion, so admitted configurations stay attacker-free and $Q_f$remains unsaturated; the feasible region is therefore largely preserved. Monotone degradation is a property of the undefended system. We check it in an independent simulation as a test of the mechanism, not as a target the model was tuned to.

5.3 Admission analysis (*Proposition 3*)

Proposition 3. (Admission condition). Let frames be i.i.d., let p_i= P (d_i = 1) be the report rate of SU, and let u_i = P (mi = 1) be the probability that the leave-one-out majority reports 1. The per-frame agreement probability a_i = P ($d_i = m_i$) satisfies.

$$1 - \rho_i - \mu_i \leq da_i \leq 1 - \rho_i + \mu_i \qquad (11a)$$

Since r_i= (A_i + 1)/ (n + 2) with $A_i \sim$ Bin (n, a_i), Hoeffding's inequality gives, for $a_i < r_{th}$,

$$P(r_i \geq r_{th}) \leq exp(-2n\,(r_{th} - a_i)^2) \qquad (11b)$$

and symmetrically P ($r_i < r_{th}$) ≤ exp (−2n ($a_i - r_{th}$)²) for $a_i > r_{th}$.

1. *Flip attacker.* With honest report rate ρ_0, a flip attacker has $\rho_a = \rho_0 + (1 - 2\rho_0)$ p. Since $\mu \approx 0$, it is admitted with high probability if and only if

$$p* = (1 - r_{th} - \rho_0)/(1 - 2\rho_0) \qquad (11c)$$

2. *Honest SU.* With p_i= (1 − π_1) $P_{f,}$, loc + π_1 P_d, i, a high-SNR SU (P_d, $i \to 1$) is excluded with high probability when:

$$\pi_1* = (1 - r_{th} - P_{f,}loc)/(1 - P_{f,}loc) \qquad (11d)$$

Evaluation at the operating point. With r_{th} = 0.55, $P_{f,}$, loc = 0.06, $\rho_0 \approx 0.10$–0.14 and $\mu \approx 0.07$, Eq. (11c) gives $p^* \approx 0.43$–0.44 and Eq. (11d) gives $\pi_1^* \approx 0.42$. Both match the simulations: attackers are admitted for $p \leq 0.4$ (Fig. 13), and the highest-SNR honest SU is excluded in 0.3% of topologies at $P(H_1) = 0.40$ versus 36% at 0.45 (Section 7.13). With n = 2000 frames and a margin of 0.05, Eq. (11b) bounds the mis-admission probability by $e^{-10} \approx 4.5 \times 10^{-5}$.

6. Simulation setup

Experiments were implemented in MATLAB [24.1.0.2628055 (R2024a) Update 4] and also run in GNU Octave. A shared core library provides the system model, energy detector, reputation, Outlier/MAD baseline, fusion rules and AUC computation, and each scenario is a script built on it. All schemes are evaluated on identical realizations (common random numbers). Fading is block-static per topology. Scenario 5 adds attackers cumulatively to a fixed topology. Scenario 4 uses independent draws, so its curves need not be monotone. Reputation is computed over the evaluation frames; a disjoint training window gives the same feasibility (Section 7.13). For the AUC, $\boldsymbol{P_{f,}}$, loc is swept over 45 values in [0.001, 0.6] at τ = 4 ms. The admission rule is applied again at every value, and the OR-fused ($\boldsymbol{Q_f}$, $\boldsymbol{Q_d}$) points are integrated with the trapezoidal rule. As a reference we use an oracle that fuses only the honest SUs; it gives an upper bound but could not be deployed.

All intervals are 95%. Feasibility uses the Wilson score interval [47], since the normal approximation collapses to ±0 at 100%:

$$\left[\frac{\hat{p} + z^2/2M \mp z\sqrt{\hat{p}(1-\hat{p})/M + z^2/4M^2}}{1 + z^2/M}\right], \hat{p} = \frac{k}{M},\ z = 1.960 \qquad (12)$$

AUC and throughput (averaged over feasible trials) use Student-t intervals [48], and scheme contrasts use paired t-intervals. The attacker fraction is β ∈ {0, 0.08, …, 0.33} (k = 0–4 of N = 12), extended to 0.42 in Table 6 to probe the majority boundary. We use M = 300 topologies in Scenarios 1–7 and 9 and in the robustness checks, and M = 60 in Scenario 8 and the ablation. A sample topology with three always-yes attackers (β = 0.25) is drawn in Figure 4. For each reported value, Supplementary Table S1 gives the section, figure or table where it appears. The main simulation parameters are listed in Table 2.

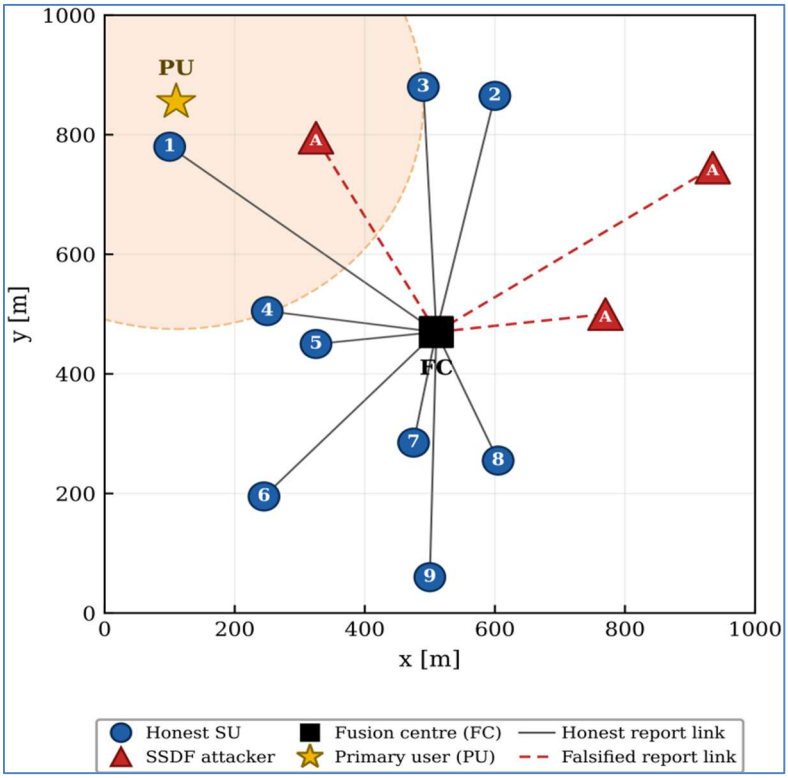

Figure 4. Simulation topology (N = 12). Nine honest secondary users (blue circles) and three SSDF always-yes attackers (red triangles; attacker fraction β = 0.25) send reports to the fusion center (FC) while the primary user (PU) holds the licensed channel. Grey solid links carry honest reports, dashed red links carry falsified reports.

Table 2 shows principal simulation parameters.

Table 2. Principal simulation parameters

Parameter	Value
Network and PHY	
Number of SUs (N) / PUs / channels	12 / 1 / 1
Propagation; mean SNR distribution	Rayleigh, block-static per topology; U [−24, −16] dB (U [−22, −14] dB in Sc. 8 and ablation)
Realized SNR range (5th–95th pct.)	−33.7 to −14.2 dB
Sampling frequency f_s; frame duration T	6 MHz; 100 ms
Sensing-time grid τ; fixed τ for ROC	1–6 ms; 4 ms
Channel bandwidth W	1 MHz
Fusion rule	OR (majority in robustness checks)
Occupancy and throughput	
Occupancy priors $P(H_0)$ / $P(H_1)$	0.70 / 0.30
Secondary-link SNR γ_s	10 dB
Idle capacity $C_0 = \log_2(1+\gamma_s)$	3.46 bit/s/Hz (3.46 Mb/s at W = 1 MHz)
Missed-detection capacity C_1	0 (neglected)
QoS and lifetime constraints	
$P_{d,min}/R_{min}$	0.90 / 1.10 Mb/s
Local false alarm $P_{f,}$, loc	0.06
Required lifetime T_{req}	1500 frames (150 s at T = 100 ms)
Sensing power P_s, report energy E_{rep}; FC energy E_{tx}	110 mW; 25 μJ; 50 μJ
Battery energy $\mathcal{E}_i$	U [0.6, 1.1] J
Trust layer and baselines	
Reputation model; threshold (r_{th})	Beta–Bernoulli (leave-one-out majority); 0.55
Forgetting factor ω: SC-Joint / DBR	1.0 / 0.99, 0.95
MAD coefficient κ (Outlier/MAD baseline only) [46]	3.0 (2.5 in Sc. 8)
Attack models	
Always-yes	Sc. 1–6
Probabilistic flip, p	0.5 (Sc. 6(b), 7), 0.9 (Sc. 8), swept 0.1–0.9 (Sc. 9)
Always-no	robustness checks (k = 1–3)
On-off duty δ; adaptive f; two-phase switch	0.5; 0.20; at 70% of frames
Attacker fraction β; count k	0–0.33 (0.42 in robustness); 0–4
Monte-Carlo protocol and statistics	
Topologies M	300 (Sc. 1–7, 9; robustness checks) / 60 (Sc. 8, ablation)
Observation window n (frames per trial)	1500 (ROC/AUC); 2000 (network outcome); 2500 (Sc. 7)
Paired common random numbers	Yes (shared topologies, noise and attacker draws)
Feasibility interval	Wilson score, 95%
AUC and throughput intervals	Student-t, 95%; paired t-intervals for scheme contrasts
Reported precision	1 decimal (%); 3 decimals (Mb/s, AUC, reputation)

7. Results and discussion

We discuss the results scenario by scenario. The nine scenarios start from the attack-free cross-layer baseline and end with network-level robustness against several attackers. Three schemes recur: SC-Joint (proposed), Outlier/MAD (a published robust-fusion baseline), and NaïveOR (undefended OR fusion). Greedy, Fixed, Random, and rate-/energy-constrained variants appear where instructive.

7.2 Scenario 1/ Cross-layer optimization without attack

In the attack-free baseline (Figure 5), Greedy attains the highest raw throughput (2.25 Mb/s) but operates at $Q_d = 0.66$, failing the detection SLA ($P_{d,min} = 0.90$) and endangering the PU. This is the failure mode that SC-Joint is designed to avoid: Greedy's operating point is infeasible, so its low energy is irrelevant. Among SLA-compliant schemes, SC-Joint delivers the highest throughput (2.13 Mb/s) at 100% feasibility while consuming the lowest energy (≈0.70 mJ/round), ahead of the Rate-Constrained variant (1.81 Mb/s) and well above Fixed (1.60 Mb/s, 2.84 mJ).

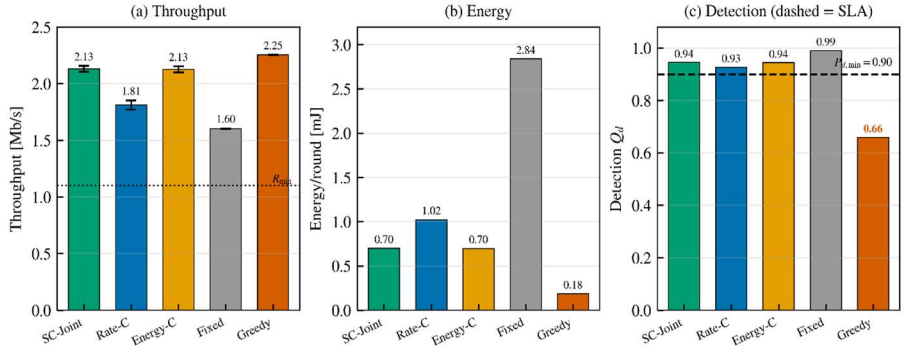

Figure 5. Scenario 1: attack-free baseline per scheme (M = 300). (a) Throughput, (b) energy per round, (c) detection probability Q_d. Rate-C and Energy-C are the throughput- and energy-constrained variants.

The Energy-Constrained variant coincides with SC-Joint here (2.13 Mb/s, 0.70 mJ). Greedy attains a higher raw throughput (2.25 Mb/s) at the lowest energy (0.18 mJ) but violates the detection SLA ($Q_d = 0.66 < 0.90$) and is therefore infeasible; SC-Joint's advantage lies in achieving the best feasible throughput–energy trade-off without breaching detection.

7.3 Scenario 2 / SSDF attack evaluation

A single always-yes attacker (Figure 6, β = 0.08) separates the schemes sharply. Only SC-Joint and the Energy variant exceed 90% feasibility (97% each); Greedy-R, Fixed and Random reach 21%, 13% and 9%, and NaïveOR 0% (Fig. 6(e)). Trust-aware admission seems effective in maintaining network viability in scenarios where unprotected schemes and detection-agnostic schemes fail to work.

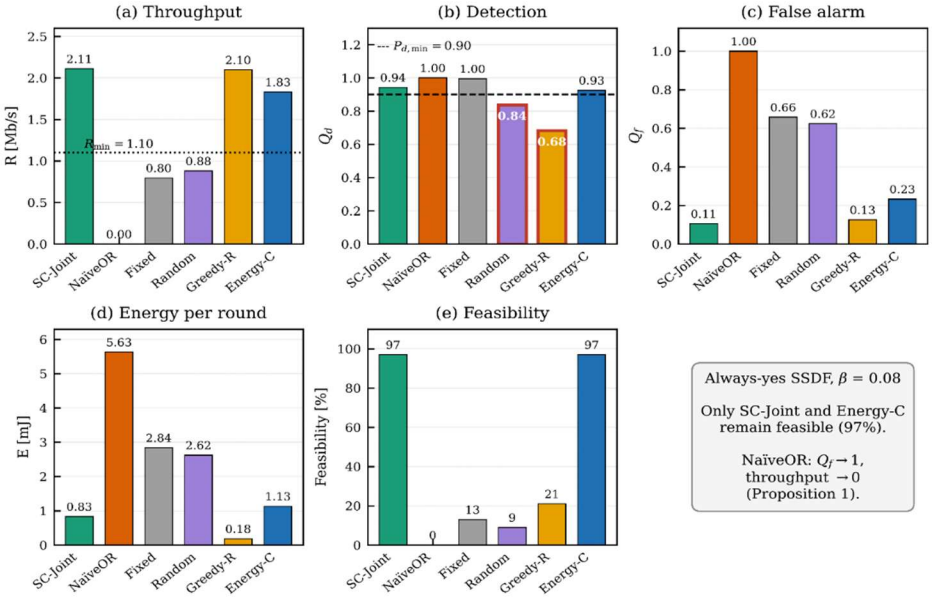

Figure 6. Scenario 2: one always-yes attacker (β = 0.08, M = 300). The figures correspond respectively to: (a) network throughput, (b) the probability of attacker detection, (c) the global false-alarm probability, (d) the energy per one round, (e) the rate of feasible assignments. Greedy-R algorithm maximizes R and doesn't consider lifetime, while Energy-C algorithm minimizes energy usage subject to the detection constraint only.

7.4 Scenario 3 / Trust reputation validation

In Figure 7, Honest nodes have stable reputations settling at around 0.80, while those of always-yes attackers are reduced almost to 0.12. Since the two distributions are placed on two sides of the 0.55 threshold, they don't overlap anymore, and a complete separation is achieved. With only a dishonest reputation at max of 0.316 less than 0.55 and a minimum of an honest reputation (0.646) higher, it's definitely a clear gap between two groups. Thus, when β = 0.25, all honest nodes remain while all attackers are rejected (100% accuracy in both cases). Naturally, with a separation margin of 0.330, any threshold lying within (0.316, 0.646) will provide the same outcome for admitting or denying the nodes. Therefore, the choice of the exact threshold value is not relevant here. The reputation convergence is very fast. Furthermore, the trusted set has converged even before the whole evaluation period was completed.

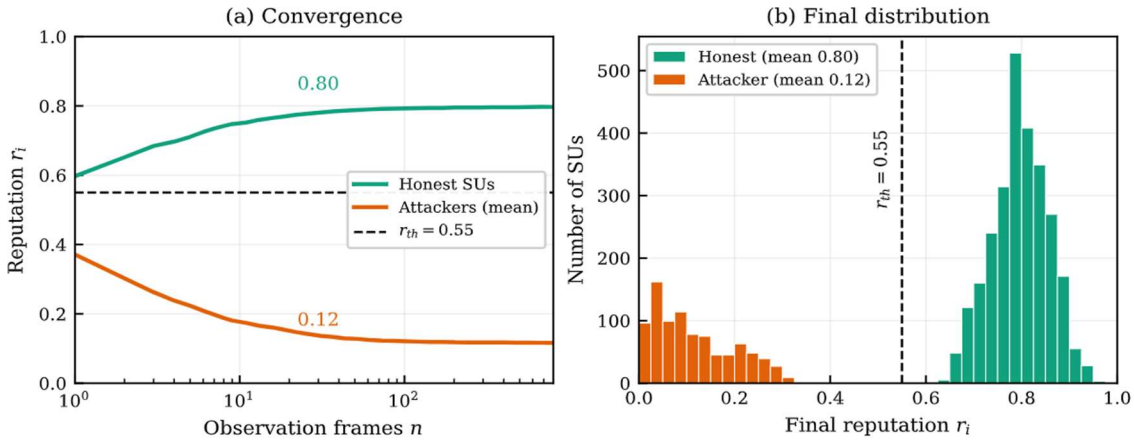

Figure 7. Leave-one-out reputation under always-yes SSDF (Scenario 3; N = 12, three attackers, M = 300). (a) Mean reputation versus observation frames. (b) Distribution of final reputations. The groups do not overlap. The lowest honest value (0.65) and the highest attacker value (0.32) fall on either side of = 0.55.

7.5 Scenarios 4 & 5 /Attacker-fraction analysis

Scenario 4 takes independent samples to cover the attacker fraction β ∈ {0, 0.08, 0.17, 0.25, 0.33} (Figure 8), but Scenario 5 simply adds 0 to 4 always-yes attackers on a fixed matched base topology thereby directly using Proposition 2 (Figure 9, Table 3). It is also a paired design which is essential since it allows the study of trust-aware admission independent of sampling noise through nested injection on a fixed topology while the independent per-condition samples may generate artificially non-monotonic curves.

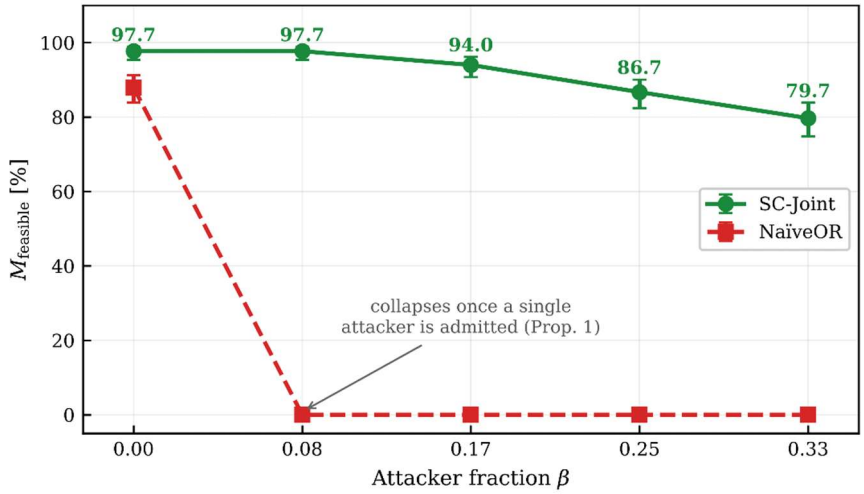

Figure 8. Scenario 4: feasibility against the attacker fraction β under always-yes SSDF, with topologies drawn independently for each β (M = 300, mean SNR U [−24, −16] dB, r_{th}= 0.55). Error bars: 95% Wilson intervals.

Scenario 4 (Figure 8) shows that β varies independently and differently under each condition as new sets of topologies are drawn each time (i.e., one condition draws a new "set" of β topologies while the other keeps the same β set). The results show feasibility decrease from 97.7% to 79.7%. For $\beta \leq 0.08$, the flat lines are the reflection of difference of sampling of different topologies. Scenario 5 (Figure 9, Table 3) is basically repeating the process from Scenario 4 but this time with paired CRNs and injecting in one CRN a set from a second CRN (nested injection). The new feasibility strictly decreases, going from 98% down to 86%, at 2.135–2.157 Mb/s. The two scenarios combined illustrate a methodological issue. It is possible to use independent draws to generate non-monotone plots; by changing that method and using paired injection, the artifacts get eliminated.

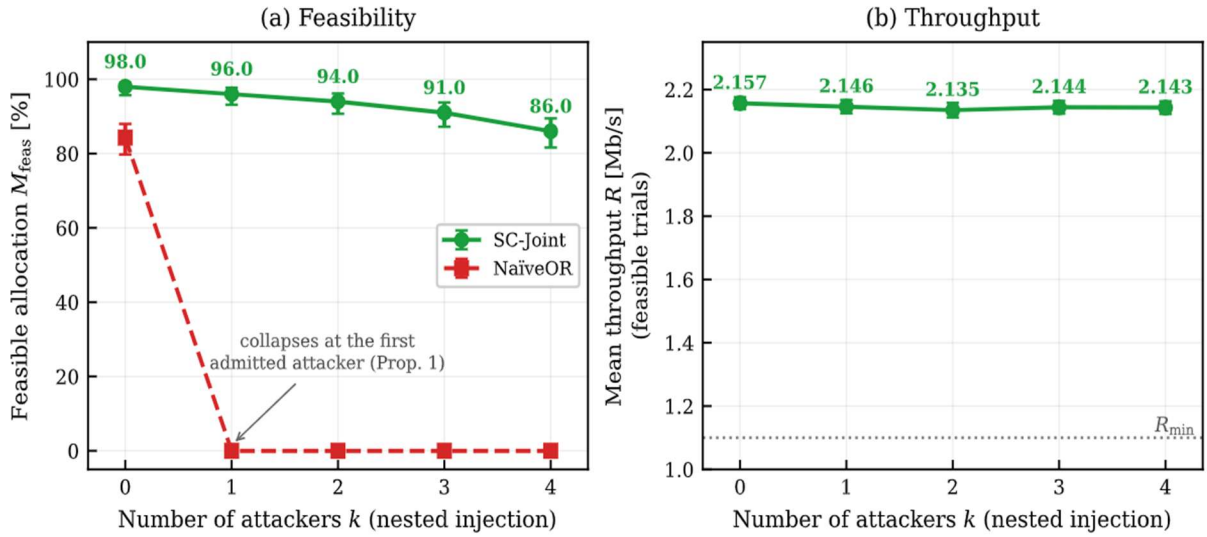

Figure 9. Scenario 5: nested injection of always-yes attackers on fixed topologies (M = 300; Wilson 95% intervals). (a) At the first admitted attacker Proposition 1 Naïve OR algorithm completely fails to give solutions and this check confirms an implementation correctness of Proposition 2 in the baseline no-defense case. SC-Joint's gradual degradation in feasibility is designed for such behavior and, in this case, because the admissible pool is a subset of the honest SUs only, is understandable. (b) SC-Joint throughput over feasible trials remains stable at 2.14–2.16 Mb/s.

Table 3 lists network feasibility for each attacker count in Scenario 5 (always-yes SSDF, paired CRN).

Table 3. Scenario 5/ Network feasibility vs attacker count (always-yes SSDF, paired CRN)

Attacker	SC-Joint Feas.	SC-Joint R	NaïveOR Feas.	Wilson [lo, hi]
0	98.0	2.157	84.3	[95.7, 99.1]
1	96.0	2.146	0.0	[93.1, 97.7]
2	94.0	2.135	0.0	[90.7, 96.2]
3	91.0	2.144	0.0	[87.2, 93.7]
4	86.0	2.143	0.0	[81.6, 89.5]

7.6 Scenario 6 / ROC and AUC Analysis

Under two always-yes attackers (Figure 10(a)), the ROCs of SC-Joint align completely with the oracle. The AUC for both is 0.959. The Outlier/MAD baseline attains 0.852 and undefended NaïveOR completely collapses to the chance level diagonal (0.500) as its fused decision saturates (Proposition 1)."

This shows again in the probabilistic flip (p = 0.5) that SC-Joint is essentially equally effective as the oracle (0.958 vs 0.959) while Outlier/MAD improves to 0.846. The undefended AUC does not hit the ceiling this time; it stays close to 0.6 (0.597; Figure 10(b)). Thus, the two attack models are indeed different experiments.

On the contrary, our major point is that through the system, faults can be endured more than detected, thus our novelty primarily is a fault-tolerant capability. That the almost perfect isolation here is only a case to illustrate how trust-aware admission can be applied is what makes this fact helpful, it is through admission that the system is kept running smoothly. AUC as a function of the number of attackers is given in Figures 11 and Table 4.

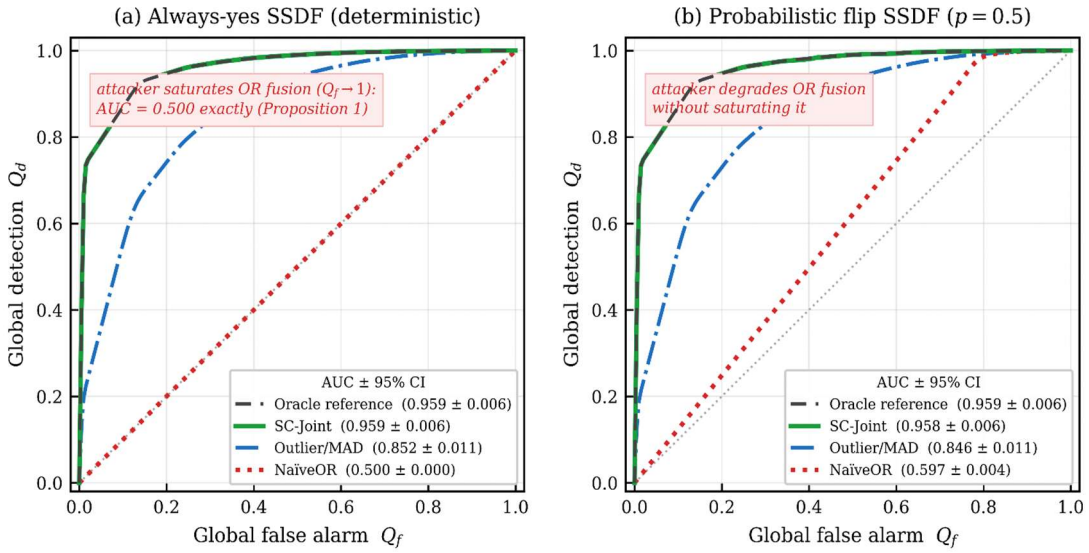

Figure 10. Scenario 6: cooperative ROC under two always-yes attackers (M = 300, τ = 4 ms). SC-Joint equals the oracle (AUC 0.959 vs 0.959), Outlier/MAD reaches 0.852, and NaïveOR falls to chance level (0.500), in line with Proposition 1. Results for the flip attack (p = 0.5), under which NaïveOR degrades more gradually, are given in Table 4.

7.7 Scenario 7/ AUC robustness versus attacker count

Figure 11 presents the results when the number of attackers increased from zero to four. SC-Joint consistently achieved above 0.94 AUC for all of them, and an oracle-like AUC for these attackers (0.965, 0.960, 0.960, 0.949 and 0.948), showing the proposed trust-aware admission mechanism does maintain the data separability even when the training data is heavily polluted. Outlier/MAD behaves non-monotonically. It climbs from 0.793 with no attacker to 0.866 with three, then drops to 0.821 with four. NaïveOR slides toward chance (0.965 → 0.523), which is how the feasibility collapse of Proposition 1 shows up at the detection layer. The proposed SCJoint method achieves better results against all attacker combinations than the baselines. See Table 4 for all numerical results.

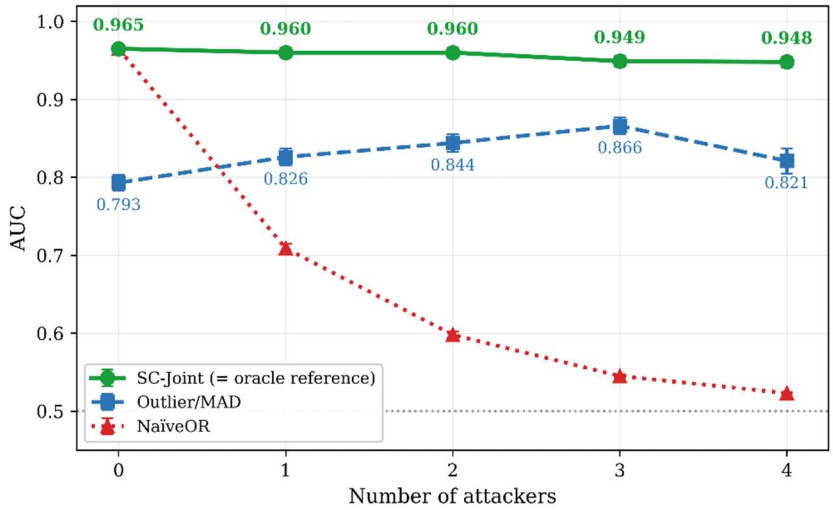

Figure 11. Scenario 7: AUC versus number of attackers under probabilistic flip (p = 0.5, M = 300, κ = 3.0; trapezoidal AUC of the OR-fused ROC). SC-Joint coincides with the oracle and stays above 0.94 (0.965 → 0.948); Outlier/MAD ranges from 0.793 to 0.866; NaïveOR decays from 0.965 to 0.523.

Table 4. Detection AUC versus number of attackers, mean [95% CI] (probabilistic flip, p = 0.5; M = 300; κ = 3.0; mean SNR U [−24, −16] dB; trapezoidal AUC of the OR-fused ROC)

Scheme	k = 0	k = 1	k = 2	k = 3	k = 4
Oracle reference	0.965 [0.005]	0.960 [0.006]	0.960 [0.006]	0.949 [0.007]	0.948 [0.007]
SC-Joint	0.965 [0.005]	0.960 [0.006]	0.960 [0.006]	0.949 [0.007]	0.948 [0.007]
Outlier/MAD	0.793 [0.010]	0.826 [0.011]	0.844 [0.011]	0.866 [0.011]	0.821 [0.016]
NaïveOR	0.965 [0.005]	0.709 [0.006]	0.598 [0.004]	0.545 [0.002]	0.523 [0.001]

7.8 Scenario 8/ Network-level feasibility and throughput

On the network-level terms that matter to a service operator, SC-Joint clearly outperforms the Outlier/MAD baseline. With the probabilistic-flip attacker (p = 0.90), SC-Joint keeps feasibility high for zero to four attackers (100, 100, 100, 98.3, 96.7%; Figure 12) at 2.21–2.24 Mb/s. For any number of attackers, the outlier/MAD metric oscillates between 66.7% and 81.7%, thus is not close to any of the attacker's scores. We assume the attacker submits false reports randomly, and this is more a robustness stress test than monotonicity. The study of Proposition 2 is included in Scenarios 4 and 5. The main contribution of this work in the network-level scenario is the fact that our mechanism can be applied under conditions where a strong baseline report-rejection mechanism fails to operate.

Scenario 8 is more feasible than Scenario 5 (98-86%) and their adversaries are nominally harder to overcome. It is attributed to environment. Scenario 8 has 60 topologies as opposed to 300 for Scenario 5, however, attack model was unable to account for all of this difference. When control channel SNR and number of attackers are held constant, always yes and probabilistic flip are as feasible (85.0% and 96.7% respectively at -20 dB, and 96.4% and 99.0% respectively at -18 dB). They are excluded by trust layer equally well, and the amount that the honest nodes can detect will be the determining factor. It is attributed to SNR and accounts for 11.7 points. Comparability Scenarios 5 and 8 are not comparable because of environment difference. Both are reported in Supplementary Table S1.

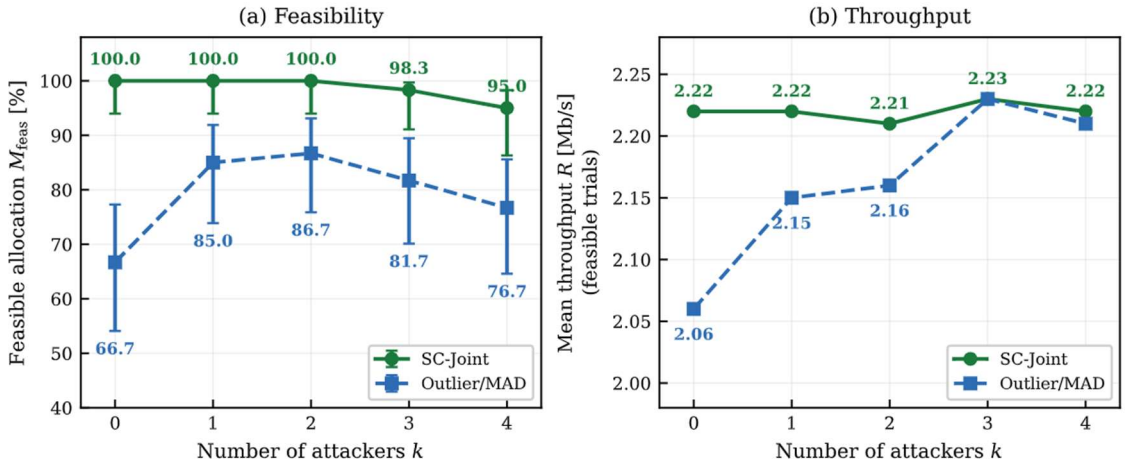

Figure 12. Scenario 8: network-level stress test under the probabilistic-flip attacker (p = 0.9, M = 60; Wilson 95% intervals). (a) Feasibility; (b) throughput over feasible trials.

7.9 Scenario 9 / Detection valley

Sweeping the falsification probability p exposes a detection valley (Figure 13). For p ≤ 0.4, the attackers' disagreement with the majority stays below $1 - r_{th} = 0.45$, so reputation admits them. SC-Joint's AUC then follows NaïveOR down to 0.64 at p = 0.3–0.4, well below Outlier/MAD (0.81–0.84). For p ≥ 0.5 the attackers are excluded and SC-Joint recovers to 0.95, above Outlier/MAD (≈0.87). The valley barely reaches the network layer: AUC falls by 32.6% while feasibility falls by at most 10.3%. SC-Joint's feasibility stays above Outlier/MAD at every p (87–97% vs 65–70%), and NaïveOR remains at zero. Adding a MAD stage fills the valley but lowers feasibility to about 70%, the Outlier/MAD level, so the deployed rule uses reputation alone.

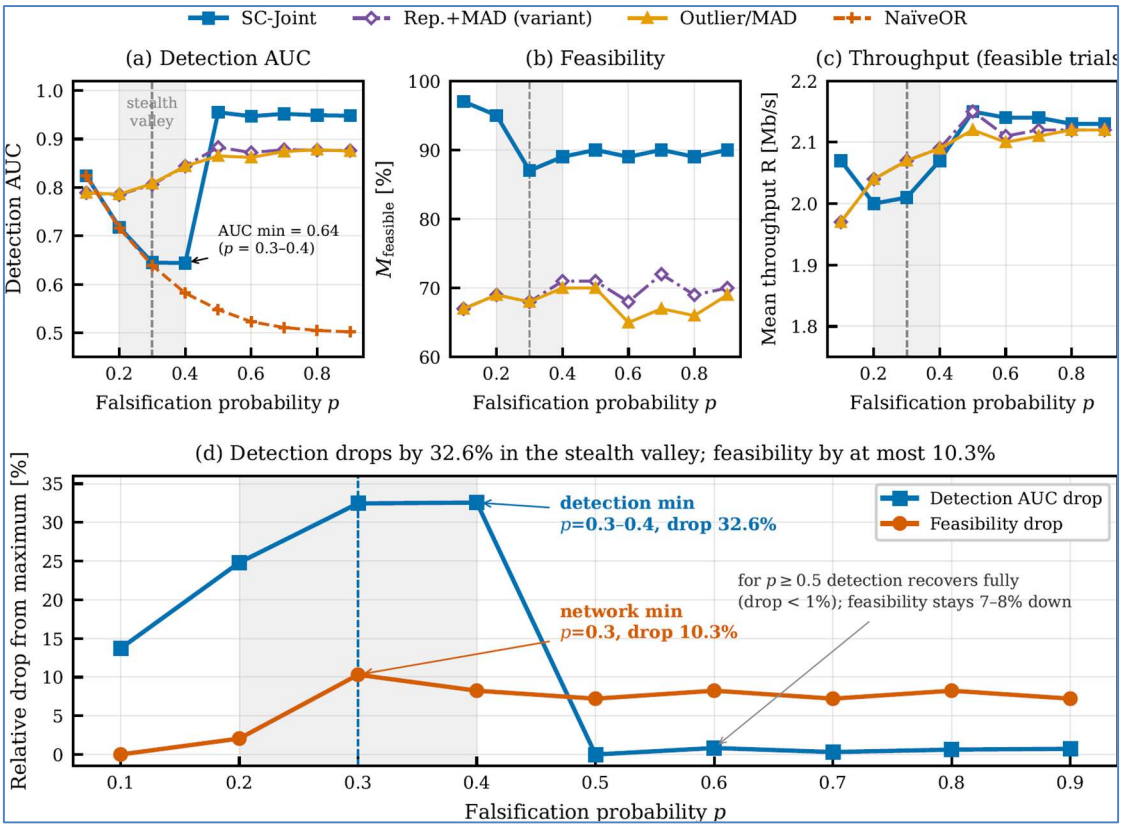

Figure 13. Stealth-attack sweep (Scenario 9; k = 3, M = 300). (a) AUC, (b) feasibility, (c) throughput of feasible trials. In the shaded valley ($p \leq 0.4$), reputation admits the attackers and SC-Joint falls to the NaïveOR level (AUC 0.64). (d) In that valley, detection drops by 32.6% while feasibility drops by at most 10.3%. For $p \geq 0.5$, SC-Joint detection recovers fully. Rep.+MAD is shown for reference only.

7.10 Admission sanity check and shrinking-pool behavior

Proposition 2 states that feasibility is monotone non-increasing under nested always-yes injection (Figure 14). We verify this directly on the same paired-CRN Scenario-5 run, in which attackers are added cumulatively and the feasible count is recorded at each stage. No attacker was admitted in any of the 300 realizations, and feasibility declined monotonically (98% → 86%) as $|\mathcal{S}|$ fell from 12 to 8, which serves as an implementation check of the admission rule and of Proposition 1's shrinking-pool prediction.

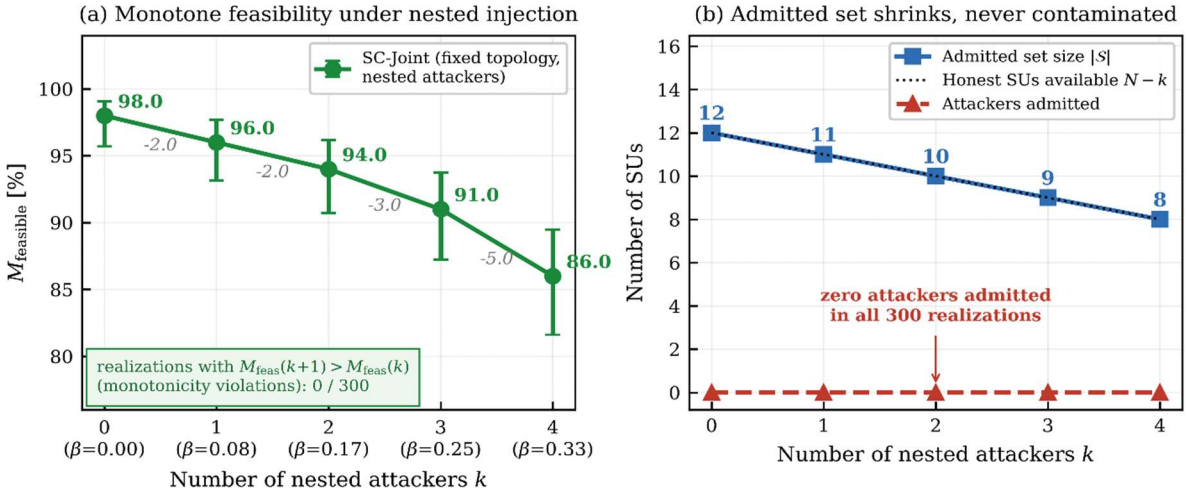

Figure 14. Scenario 5 admission check: SC-Joint under nested always-yes injection (paired CRN, M = 300). (a) Feasibility decreases monotonically from 98% to 86% as the honest pool gets smaller; (b) the admitted set follows $|\mathcal{S}| = N - k$, with zero attackers admitted in all 300 realizations.

7.11 Ablation / Contribution of each component

The SC-Joint method is based on the same topology, seed selection and constraints as Table 4 (flip p=0.5, 4 attackers, M=60), but separated into its components, i.e. trust, reputation and both trust-reputation. As shown in Table 5, trust is most important component and the admission of reputation yields 100.0% [94.0, 100.0] feasible. The accuracy was 85.0% [73.9, 91.9] with optimization without trust and 81.7% [70.1, 89.4] with MAD. None of them overlap. The result of adding MAD over reputation is a lower feasibility of 93.3% [84.1, 97.4]. Its two-sided band also throws out 0.77 honest nodes per realization on average, which is why the deployed rule does not include MAD. If the sensing time is fixed at $\tau = 4$ ms, feasibility falls to 76.7% [64.6, 85.6]. All of this loss is due to the lifetime constraint; once that constraint is removed, the fixed-τ variant also reaches 100%. Optimizing τ thus mostly protects lifetime and gains only +0.05 Mb/s of throughput (2.215 vs 2.165 Mb/s).

Table 5. Ablation (flip p = 0.5, four attackers, M = 60; identical topologies and seeds)

Variant	τ optimized	T_{life} enforced	Feasibility [%]	Wilson 95% [lo, hi]	R [Mb/s]
Optimizer-only	yes	yes	85.0	[73.9, 91.9]	2.061
MAD-only	yes	yes	81.7	[70.1, 89.4]	2.069
Reputation + MAD (variant)	yes	yes	93.3	[84.1, 97.4]	2.213
Trust + fixed τ	no (4 ms)	yes	76.7	[64.6, 85.6]	2.169
Trust + fixed τ, no T_{life}	no (4 ms)	no	100.0	[94.0, 100.0]	2.165
SC-Joint	yes	yes	100.0	[94.0, 100.0]	2.215

7.12 Reputation robustness and scope conditions

Leave-one-out reputation can only be relied on while honest reports are in the majority. We evaluate our approach against five different adversaries across various malicious fraction β and compared with a decayed-reputation version of our approach ω (Table 6). In each of the cases the always-yes attacker separates cleanly (0% false admission rate for all β).

The on-off attacker degrades gradually as intermittent honesty dilutes accumulated evidence (5% at $\beta = 0.25$, 28% at 0.33, 70% at 0.42); a colluding coalition is admitted once it sways the consensus (57% at $\beta = 0.25$, 98% at 0.42); and at $\beta = 0.42$ the majority assumption itself breaks (44.5% honest exclusion). The adaptive mimic is never rejected (100% at every β) but causes negligible harm: because it is constrained to behave honestly, its effect on detection is negligible ($\Delta P_{d,} \leq 0.007$). Effect of forgetting. Equal frame weighting leaves the scheme open to a two-phase attacker that accumulates credit before attacking: it is admitted 88.3% of the time at $\omega = 1$, but 0% under $\omega = 0.99$.

Table 6. Attacker false-admission rate FN [%] of the reputation rule versus malicious fraction β ($\omega = 1$; M = 60 per model–β pair) and versus forgetting factor ω at $\beta = 0.25$ (M = 60); -: not evaluated.

Attack (FN)	β=0.08	0.17	0.25	0.33	0.42	ω =0.99	ω =0.95
Always-yes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
On-off	0.0	1	5	28	70	12.2	21.1
Colluding	0.0	17	57	85	98	-	-
Adaptive	100	100	100	100	100	100	100
Two-phase	-	-	88.3	-	-	0.0	0.0

The cost is weaker duty-cycled detection (on-off false admission rises from 2.2% to 12.2% at $\omega = 0.99$ and 21.1% at $\omega = 0.95$). The adaptive mimic is admitted at every ω, suggesting that its evasion is structural rather than a consequence of unbounded memory. Scope condition. At these SNRs most SUs carries little information,

so the leave-one-out majority is dominated by idle reports, and separation relies on that asymmetry. The margin between the honest minimum and the attacker maximum shrinks from +0.745 at $P(H_1) = 0.10$ to +0.386 at 0.30, and becomes negative at higher occupancy (−0.005 at 0.50, −0.350 at 0.70). For this reason, we state idle-dominant occupancy as a scope condition of the mechanism, not as a general property (Section 8).

7.13 Robustness checks

Further robustness checks are summarized in Table 7.

- At the operating point we evaluate, the reputation is in effect a report-rate threshold (Proposition 3); an absolute rule $\rho_i \leq 0.45$ admits exactly the same set in every topology. What separates it from Outlier/MAD is that its admission rule is absolute rather than relative. Even with no attack, the MAD band cuts 2.17 honest SUs per topology, and in 82.7% of cases one of them is the strongest SU.
- Once $P(H_1) \geq 0.45$, the rule starts to exclude high-SNR honest SUs, which is what Eq. (11d) predicts.
- Always-no attackers are admitted but cost under 1 point of feasibility under OR fusion.
- Evaluating on a disjoint training window does not change feasibility.
- SC-Joint selects a single SU in most trials (K = 1 in 74–95%).
- NaïveOR's attack-free feasibility depends on the placement of R_{min}, whereas SC-Joint's does not.
- Under majority fusion, SC-Joint is still the best of the schemes tested. Paired on identical topologies, it beats Outlier/MAD by 0.089–0.161 in AUC, and for $k \leq 3$ it cannot be told apart from oracle exclusion.

Table 7. Robustness checks (SNR −20 dB, M = 300)

Check	Condition	SC-Joint	Reference
Report-rate threshold ($\rho \leq 0.45$)	k = 3 flip	89.7% feasible	Identical admitted set in 100% of topologies
Outlier/MAD honest trimming	no attack	----------	2.17 honest SUs trimmed; strongest SU in 82.7%
High $P(H_1)$, no attack	$P(H_1) = 0.45$ / 0.50	84.0 / 72.0%	All admitted: 95.3 / 94.7%
Always-no attackers	k = 3	89.0%	Oracle: 89.3%
Disjoint training window	B = 300 / 2000	88.0 / 89.7%	Same window: 88.0%
Selected set size	k = 0–4	K = 1 in 74–81% (−20 dB), 88–95% (−18 dB)	----------
R_{min}= 0.90 / 1.20, no attack	----------	98.0 / 96.0%	NaïveOR: 98.0 / 2.0%
Majority fusion	k = 3 flip	76.0%	Outlier 56.3%, NaïveOR 0%
Paired ΔAUC (95% CI)	k = 0–4	vs Outlier: +0.089 to +0.161; all CIs > 0 (lowest bound +0.078)	vs oracle: ≤ 0.001; CI includes 0 for $k \leq 3$

8. Limitations

1. *Scope:* The analysis assumes OR fusion; majority fusion is tested only as a robustness check (Section 7.13). The model is a single channel with one PU, twelve SUs and block-static Rayleigh fading, and it is evaluated in simulation only. Thresholds are fixed $r_{th} = 0.55$; κ = 3.0 or 2.5 by family), so results are not threshold-invariant. Because the optimizer mostly selects one SU (K = 1 in 74–95% of trials), the results show that trust keeps the strongest SU honest; they do not show gains from large cooperating sets. Outlier/MAD is a binary-report adaptation of [32].
2. *Trust layer*: (i) The reputation reduces to a report-rate threshold (Proposition 3), and flip attackers with $p < p^* \approx 0.44$ are admitted (Section 7.9). (ii) Colluding or majority-forming coalitions evade the leave-one-out consensus, and duty-cycled attackers weaken it (Table 6). Always-no attackers are admitted but

harmless under OR fusion. (iii) Evidence never decays: a two-phase attacker is admitted in 88.3% of cases. Forgetting with $\omega = 0.99$ removes this exposure but weakens duty-cycled detection (2.2% → 12.2%). (iv) For $P(H_1) \geq 0.45$ the highest-SNR honest SU is excluded in 36–63% of topologies and feasibility drop by up to 23 points, consistent with $\pi_1^* \approx 0.42$ (Eq. (11d)). "Oracle-free" therefore means operating without ground-truth labels; it does not mean that every attacker can be identified. Countering coordinated adversaries, for example by fusing the consensus with physical-layer fingerprints or with adaptive forgetting, is left to future work.

9. Future work

Four directions follow from the limitations above.

1. *Coordinated adversaries.* One next step is to make the reputation layer work against colluding and majority-forming attackers. A possible route is to combine the consensus signal with an independent physical-layer fingerprint, so that resilience no longer requires an honest majority [16], [49].
2. *Adaptive trust parameters.* The fixed threshold r_{th} and forgetting factor ω could give way to a learned policy that follows the attack regime and PU activity as they are observed [20]. Any such policy would have to deal with the trade-off seen for the decayed reputation (Table 6) without excluding honest SUs when $P(H_1)$ is high (Eq. (11d)).
3. *Fusion rules and denser deployments.* It remains to evaluate the framework with k-out-of-N and likelihood-ratio fusion, and in higher-SNR or denser networks, where larger cooperating sets become optimal.
4. *Reinforcement-learning control.* Joint admission and sensing-time selection could be cast as a sequential decision process that maximizes long-horizon throughput under SLA and trust constraints [26], [28], [42], including multi-channel, multi-PU settings.

10. Conclusions

This paper presented SC-Joint, an integration framework that treats SSDF defense in cooperative CRNs as an admission constraint inside a cross-layer problem. SUs are admitted before fusion by an oracle-free reputation, after which the cooperating set and the sensing time are chosen together to maximize throughput subject to detection, throughput and lifetime constraints. Success is thus judged by whether the network continues to meet its SLA, and not only by whether attackers are detected. Propositions 1–2 explain why this matters with OR fusion. A single always-yes SU is enough to drive throughput to zero, and the undefended baseline loses feasibility as more attackers join. Proposition 3 states when the reputation admits or excludes an SU, and its two closed-form thresholds anticipate the simulated behavior without any fitting: attackers with falsification probability below $p^* \approx 0.44$ are admitted, and high-SNR honest SUs are excluded once $P(H_1)$ exceeds $\pi_1^* \approx 0.42$.

Three findings come out of the evaluation. The first is that admission keeps the network running. Once a single attacker is admitted, undefended OR fusion has no feasible allocation left, while SC-Joint maintains 86–98% feasibility at about 2.14–2.16 Mb/s, and 97–100% under network stress. The AUC of the proposed algorithm exceeds the baseline algorithms of Outlier and MAD by 0.089 to 0.161 respectively. With paired experiment settings on topologies, SC-Joint algorithm shows oracle exclusion. Second, most of the gain is due to trust-based admission. In ablation experiments, admission is the essential factor and optimization to reduce the sensing time mainly safeguards the network lifetime (fixed τ: 76.7% feasibility, becoming 100% feasibility when no lifetime constraint is set). Regarding feasibility, the nested search is consistent with the exhaustive enumeration (0/40 disagreements) and gives a mean throughput difference of 0.42%. Besides, detection metric trends and network metric trends may not necessarily be congruent. In the stealth search, the AUC drops to 0.64 at p=0.3-0.4, a decrement of 32.6%, while maximum feasibility falls by 10.3% at most, all relative to the maxima. The valley matches the admission region that proposition 3 anticipates, a proof that detection metrics by themselves are not indicative of network resilience.

On the operating point the evaluator the reputation is a simple gatekeeper report-rate threshold, the optimizer picks a single SU. Under OR fusion noncolliding attackers are covered as long as $P(H_1) < 0.45$. Always-no attackers are not harmed. Stealthy attackers less than p* and colluding or forming majority coalitions are outside the envelope. It is likely that trust estimation should deal with coordinated adversaries and the parameters should adapt to changes.

References

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," IEEE Personal Communications, vol. 6, no. 4, pp. 13–18, 1999, https://doi.org/10.1109/98.788210.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, pp. 201–220, 2005, https://doi.org/10.1109/JSAC.2004.839380.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," Computer Networks, vol. 50, no. 13, pp. 2127–2159, 2006, https://doi.org/10.1016/j.comnet.2006.05.001.

[4] H. Urkowitz, "Energy detection of unknown deterministic signals," Proceedings of the IEEE, vol. 55, no. 4, pp. 523–531, 1967, https://doi.org/10.1109/PROC.1967.5573.

[5] R. Chen, J.-M. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in Proc. IEEE INFOCOM 2008 - The 27th Conference on Computer Communications pp. 1876–1884, http://doi.org/10.1109/INFOCOM.2008.251.

[6] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 428–445, 2013. https://doi.org/10.1109/SURV.2011.122211.00162

[7] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in Proc. IEEE DySPAN, 2005, pp. 131–136, http://dx.doi.org/10.1109/dyspan.2005.1542627 .

[8] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-22, no. 1, pp. 98–101, 1986, https://doi.org/10.1109/TAES.1986.310699.

[9] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," IEEE Transactions on Wireless Communications, vol. 7, no. 4, pp. 1326–1337, 2008, http://dx.doi.org/10.1109/TWC.2008.060869.

[10] E. C. Y. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, "Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view," IEEE Transactions on Vehicular Technology, vol. 58, no. 9, pp. 5294–5299, 2009, https://doi.org/10.1109/tvt.2009.2028030.

[11] R. Swaroop et al., "Energy-efficient cooperative spectrum sensing for cognitive radio networks over fading channel: Design and trade-offs," Engineering Research Express, Volume 7, Number 2, June 2025, https://doi.org/10.1088/2631-8695/ade2b9.

[12] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660–670, 2002, https://doi.org/10.1109/TWC.2002.804190.

[13] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of Byzantine attacks in cognitive radio networks," IEEE Transactions on Signal Processing, vol. 59, no. 2, pp. 774–786, 2011, https://doi.org/10.1109/TSP.2010.2091277.

[14] L. Zhang, G. Ding, Q. Wu, Y. Zou, Z. Han, and J. Wang, "Byzantine attack and defense in cognitive radio networks: A survey," IEEE Communications Surveys & Tutorials, vol. 17, no. 3, pp. 1342–1363, 2015, https://doi.org/10.1109/COMST.2015.2422735.

[15] Z. Xu et al., "Joint spectrum sensing and spectrum access for defending massive SSDF attacks: A novel defense framework," Chinese Journal of Electronics, Volume31, Issue2, March 2022, Pages 240-254, 2022, https://doi.org/10.1049/cje.2021.00.090Digital Object Identifier (DOI).

[16] M. Liu et al., "Attacking spectrum sensing with adversarial deep learning in cognitive radio-enabled Internet of Things," IEEE Transactions on Reliability, Volume: 72, Issue: 2, Page(s): 431 – 444, June 2023, https://doi.org/10.1109/TR.2022.3179491.

[17] N. Parhizgar et al., "Defense against spectrum sensing data falsification attack in cognitive radio networks using machine learning," in Proc. 30th Int. Conf. Electrical Engineering (ICEE), 2022, https://doi.org/10.1109/ICEE55646.2022.9827418.

[18] Y. Li et al., "Isolation Forest and spectral clustering based cooperative spectrum sensing against SSDF attack in CWSNs," IEEE Sensors Journal, Volume: 25, Issue: 11, 01 June 2025, https://doi.org/10.1109/JSEN.2025.3559631.

[19] A. Tiwari et al., "Reputation-based classification for malicious user detection in cognitive radio networks," in Proc. 17th Int. Conf. Communication Systems and Networks (COMSNETS), 2025, https://doi.org/10.1109/COMSNETS63942.2025.10885652.

[20] G. Xie et al., "Adaptive trust threshold model based on reinforcement learning in cooperative spectrum sensing," Sensors 23(10), 4751, 2023, https://doi.org/10.3390/s23104751.

[21] N. Dewangan et al., "Passive reputation method with hybrid optimized deep model for spectrum sensing in cognitive radio network based on blockchain," International Journal of Communication Systems, Volume37, Issue7,2024, https://doi.org/10.1002/dac.5704Digital Object Identifier (DOI).

[22] K. Zeng, P. Pawełczak, and D. Cabrić, "Reputation-based cooperative spectrum sensing with trusted nodes assistance," IEEE Communications Letters, vol. 14, no. 3, pp. 226–228, 2010, https://doi.org/10.1109/LCOMM.2010.03.092240.

[23] M. Dixit et al., "DRL-SecRoute: A synergetic deep reinforcement learning paradigm for mitigating Byzantine faults and SSDF attacks through heuristic spectrum cognizance in next-generation cognitive radio networks," The Journal of Supercomputing, Volume 82, article number 401, 2026, https://link.springer.com/article/10.1007/s11227-026-08540-3.

[24] R. Kumar et al., "Reinforcement learning for performance improvement in cooperative spectrum sensing," Physical Communication, p: 102112, Volume 59, 2023, https://doi.org/10.1016/j.phycom.2023.102112.

[25] N. Biswas et al., "Buffer-aware user selection and resource allocation for an opportunistic cognitive radio network: A cross-layer approach," IEEE/ACM Transactions on Networking, Volume: 30, Issue: 5, October 2022, https://doi.org/10.1109/TNET.2022.3159819.

[26] T. Sheela et al., "Cross-Layer Optimization in AI-Powered Cognitive Radio Networks for Dynamic Spectrum Access" in Proc. 9th Int. Conf. Electronics, Communication and Aerospace Technology (ICECA), 2025, https://doi.org/10.1109/ICECA66444.2025.11382605.

[27] S. Latif et al., "An efficient Pareto optimal resource allocation scheme in cognitive radio-based Internet of Things networks," Sensors, *22*(2), p:451, 2022, https://doi.org/10.3390/s22020451.

[28] M. Gupta et al., "RORA: Reinforcement learning based optimal distributed resource allocation strategies in vehicular cognitive radio networks for 6G," Vehicular Communications, Volume 52, p:100882, 2025, https://doi.org/10.1016/j.vehcom.2025.100882.

[29] R. Lin et al., "Physical-layer security enhancement in energy-harvesting-based cognitive Internet of Things: A GAN-powered deep reinforcement learning approach," IEEE Internet of Things Journal, Volume: 11, Issue: 3, Page(s): 4899 – 4913, 2024, https://doi.org/10.1109/JIOT.2023.3300770.

[30] A. Taherpour et al., "Adaptive learning for IRS-assisted wireless networks: Securing opportunistic communications against Byzantine eavesdroppers," IEEE Journal on Selected Areas in Communications, Volume: 44, Page(s): 3926 – 3948, 05 March 2026, https://doi.org/10.1109/JSAC.2026.3670779.

[31] M. Tatar et al., "Securing integrated sensing and communication against a mobile adversary: A Stackelberg game with deep reinforcement learning," IEEE Journal on Selected Areas in Communications, Volume: 44, Page(s): 942 – 958, 2025, https://doi.org/10.1109/JSAC.2025.3611404.

[32] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," IEEE Trans. Wireless Commun., vol. 9, no. 8, pp. 2488–2497, 2010. https://doi.org/10.1109/TWC.2010.061510.090395.

[33] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attacks," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 16–29, Jan. 2009. https://doi.org/10.1109/TSP.2008.2007335

[34] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "Distributed Bayesian detection in the presence of Byzantine data," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5250–5263, Oct. 2015, https://doi.org/10.1109/TSP.2015.2450191.

[35] A. Paul et al., "Joint spectrum sensing and D2D communications in cognitive radio networks using clustering and deep learning strategies under SSDF attacks," Ad Hoc Networks, Volume 143, 15 April 2023, Page 103116, 2023. https://doi.org/10.1016/j.adhoc.2023.103116.

[36] J. Wu et al., "Sequential spectrum sensing against random Byzantine attack in cognitive radio networks," Transactions on Emerging Telecommunications Technologies, Volume34, Issue7, 2023, https://doi.org/10.1002/ett.4796.

[37] W. Wang, H. Li, Y. Sun, and Z. Han, "Attack-proof collaborative spectrum sensing in cognitive radio networks," in Proc. CISS, 2009, pp. 130–134, https://doi.org/10.1109/CISS.2009.5054704.

[38] M. Mahvash et al., "Blockchain-inspired trust management in cognitive radio networks with cooperative spectrum sensing," Pervasive and Mobile Computing, 101999, Volume 106, 2024, https://doi.org/10.1016/j.pmcj.2024.101999.

[39] C. Chen et al., "Regional autonomous security cooperative spectrum sensing method based on trust value," IET Communications, 17:1721–1729, 2023. https://doi.org/10.1049/cmu2.12647.

[40] M. Giri et al., "Extreme learning machine-based identification of malicious users for secure cooperative spectrum sensing in cognitive radio networks," Wireless Personal Communications, v. 130, n. 3, p. 1993, 2023, https://doi.org/10.1007/s11277-023-10368-6.

[41] H. Chen et al., "Latency-aware cross-layer signaling for energy-efficient CRNs," International Journal of Distributed Sensor Networks, Volume. 25, Article ID 1778624, 13 pages, September 2025, https://doi.org/10.1155/dsn/1778624.

[42] X. Zhang et al., "Decentralized routing and radio resource allocation in wireless ad hoc networks via graph reinforcement learning," IEEE Transactions on Cognitive Communications and Networking, Volume: 10, Issue:3, June 2024, https://doi.org/10.1109/TCCN.2024.3360517.

[43] M. A. Ferrag et al., "Edge learning for 6G-enabled Internet of Things: A comprehensive survey of vulnerabilities, datasets, and defenses," IEEE Communications Surveys & Tutorials, Volume: 25, Issue: 4, Page(s): 2654 – 2713, 2023, https://doi.org/10.1109/COMST.2023.3317242.

[44] I. E. Ezhilarasi et al., "A survey on cognitive radio network attack mitigation using machine learning and blockchain," EURASIP Journal on Wireless Communications and Networking, volume 2023, article number 98, 2023, https://doi.org/10.1186/s13638-023-02290-z.

[45] A. Jøsang and R. Ismail, "The Beta reputation system," in *Proc. 15th Bled Electronic Commerce Conf.*, Bled, Slovenia, 2002, pp. 324–337.

[46] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. Amer. Statist. Assoc.*, vol. 88, no. 424, pp. 1273–1283, 1993. https://doi.org/10.1080/01621459.1993.10476408.

[47] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *J. Amer. Statist. Assoc.*, vol. 22, no. 158, pp. 209–212, 1927. https://doi.org/10.1080/01621459.1927.10502953.

[48] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982. https://doi.org/10.1148/radiology.143.1.7063747

[49] B. D. Son et al., "Adversarial attacks and defenses in 6G network-assisted IoT systems," IEEE Internet of Things Journal, Volume: 11, Issue: 11, Page(s): 19168 – 19187, 2024, https://doi.org/10.1109/JIOT.2024.3373808.