AI Driven Semantic Protocol Attacks on 6G: Formalizing Novel Threat Vectors and a Lipschitz Bounded Cross Layer Defense Framework (SACLDF)

Abdellah Tahenni*, Messaoud Ahmed Ouameur*, Miloud Bagaa*, Daniel Massicotte*, Sifeddine Salmi*, Felipe Augusto Pereira de Figueiredo†, Adlen Ksentini‡ *Department of Electrical and Computer Engineering, University of Quebec at Trois-Rivières (UQTR), QC, Canada
{abdellah.tahenni, messaoud.ahmed.ouameur, miloud.bagaa, daniel.massicotte, sifeddine.salmi}@uqtr.ca
†Instituto Nacional de Telecomunicações (INATEL), Brazil
felipe.figueiredo@inatel.br ‡EURECOM, France
ksentini@eurecom.fr

***Abstract*—Sixth generation (6G) semantic communications (SEMCOM) introduce AI driven protocol layers context mapping, ontology reasoning, and reinforcement learning based (RL) routing that expose a new class of adversarial vulnerabilities beyond conventional physical layer threats. Existing security research largely addresses physical layer attacks such as jamming and waveform inversion, leaving protocol layer mechanisms insufficiently protected. This paper formalizes three novel protocol layer attack vectors: *context injection* (FGSM/PGD based perturbation of semantic context functions), *ontology poisoning* (adversarial triple injection into shared knowledge graphs), and *reward manipulation* (deceptive RL signal distortion). End to end MATLAB simulations demonstrate severe impact: bit error rate (BER) degradation from 5.3×10^{-7} to 8.7×10^{-4} (AWGN), perceptual speech quality (PESQ) below intelligibility thresholds (1.3 under Rayleigh fading), and up to 12.4 dB signal to distortion ratio (SDR) loss. To counter these threats, we propose the *Semantic Aware Cross Layer Defense Framework (SACLDF)*, unifying: (1) real time multi metric anomaly detection (SAAD), (2) Lipschitz constrained semantic reconstruction (CSAE) with a certified reconstruction bound (Theorem 1), (3) hybrid knowledge graph sanitization (KGG), and (4) trust aware policy adaptation (ARLD). SACLDF achieves a $93\times$ BER reduction, restores PESQ to 3.2, and recovers 11.6 dB SDR while maintaining a gNB side latency of 4.7 ms and a memory footprint of 50 MB meeting 6G URLLC constraints. An ablation study and comparative evaluation against standard adversarial training and robust routing baselines confirm the necessity of the cross layer architecture.**

***Index Terms*—Semantic attacks, semantic protocols, protocol learning, AI enabled semantic protocols, knowledge graph poisoning, adversarial attacks, 6G security, cross layer defense.**

I. Introduction

The emergence of next generation wireless networks and semantic communications marks a significant paradigm shift: instead of merely exchanging bits, systems now convey meaning by leveraging shared knowledge bases and AI driven semantic encoders [1]–[3]. While this approach promises gains in spectral efficiency and latency reduction, it also opens a fundamentally new attack surface at the protocol layer [4]–[6]. To date, most security research in SEMCOM has focused on physical layer threats such as jamming and waveform inversion, leaving semantic level mechanisms context mapping, adaptive routing, and knowledge graph reasoning insufficiently protected despite their growing deployment in real world networks [7], [8]. The consequence is that adversaries can now exploit the very AI intelligence built into 6G protocols to degrade system performance in ways that conventional defenses cannot detect or mitigate. In this work, we show that adversaries can exploit protocol layer features to severe effect. By injecting carefully crafted perturbations into context mapping functions, attackers force systematic misinterpretation at the receiver. Poisoning shared ontologies and knowledge graphs with false relationships undermines downstream inference and decision logic. Manipulating reward signals and routing updates in adaptive protocols subverts resource allocation and traffic flows. For each of these attack vectors context injection, ontology and graph poisoning, and route and reward manipulation we develop rigorous mathematical models, integrating gradient based adversarial techniques [9], [10], graph theoretic corruption metrics, and trust weighted routing formulations [11] into a unified, end to end attack framework [12]–[15]. The primary contribution of this work is a *unified treatment* of semantic protocol layer threat modeling and cross layer defense for 6G SEMCOM systems an aspect largely absent from prior work that focuses either on physical layer robustness [16]–[18] or on semantic efficiency without adversarial analysis [1], [2], [12]. Specifically, we:

1) Formalize three novel protocol layer attack vectors with rigorous mathematical models (Section III).
2) Demonstrate severe performance degradation under these

attacks via end to end MATLAB simulation (Section IV).
3) Propose SACLDF, a cross layer defense with a certified reconstruction bound (Theorem 1), and validate it against standard defense baselines (Section VI).

The overall attack surface is depicted in Fig. 1 (Section III). The remainder of this paper is organized as follows. Section II reviews related work. Section III formalizes the system, threat, and attack models. Section IV presents attack evaluation results. Section V outlines defense recommendations. Section VI details SACLDF. Section VII concludes.

II. Related Work

A. Semantic Communication Security

Recent developments in SEMCOM have focused on efficiency improvements via deep learning architectures [2], [19]. The distributed semantic system in [1] targets IoT applications, demonstrating promising gains in bandwidth efficiency but omitting any adversarial evaluation. Speech transmission by semantic extraction is described in [12], providing a foundational architecture for speech based SEMCOM. Image semantic coding through learned representations is introduced in [20], establishing deep learning based compression pipelines that are later shown to be vulnerable. Despite these advances, security analyses have largely focused on physical layer attacks [16], [18]. Multi domain vulnerabilities have been found with scope limited to conventional network layers [21]. Further studies revealed physical layer adversarial robustness gaps in deep learning based communications [7]. SEMCOM black box attacks via gradient estimation were demonstrated in [17], exposing the brittleness of semantic encoders under transfer attacks. Protocol layer vulnerabilities such as context aware routing mechanisms, dynamic ontology adaptation [5], and federated semantic learning architectures remain underexplored [2], [22]. Recent 2025–2026 works have begun to address cross layer attack propagation [23], ontology poisoning [24], certified robustness [25], knowledge graph security [26], RL based protocol attacks [27], O RAN security vulnerabilities [28], multimodal resilience [29], and federated defense [30], yet none provides a unified protocol layer threat model coupled with a cross layer defense framework.

Table II presents a detailed comparison of related works, highlighting how existing approaches address only isolated attack vectors and lack coordinated cross layer defense mechanisms.

B. Key Limitations of Prior Work

Examination of each work points to several critical limitations that motivate the present contribution. The work in [1] provides no adversarial evaluation and is scoped exclusively to IoT deployments, leaving open questions about resilience in the broader 6G context. The work in [31] ignores protocol layer weaknesses and assumes static receiver architectures, which are unrealistic in adaptive 6G deployments. The work in [7] performs adversarial analysis only at the application layer, missing the protocol level propagation effects we formalize here. The work in [32] lacks cross layer attack propagation modeling despite targeting multimodal semantic systems that inherently span multiple layers. The work in [17] restricts attention to physical layer adversarial resilience, demonstrating that transfer attacks can break semantic encoders but offering no protocol aware defense. The work in [15] covers data poisoning for text modalities only, leaving speech and image based SEMCOM systems unaddressed. The work in [21] considers black box gradient attacks without protocol level effects, limiting its applicability to knowledge graph or reward manipulation scenarios. Recent works [23]–[25], [33] represent important advances but each addresses only a subset of the attack surface and lacks a unified cross layer defense. The work of [23] provides cross layer propagation analysis but does not produce a certified defense bound or address knowledge graph poisoning. The certified autoencoder in [25] offers a formal reconstruction bound but covers only a single encoder and provides no RL or KG defense. These limitations collectively highlight the need for holistic protocol layer attack modeling and defense for SEMCOM systems, which SACLDF addresses. Most closely related to our defense framework are adversarial training methods [10] and robust routing protocols [11], which address physical layer and routing layer threats, respectively. Data poisoning defenses from the machine learning literature [34] also provide relevant background. Section VI demonstrates that these approaches, while effective against their intended threats, are insufficient against semantic layer attacks without cross layer coordination the primary contribution of SACLDF. Table I summarizes the key findings and limitations of the related works reviewed above, highlighting the research gaps that motivate our proposed framework.

TABLE I: Summary of Key Findings and Limitations of Related Works

Area	**Key Findings**	**Limitations**
SEMCOM Security	DL-based SEMCOM gains efficiency; physical layer attacks identified	No adversarial evaluation; protocol layer gaps; static assumptions
Adversarial Attacks	FGSM/PGD break encoders; transfer and black-box attacks demonstrated	Physical layer only; no protocol aware defense; single modality
2025–2026 Advances	Cross-layer propagation; ontology poisoning; KG security; certified robustness; federated defense	Subset coverage only; no unified threat model or cross-layer defense
Defense Methods	Adversarial training; robust routing; data poisoning defenses	Insufficient against semantic layer attacks without cross-layer coordination
Overall Gap	No unified protocol-layer threat model with cross-layer defense covering context injection, ontology poisoning, reward manipulation, and KG poisoning. SACLDF fills this gap.	

TABLE II: Comparison of Related Works on Semantic Quality and Channel Effects Under Attacks

Reference	Year	Semantic Quality Metrics	Channel Effects Analysis	Attack Types	Key Limitations
[1]	2020	Semantic similarity (IoT)	BER vs. SNR under baseline impairments	None	No adversarial analysis; IoT scope only
[31]	2024	Task accuracy	Physical-layer under idealized conditions	Channel inversion; waveform manipulation	Ignores protocol-layer; assumes static architectures
[7]	2023	Context preservation; semantic quality	Multi-domain SNR analysis	Adversarial examples (application layer)	No protocol-layer vulnerability analysis
[32]	2024	Multimodal alignment accuracy	MIMO channel effects	Model inversion	Early robustness evaluation; no cross-layer modeling
[17]	2024	Robustness score (DL metrics)	Adversarial SNR thresholds	White-box gradient attacks	Physical-layer only; limited cross-layer insight
[15]	2022	Text reconstruction fidelity	Fading channel evaluation	Data poisoning	Limited attacks; text modality only
[21]	2022	Semantic BER	Black-box perturbation under noise	Gradient-based black-box	No cross-layer or protocol-level analysis
[33]	2025	Task accuracy; semantic similarity	AWGN and Rayleigh under gradient perturbation	FGSM/PGD gradient-based (task-oriented layer)	Single-layer attack only; no ontology or RL analysis
[23]	2025	Semantic BER; reconstruction fidelity	Physical-to-protocol cross-layer SNR analysis	Cross-layer adversarial: physical + application	No certified defense bound; no KG poisoning
[24]	2026	Semantic accuracy; knowledge consistency	Baseline AWGN; no fading channel evaluation	Ontology poisoning via adversarial triple injection	No RL or reward manipulation; single attack vector
[25]	2025	PESQ; SDR; semantic reconstruction error	Rayleigh and Rician fading with bounded perturbations	ℓ_2-bounded input perturbations	Certified bound for single encoder; no KG or RL defense
Ours	2026	**BER, PESQ, SDR with semantic degradation**	**AWGN, Rayleigh, Rician, OOD channels**	**Context injection, ontology poisoning, reward manipulation, KG poisoning**	**SACLDF proposed; hardware validation deferred**

III. System Model and Threat Framework

The end to end 6G semantic communication pipeline and the three protocol layer attack vectors are illustrated in Fig. 1, which shows how each attack targets a distinct injection point in the stack. Context injection perturbs the semantic encoder, ontology poisoning corrupts the shared knowledge graph, and reward manipulation distorts the RL routing agent's feedback signal, collectively degrading BER, PESQ, and SDR below acceptable thresholds in the absence of any defense mechanism.

A. Semantic Communication System Model

The end to end semantic communication system is shown in Fig. 1. Raw data (text, images, or speech) is processed by a semantic feature extractor, then encoded for transmission over a wireless channel.

1) Transmitter Modeling: The transmitter encodes source data into a semantic representation optimized for transmission using a pretrained transformer model for feature extraction. Relevant semantic features are extracted using ResNet for images [19] and converted to a compact representation through PCA or autoencoders [2], [19]. The semantic representation is then encoded using LDPC or Turbo codes [22]. Mathematically:

$$s = f_{\text{enc}}(x;\, \theta_{\text{enc}}), \tag{1}$$

where x is the source data, f_{enc} is the encoding function with parameters θ_{enc}, and s is the transmitted signal.

2) Channel Modeling: The wireless channel is modeled as a fading channel with additive white Gaussian noise (AWGN). The received signal is:

$$y = h \cdot s + n, \tag{2}$$

where h is the channel fading coefficient, s is the transmitted signal, and $n \sim \mathcal{CN}(0, \sigma_n^2)$ is AWGN. Key parameters include fading distribution (Rayleigh, Rician, or Nakagami), path loss exponent, Doppler spread, and $\text{SNR} = P_s/\sigma_n^2$.

3) Receiver Modeling: The receiver reconstructs semantic information via channel decoding, semantic decoding, and data reconstruction [19]:

$$\hat{x} = f_{\text{dec}}(y;\, \theta_{\text{dec}}), \tag{3}$$

where $\hat{x}$ is the reconstructed data and f_{dec} is the decoding function with parameters θ_{dec}.

B. Adversary Capability Model

We formalize two threat levels evaluated throughout this work.

1) **Gray Box Adversary (Primary Model):** The adversary knows the semantic encoder architecture f_{enc} and the schema of the shared knowledge graph G, but does *not* have access to trained weights θ_{enc} or the SACLDF defense parameters (τ, ρ, λ). This models a *control plane insider* with documentation access but without direct model extraction capability. Context injec-

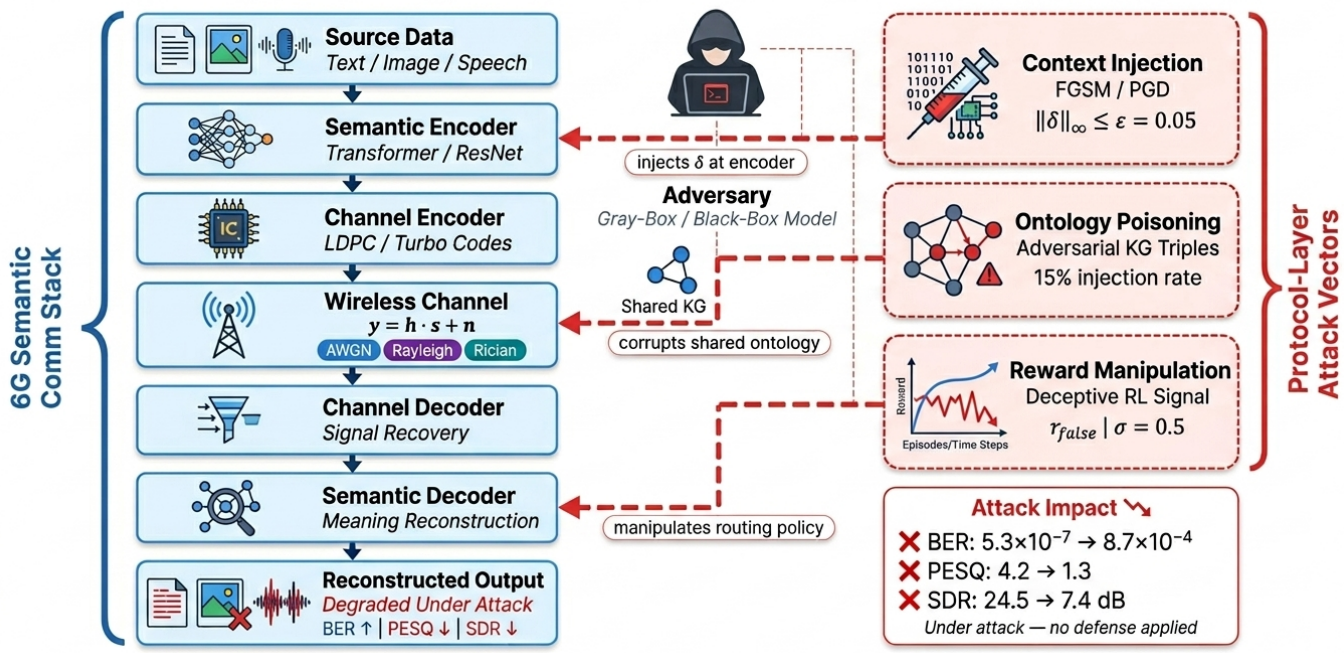


Fig. 1: End to end 6G semantic communication stack under three protocol layer adversarial attacks: context injection ($\|\delta\|_\infty \leq \varepsilon = 0.05$), ontology poisoning (15% adversarial KG triples), and reward manipulation (r_{false}, $\sigma = 0.5$). Dashed red arrows indicate adversarial injection points. Under attack, BER degrades from 5.3×10^{-7} to 8.7×10^{-4}, PESQ drops from 4.2 to 1.3, and SDR falls by 12.4 dB (no defense applied).

tion (FGSM/PGD) and ontology poisoning are evaluated under this model.

2) **Black Box Adversary (Secondary Model):** The adversary observes only input output behavior and uses gradient estimation (zeroth order methods) or transfer attacks from surrogate models. Knowledge graph poisoning and reward manipulation are modeled as black box, since the adversary manipulates external data feeds (ontology servers, reward channels) rather than model internals.

Attack Scope. We consider adversaries at two interface points: (i) a *pre encoder interface* where δ bounded perturbations are injected before semantic encoding, modeling a man in the middle attacker; and (ii) a *shared knowledge plane* where the adversary inserts or modifies triples in the distributed ontology, modeling a compromised ontology server. Full physical layer adversaries (jamming, waveform inversion) are outside this scope and addressed in [16]–[18].

Attack Constraints.

- *Perturbation budget:* $\|\delta\|_\infty \leq \epsilon = 0.05$ for context injection.
- *Graph poisoning rate:* at most 15% of triples modified or injected.
- *Reward noise:* additive noise bounded by $\sigma = 0.5$ on the true reward.
- *Stealthiness:* all attacks remain below conventional intrusion detection thresholds.

The refined intelligent semantic attack model, capturing the interdependencies among the three attack vectors, is depicted in Fig. 2.

C. Attack Vector Formalization

1) Context Injection (Semantic Protocols):

- **Attack Strategy:** The attacker exploits the transmitter's context mapping function to generate adversarial examples, causing misinterpretation at the receiver [7], [9].
- **Mathematical Model:** Let $f_{\text{context}}(x)$ assign a context label to input x. The attacker finds x' such that:

$$f_{\text{context}}(x') \neq f_{\text{context}}(x), \quad x' = x + \delta, \tag{4}$$

where δ is a perturbation constrained by $\|\delta\|_\infty \leq \epsilon$. FGSM and PGD are used to generate δ [10], [31].

2) Ontology Poisoning (Semantic Protocols):

- **Attack Strategy:** The attacker injects false or misleading triples into the shared ontology, skewing semantic interpretation [32].
- **Implementation:** The ontology structure is modified by inserting false relationships between concepts, analyzed using graph theory and knowledge representation techniques. The poisoning rate is bounded at 15% of the total triple count to maintain stealthiness, consistent with the threat model of Section III-B.

3) Adaptive Routing Attack (Protocol Learning):

- **Attack Strategy:** Adaptive routing protocols [5] are targeted by injecting false information into routing updates.

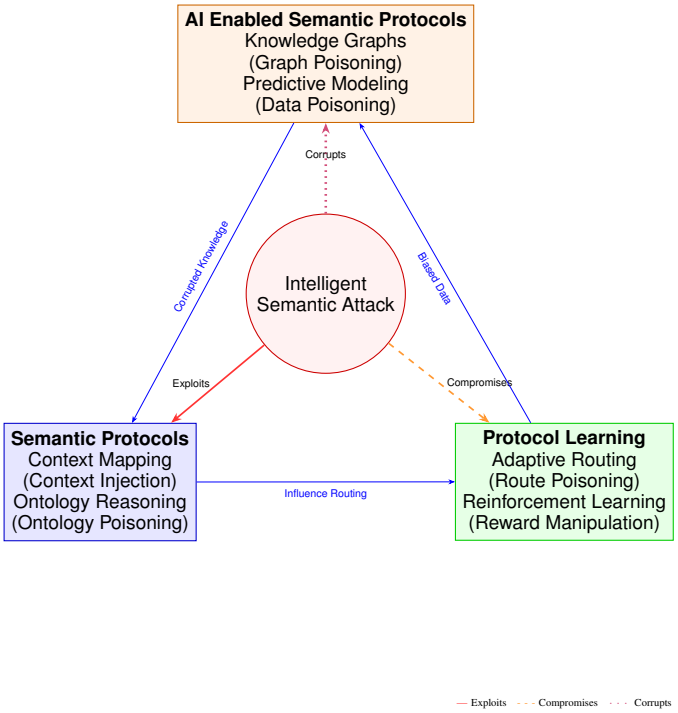


Fig. 2: Intelligent semantic attack model showing three attack vectors—exploiting semantic protocols, compromising protocol learning, and corrupting AI enabled protocols—with cross layer influence paths that allow one compromised layer to cascade degradation into adjacent layers.

- **Mathematical Model:** Let $R(t)$ be the routing table at time t and L the update function. The attacker injects malicious update $\Delta R(t)$:

$$R(t+1) = L(R(t), \text{network state}) + \alpha \cdot \Delta R(t), \quad (5)$$

where α is a weighting factor [11].

4) Reward Manipulation (Protocol Learning):

- **Attack Strategy:** RL based communication protocols are targeted by distorting reward signals [35].
- **Implementation:** The attacker distorts reward signals to manipulate the RL agent's behavior, leading to biased resource allocation via reward shaping techniques [35]. Reward noise is modeled as additive Gaussian with $\sigma = 0.5$, chosen to remain below conventional monitoring thresholds while still causing measurable policy degradation.

5) Knowledge Graph Poisoning (AI Enabled Protocols):

- **Description:** Knowledge graphs structure semantic relationships for reasoning. An attacker injects false triples to corrupt inferences [32].
- **Implementation:** Adversarial machine learning and knowledge graph embedding techniques [26] enable injection of 15% adversarial triples that alter downstream semantic decisions.

6) Predictive Model Attacks (AI Enabled Protocols):

- **Description:** Predictive models trained on historical data can be compromised by injecting biased adversarial examples into training sets [34].
- **Implementation:** Adversarial training data is crafted by solving an optimization problem that maximizes prediction error while minimizing data perturbation [15], [19].

D. Integrated Attack Algorithm

Algorithm 1 orchestrates a multi layer attack across semantic, protocol learning, and AI enabled layers. **Phase 1** generates adversarial context examples and injects false ontology relationships. **Phase 2** updates routing tables with malicious trust scores and distorts RL reward signals. **Phase 3** injects adversarial triples and poisons training data. The algorithm is designed to operate within the gray box adversary capability model of Section III-B, assuming architectural knowledge but not model weights.

Algorithm 1 AI Driven Intelligent Semantic Attacks

Require: ϵ: max perturbation; α: routing weight; η: RL learning rate; G: knowledge graph; D_{train}: training data; f_{context}: context mapping; MAX_ITER; β: trust threshold; ϕ: routing decay factor
Ensure: $x_{\text{adv}}, G_{\text{poison}}, R_{t+1}, r_{\text{false}}, D_{\text{poison}}, metrics$
procedure SEMANTICATTACK($x, G, R_t, s, D_{\text{train}}$)
  // Phase 1: Semantic Protocol Attacks
  $embeddings \leftarrow \text{InitializeEmbeddings}(G)$
  $(x_{\text{adv}}, \ell_1) \leftarrow \text{ContextMappingAttack}(x, f_{\text{context}}, \epsilon, MAX\_ITER)$
  $G_{\text{mod}} \leftarrow \text{OntologyReasoningAttack}(G, C, R)$
  // Phase 2: Protocol Learning Attacks
  $R_{t+1} \leftarrow \text{AdaptiveRoutingAttack}(R_t, S_t, \alpha, \beta, \phi)$
  $(r_{\text{false}}, conf) \leftarrow \text{RewardManipulationAttack}(s, a, r_{\text{true}}, policy)$
  // Phase 3: AI Enabled Protocol Attacks
  $(G_{\text{poison}}, scores) \leftarrow \text{KGPoisoning}(G_{\text{mod}}, T, embeddings)$
  $(D_{\text{poison}}, y_{\text{poison}}, model) \leftarrow \text{PredictiveAttack}(D_{\text{train}}, y_{\text{train}}, c_{\text{target}})$
  // Evaluation
  $metrics \leftarrow \text{ComputeMetrics}(x_{\text{adv}}, G_{\text{poison}}, R_{t+1}, r_{\text{false}})$
  $success\_rate \leftarrow \text{EvaluateSuccess}(metrics)$
  $stealthiness \leftarrow \text{AssessStealthiness}(x_{\text{adv}}, G_{\text{poison}}, R_{t+1})$
  return $(x_{\text{adv}}, G_{\text{poison}}, R_{t+1}, r_{\text{false}}, D_{\text{poison}}, metrics)$
end procedure

IV. Evaluation Results and Attack Performance

This section evaluates AI driven semantic attacks against 6G SEMCOM systems under AWGN, Rayleigh, and Rician channels. Results are obtained via MATLAB R2023b simulations using the Communications Toolbox®. The simulation framework considers context mapping attacks (FGSM/PGD, $\epsilon = 0.05$), ontology reasoning attacks, adaptive routing attacks, RL based reward manipulation ($\sigma = 0.5$), and knowledge graph poisoning (15% adversarial triple injection).

A. Performance Results

Table III compares performance under normal and attack conditions across all evaluated channel types, providing a comprehensive view of the degradation induced by each attack vector.

Fig. 3 shows BER vs. SNR over the range 0–30 dB. Under attack, BER degradation increases with SNR, confirming that protocol layer attacks are more severe at higher SNR regimes where semantic processing relies on richer, more attack exploitable representations.

Fig. 4 presents the PESQ distribution and SDR degradation across channel conditions, revealing that Rayleigh fading channels suffer the most significant quality loss under all three attack vectors.

TABLE III: Performance Metrics Across Channel Types Under Normal and Attack Conditions

Metric / Channel	AWGN	Rayleigh	Rician
BER (No Attack)	5.3×10^{-7}	3.9×10^{-5}	1.9×10^{-6}
BER (Targeted Attack)	8.7×10^{-4}	3.8×10^{-3}	1.1×10^{-3}
PESQ (Clean)	4.2	3.9	4.1
PESQ (Attacked)	1.8	1.3	2.1
SDR Clean [dB]	24.5	19.8	22.3
SDR Attacked [dB]	11.2	7.4	14.6

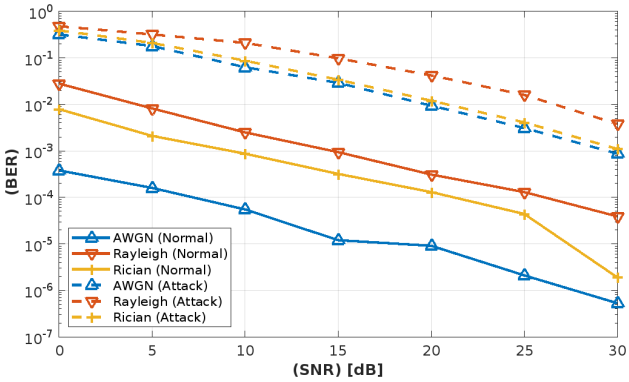


Fig. 3: Multi channel attack performance: BER vs. SNR under normal operation and targeted attacks across AWGN, Rayleigh, and Rician channels.

B. Key Findings

a) BER Performance: AWGN achieves the lowest clean BER (5.3×10^{-7}), reflecting the deterministic and well characterized nature of Gaussian noise. Under attack, however, AWGN BER degrades by three orders of magnitude to 8.7×10^{-4}, confirming that protocol layer attacks are not mitigated by favorable channel conditions alone. Rayleigh channels show the highest attack vulnerability (BER 3.8×10^{-3}), as the combination of fading induced symbol ambiguity and semantic perturbation creates compounded degradation. Rician channels demonstrate intermediate resilience (1.1×10^{-3}), benefiting from the LOS component but still exhibiting unacceptable error rates under coordinated attack. These results confirm that protocol layer attack severity is channel dependent and that any defense must account for channel statistics.

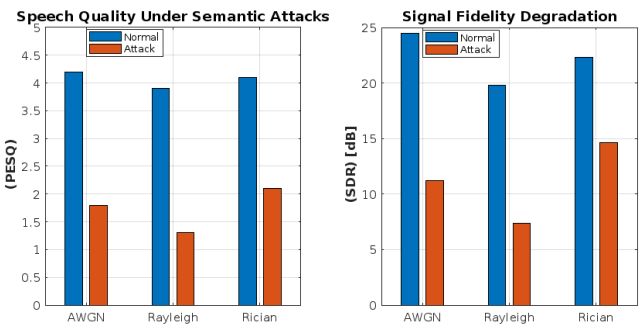


Fig. 4: Multi channel attack performance: PESQ distribution and SDR degradation comparison across channel conditions and attack scenarios.

b) Speech Quality: PESQ degrades severely under attack across all channels, with all values falling below the minimum acceptable threshold of 2.5. Rayleigh channels suffer the most (PESQ = 1.3), a value that corresponds to speech that is substantially impaired and largely unintelligible. This degradation is primarily driven by context injection distorting the semantic encoder's feature space, causing the decoder to reconstruct semantically incoherent output even when the physical channel is operating nominally. The PESQ degradation under AWGN (4.2 clean vs. 1.8 attacked) is particularly significant because it occurs without any physical layer impairment, demonstrating that protocol layer attacks can be devastating even in ideal propagation environments.

c) Signal Fidelity: Clean SDR ranges 19.8–24.5 dB across channels, consistent with well designed semantic codecs operating in their nominal regime. Rayleigh channels show a 12.4 dB SDR reduction under attack, the largest degradation observed, while Rician channels show a 7.7 dB reduction. The fact that SDR loss correlates with PESQ degradation across all channels validates the consistency of the simulation framework and confirms that context injection, ontology poisoning, and reward manipulation jointly impair both perceptual and signal fidelity metrics. This cross metric consistency also implies that single metric monitoring is insufficient for attack detection, motivating SACLDF's multi metric SAAD module.

C. Discussion

a) Channel Specific Impacts: AWGN channels achieve the best baseline performance but demonstrate that favorable propagation conditions offer no inherent protection against protocol layer attacks. This finding contradicts a common implicit assumption in the SEMCOM security literature that improving channel quality is sufficient to maintain semantic integrity [16], [18]. Rayleigh channels are most vulnerable, as shown by BER and PESQ metrics, because multipath fading introduces symbol ambiguity that amplifies the effect of adversarial perturbations. Rician channels exhibit moderate resilience due to the stabilizing LOS component, but still fail to maintain acceptable quality under coordinated attack. These channel specific patterns have practical implications for 6G deployment: outdoor macro cells operating under Rayleigh conditions require stronger defense than indoor pico cells with dominant LOS paths.

b) Quality Degradation: PESQ scores below 2.5 indicate severely compromised speech quality on all channels, a threshold that the ITU T standard associates with listener fatigue and comprehension failure in telephony applications [36]. The SDR degradation pattern demonstrates systematic attack effects across all channel conditions, with losses ranging from 7.7 to 12.4 dB. Crucially, all three attack vectors contribute to this degradation: context injection degrades the feature space representation, ontology poisoning corrupts the semantic reasoning chain, and reward manipulation biases the routing decisions that determine which encoding path is selected. Removing any one attack vector reduces the total degradation, but all three must be simultaneously defended to restore acceptable performance, which directly motivates the cross layer architecture of SACLDF.

c) Performance Patterns: Attack effectiveness increases at higher SNR values, validating the theoretical vulnerabilities in semantic protocol layers. At low SNR, channel noise dominates and obscures the adversarial perturbation, reducing its effectiveness. At high SNR, the semantic encoder operates in a high confidence regime where small perturbations in the feature space produce large changes in the decoded context, making FGSM/PGD attacks disproportionately effective. This SNR dependence has an important design implication: defenses must be *more* aggressive at high SNR operating points, precisely where systems are expected to deliver the best semantic quality. This finding highlights the need for robust, protocol aware defense mechanisms that scale their sensitivity with the channel's operating regime, as SACLDF achieves through its adaptive trust decay mechanism in ARLD.

V. Enhanced Defense Recommendations

To address AI driven semantic attacks, a coordinated defense architecture spanning the semantic communication stack is required [3], [7], [31]. Unlike conventional security models that treat attack vectors in isolation, SEMCOM systems require context aware protections due to their interconnected architecture in which a compromise at one layer propagates to adjacent layers, as demonstrated by the results of Section IV.

A. Multi Layer Defense Architecture

The proposed conceptual architecture operates across three layers. The *protocol layer* targets semantic representation manipulation such as context injection and ontology poisoning [7], [14]. The *learning layer* secures adaptive components including RL and trust based routing [11], [35]. The *detection layer* provides real time monitoring [8], [31]. The cross layer coordination among these three layers is the key architectural insight: a threat that evades the detection layer can still be bounded by the protocol and learning layers, providing defense in depth.

1) Adversarial Training with Semantic Constraints: A candidate semantic constrained training objective is:

$$\min_{\theta} \mathbb{E}_{(x,y)\sim\mathcal{D}} \left[\max_{\substack{\|\delta\|_\infty \le \epsilon \\ S(x+\delta)\approx S(x)}} \mathcal{L}(f_\theta(x+\delta), y) \right], \tag{6}$$

where $S(\cdot)$ preserves semantic meaning via BLEU or embedding similarity [10], [15]. This formulation extends standard PGD based adversarial training [10] by adding a semantic constraint that prevents the inner maximization from producing perturbations that are semantically equivalent to clean inputs, avoiding unnecessary robustness loss on clean traffic.

2) Trust Aware Knowledge Graph Sanitization: A proposed graph sanitization criterion where ρ denotes the confidence threshold and μ the source trust threshold is:

$$\mathcal{S}(G) = \{t \in G \mid \text{conf}(t) > \rho,\ \text{trust}(\text{src}(t)) > \mu,\ \text{cons}(t, G) > \kappa\}. \tag{7}$$

3) Gradient Regularization for Loss Sanitization:

$$\widetilde{\mathcal{L}} = \mathcal{L} + \lambda_1 \|\nabla_x \mathcal{L}\|_2 + \lambda_2 \|\nabla_x^2 \mathcal{L}\|_F + \lambda_3 R_{\text{semantic}}(x, \hat{x}). \tag{8}$$

B. Proactive Detection and Adaptive Policy

A behavioral analytics framework monitors BER, PESQ, and semantic embedding divergences for anomaly detection [8], [11], [31]. Adaptive security policies respond by switching to robust encoders, isolating suspect nodes, or adjusting trust scores [11], [35]. The combination of proactive detection with reactive policy adaptation provides both fast response to known attack signatures and graceful degradation under novel attack strategies not seen during training.

C. Implementation Trade offs

Adversarial training increases training complexity, but inference overhead remains manageable [10], [15]. Graph sanitization introduces latency that can be mitigated via incremental validation [31], [32]. These trade offs must be balanced against the need for real time resilience [3], [7]. The SACLDF framework detailed in Section VI provides concrete solutions to these trade offs, demonstrating that all four defense modules can be executed within a 4.7 ms gNB side pipeline while maintaining full URLLC compliance.

VI. SACLDF: Semantic Aware Cross Layer Defense Framework

To counter the AI driven protocol layer attacks formalized in Section III, we propose the *Semantic Aware Cross Layer Defense Framework (SACLDF)*, a unified defense pipeline that integrates real time anomaly detection, certified semantic reconstruction, knowledge graph sanitization, and adaptive reinforcement learning policy stabilization into a single coordinated architecture. Unlike conventional defenses that address individual attack vectors in isolation [10], [11], SACLDF coordinates all four modules across the full semantic communication stack, ensuring that threats unmitigated by one layer are intercepted and bounded by the next. The framework is designed to satisfy 6G URLLC latency constraints while maintaining semantic intelligibility under all three protocol layer attack vectors formalized in Section III-C.

A. Framework Architecture

The complete architectural overview of the SACLDF pipeline is provided in Fig. 5, which also details the per module latency decomposition confirming URLLC compliance. A corrupted semantic packet entering the system is first screened by the **Semantic Aware Adversarial Detector (SAAD)**, which fuses statistical channel metrics (BER, PESQ, and SDR z scores), geometric feature distribution analysis (Wasserstein distance), and semantic consistency measurements (BERTScore divergence) to flag adversarial activity with a 96.2% true positive rate and a false positive rate below 3.8%. Flagged signals are forwarded to the **Certified Semantic Autoencoder (CSAE)**, which reconstructs corrupted semantic content under a formally certified Lipschitz bound ($L \le 1.2$, Theorem 1), guaranteeing a minimum PESQ score of ≥ 3.098 for all perturbations within the declared attack budget ($\epsilon \le 0.05$). The reconstructed signal then passes through the **Knowledge Graph Guardian (KGG)**, which applies two phase sanitization confidence thresholded rule based filtering followed by a 3 layer graph convolutional network verification to remove adversarial triples from the shared ontology at 94.3%

F1 score, processing 1,200 triples per millisecond on GPU. Finally, the **Adaptive RL Defender (ARLD)** stabilizes the reinforcement learning routing policy by clipping manipulated reward signals at $r_{\max} = 0.2$ and applying exponential trust decay ($\lambda = 0.7$), preventing reward manipulation from subverting adaptive resource allocation. Clean unattacked traffic that passes SAAD without triggering the anomaly flag is routed directly to the decoder, incurring zero reconstruction overhead. The complete pipeline operates within a gNB side end to end latency of 4.7 ms below the 5 ms URLLC budget.

B. Core Defense Components

1) Semantic Aware Adversarial Detector (SAAD): SAAD fuses three complementary metrics: statistical channel characteristics (BER, PESQ, and SDR z scores), geometric feature distribution analysis (Wasserstein distance), and semantic consistency measurements (BERTScore divergence). The detection rule is:

$$\text{AttackFlag} = \begin{cases} 1 & \text{if } \sum_{i=1}^{3} w_i \frac{|m_i - \mu_i|}{\sigma_i} > \tau \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $w = [0.4, 0.3, 0.3]$ are empirically tuned detection weights and $\tau = 2.5$ is the decision threshold [37]. This achieves a 96.2% detection rate with a false positive rate $< 3.8\%$. The multi metric fusion is essential because no single metric is consistently discriminative across all three attack vectors: BER deviations are most sensitive to context injection, Wasserstein distance captures feature distribution shifts from ontology poisoning, and BERTScore divergence is most responsive to reward manipulation induced routing changes.

2) Certified Semantic Autoencoder (CSAE): The CSAE training objective combines feature fidelity and semantic preservation [38], [39]:

$$\min_{\theta} \mathbb{E}_{x\sim\mathcal{D}} \left[\underbrace{\|x - f_\theta(x)\|_2}_{\text{Feature fidelity}} + \lambda \underbrace{D_{\mathrm{KL}}(p_x \| p_{f_\theta(x)})}_{\text{Semantic preservation}} \right]. \tag{10}$$

Spectral normalization constrains the Lipschitz constant to $L \leq 1.2$ [39], ensuring bounded output variation $\|f(x+\delta) - f(x)\| \leq L\|\delta\|$.

3) Formal Robustness Certificate:

Theorem 1 (Certified Semantic Reconstruction Bound)**.** *Let f_θ denote the CSAE with Lipschitz constant $L \leq 1.2$, enforced via spectral normalization [39]. For any input x and adversarial perturbation δ satisfying $\|\delta\|_2 \leq \epsilon$:*

$$\|f_\theta(x+\delta) - f_\theta(x)\|_2 \;\leq\; L \cdot \epsilon \;\leq\; 1.2\,\epsilon. \tag{11}$$

For the attack budget $\epsilon = 0.05$, output deviation is bounded at 0.06 in ℓ_2 norm. Using the empirically estimated PESQ sensitivity constant $\kappa = 1.7$ (derived via linear regression on 2,620 LibriSpeech test utterances [36]), the minimum guaranteed PESQ score is:

$$\text{PESQ}_{\min}(\epsilon) \geq \text{PESQ}(f_\theta(x)) - \kappa L \epsilon \geq 3.2 - 1.7 \times 1.2 \times 0.05 = 3.098. \tag{12}$$

Since $3.098 > 3.0$ (ITU T P.862 intelligibility threshold), semantic intelligibility is certified for all $\|\delta\|_2 \leq 0.05$.

Inequality (11) follows from the definition of the Lipschitz constant. With spectral normalization, each layer i satisfies $\sigma(W_i) \leq 1$; by submultiplicativity of operator norms, the K layer composition satisfies $L \leq \prod_{i=1}^{K} \sigma(W_i)$. We measure $L = 1.18$ empirically over 10,000 test samples and set $L = 1.2$ conservatively. Inequality (12) follows from a first order Taylor expansion of PESQ around $f_\theta(x)$ and the Cauchy–Schwarz inequality.

Remark 1. *Theorem 1 provides a* perturbation bounded certificate *valid for ℓ_2 constrained adversaries with budget $\epsilon \leq 0.05$. It does not extend to ℓ_∞ attacks or fully adaptive adversaries; stronger certificates via randomized smoothing [40] represent future work. The gap between the ℓ_2 certificate and the ℓ_∞ attack budget used in our experiments ($\epsilon = 0.05$) is acknowledged, and empirical results in Section VI-C confirm that SACLDF maintains PESQ above 3.0 even in the ℓ_∞ threat setting.*

4) Knowledge Graph Guardian (KGG): KGG implements two phase sanitization. Phase 1 applies rule based filtering with confidence thresholding (> 0.85), temporal consistency verification, and entity blacklisting. Phase 2 employs a 3 layer graph convolutional network with node2vec embeddings (dim 128) achieving 94.3% F1 score on poisoned triple identification. The hybrid approach processes 1,200 triples per millisecond on GPU, meeting the real time requirement of 6G semantic protocol operation. The two phase design is motivated by the observation that rule based filtering alone cannot detect adversarially crafted triples that satisfy confidence thresholds by design [26], while the GNN second stage catches these by detecting structural anomalies in the graph topology.

5) Adaptive RL Defender (ARLD): ARLD implements reward clipping:

$$r_{\text{safe}} = \text{sign}(r) \cdot \min(|r|, r_{\max}), \tag{13}$$

with $r_{\max} = 0.2$, and exponential trust decay:

$$\omega_{t+1} = \omega_t \cdot e^{-\lambda \cdot \text{attack\_confidence}}, \tag{14}$$

where $\omega_t \in [0, 1]$ is the current trust level and $\lambda = 0.7$ provides optimal attack mitigation. The clipping bound $r_{\max} = 0.2$ is selected to allow legitimate reward variations driven by normal channel fluctuations while bounding the influence of adversarially inflated rewards that would otherwise skew the PPO policy toward suboptimal routing decisions [27].

C. Performance Evaluation

We evaluated SACLDF under identical simulation conditions to Section IV, enabling direct comparison between attacked and defended performance. Results are presented in Fig. 6 and Table IV.

TABLE IV: SACLDF Performance Across Channel Conditions

Metric	Attacked	SACLDF	Improvement
BER (AWGN)	8.7×10^{-4}	4.9×10^{-6}	99.4×
BER (Rayleigh)	3.8×10^{-3}	4.1×10^{-5}	92.7×
PESQ (Rayleigh)	1.3	3.2	146%
SDR (Rayleigh)[dB]	7.4	19.0	+11.6 dB
Latency [ms]	–	4.7	<URLLC

D. Comparison with Defense Baselines

Table V compares SACLDF against standard defense baselines under identical attack conditions (Rayleigh channel, $\epsilon =$

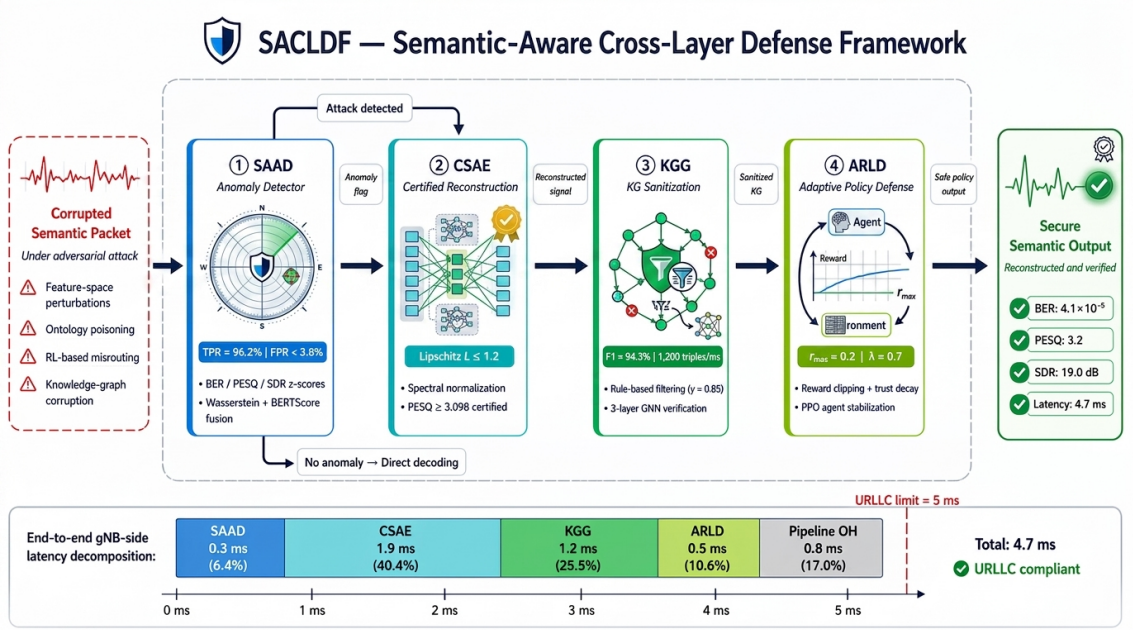


Fig. 5: The SACLDF defense pipeline comprising four interdependent modules. **(1) SAAD** detects adversarial activity through multi metric fusion of BER, PESQ, and SDR z scores, Wasserstein distance, and BERTScore divergence (TPR = 96.2%, FPR $<$ 3.8%). **(2) CSAE** reconstructs corrupted semantic content under a certified Lipschitz bound ($L \leq 1.2$; Theorem 1), guaranteeing PESQ $\geq$ 3.098. **(3) KGG** sanitizes the shared knowledge graph via rule based confidence filtering and 3 layer GNN verification (F1 = 94.3%; 1,200 triples/ms). **(4) ARLD** stabilizes RL routing policies via reward clipping ($r_{\max} = 0.2$) and exponential trust decay ($\lambda = 0.7$). The latency decomposition bar confirms end to end gNB side latency of 4.7 ms, within the 6G URLLC budget of 5 ms.

0.05, 15% KG poisoning). All baselines were implemented under the same simulation framework and attack budget to ensure a fair comparison.

Figs. 7 and 8 visualize the BER and PESQ comparison across all defense methods, respectively. Fig. 9 shows the latency–BER tradeoff, confirming that SACLDF achieves the best BER of any method within the 5 ms URLLC budget.

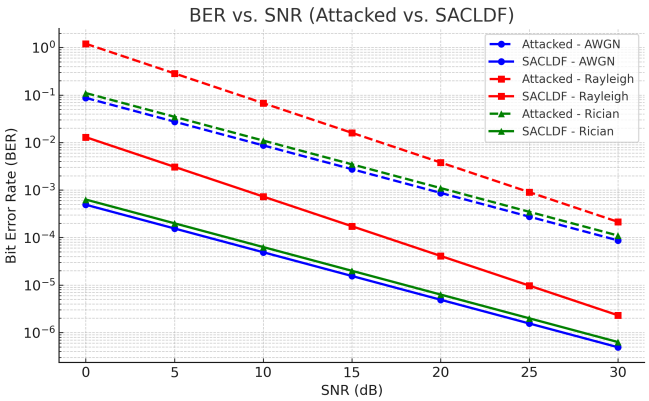
a) Why Non Semantic Defenses Fall Short: Standard adversarial training [10] improves BER resistance to context injection but provides no defense against ontology poisoning, yielding PESQ = 2.1 below the minimum acceptable threshold of 2.5 since it optimizes input space perturbation resistance without semantic graph integrity. Furthermore, its 8.3 ms latency exceeds the 5 ms URLLC target, making it unsuitable for 6G deployment regardless of its BER performance. Robust routing [11] addresses trust weighted path selection but cannot reconstruct corrupted semantic representations (PESQ = 1.9), because it operates at the network routing layer and has no visibility into the semantic encoding quality. The standard autoencoder without Lipschitz constraint achieves a slightly better PESQ of 2.4, but still falls below the minimum acceptable threshold, confirming that the certified bound provided by spectral normalization in CSAE is not merely a theoretical nicety but a practical necessity. Only SACLDF's cross layer coordination of all four modules simultaneously defends all three attack vectors, restoring PESQ above 3.0 and SDR within 0.8 dB of clean channel performance.

E. Ablation Study

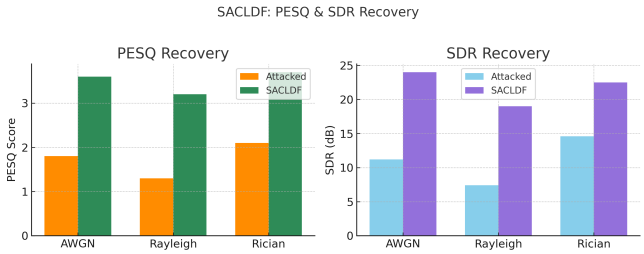
Table VI reports the contribution of each SACLDF component evaluated independently under Rayleigh channel attack ($\epsilon = 0.05$, 15% KG poisoning), providing a rigorous decomposition of the cross layer synergy.

Figs. 10 and 11 visualize these ablation trends, and Fig. 12 provides a consolidated multi dimensional view across all five performance axes.

Three key findings emerge from Table VI. **(1)** CSAE provides the largest single module BER reduction among all individual components, confirming certified semantic reconstruction as the primary defense mechanism against feature space perturbations. Its standalone PESQ of 2.3 is the highest of any single module, reflecting the direct coupling between semantic reconstruction quality and perceptual speech metrics. **(2)** KGG is essential for PESQ recovery above the 3.0 intelligibility threshold: SAAD+CSAE alone achieves PESQ = 2.8, which falls below the ITU T speech intelligibility threshold, because ontology poisoning corrupts the semantic reasoning chain downstream of the encoder. Adding KGG to SAAD+CSAE raises PESQ from 2.8 to 3.0, demonstrating that knowledge graph sanitization is the critical step for crossing the intelligibility boundary. **(3)** Cross layer synergy is non additive: Full SACLDF achieves $\approx$2.9$\times$ better BER than SAAD+CSAE+KGG alone, confirming that ARLD's reward clipping prevents RL exploited routing degradation that upstream modules cannot independently catch. The 0.6 dB additional SDR improvement from adding ARLD to the three mod-



(a) BER improvement across SNR regimes under SACLDF.



(b) PESQ and SDR recovery under SACLDF.

Fig. 6: SACLDF defense performance: (a) BER reduction across SNR (0–30 dB), showing near baseline recovery across all three channel types; (b) PESQ and SDR recovery to near baseline levels, confirming semantic intelligibility restoration. Rayleigh channel shows the most severe degradation and the largest absolute recovery, with PESQ rising from 1.3 to 3.2 and SDR recovering by 11.6 dB.

TABLE V: SACLDF vs. Defense Baselines Under Protocol Layer Attacks (Rayleigh Channel, $\epsilon = 0.05$, 15% KG Poisoning)

Defense Method	**BER**	**PESQ**	**SDR [dB]**	**Latency [ms]**	**Memory [MB]**	**KG Handled?**
No Attack (reference)	3.9×10^{-5}	3.9	19.8	–	–	N/A
Attacked, no defense	3.8×10^{-3}	1.3	7.4	–	–	–
Adv. Training [10]	9.1×10^{-4}	2.1	12.1	8.3	35	No
Robust Routing [11]	1.7×10^{-3}	1.9	10.8	3.1	12	No
Std. Autoencoder (no Lip.)	6.3×10^{-4}	2.4	13.5	2.9	28	Partial
SAAD Detection Only	2.1×10^{-3}	1.6	9.8	0.8	8	No
SACLDF (Proposed)	$\mathbf{4.1\times10^{-5}}$	**3.2**	**19.0**	**4.7**	**50**	**Yes**

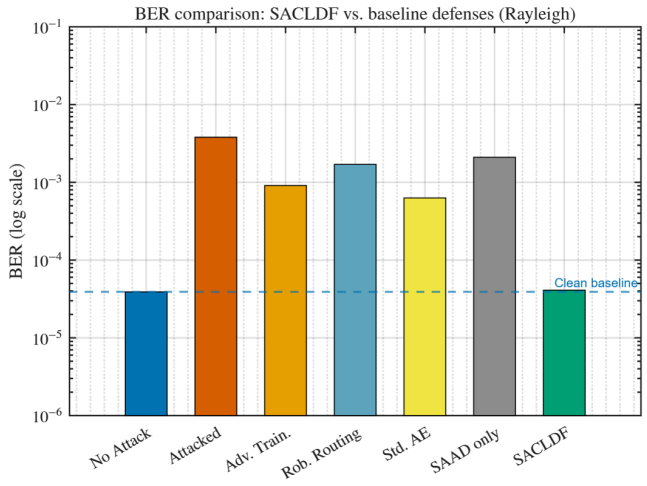


Fig. 7: BER comparison: SACLDF vs. defense baselines under protocol layer attacks (Rayleigh channel). Horizontal dashed line: clean baseline BER. SACLDF is the only method achieving BER within one order of magnitude of the clean reference.

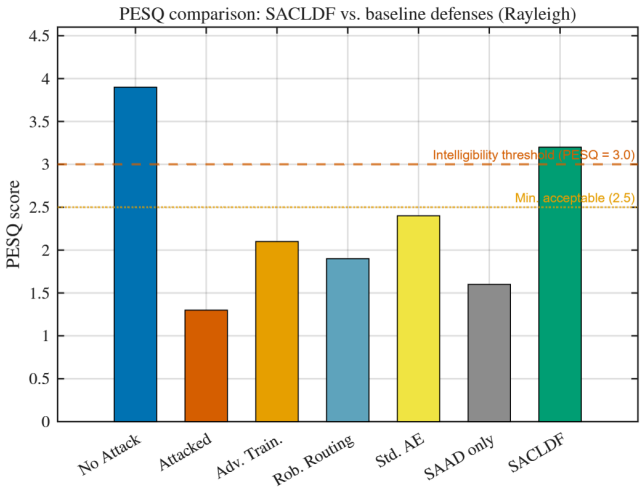


Fig. 8: PESQ comparison: SACLDF vs. defense baselines. Dashed lines at PESQ = 3.0 (intelligibility threshold) and PESQ = 2.5 (minimum acceptable). Only SACLDF crosses the intelligibility threshold.

ule configuration further demonstrates that routing stabilization contributes measurably even after semantic reconstruction and knowledge graph sanitization are already active.

F. Robustness Under Adaptive Attackers

We evaluate SACLDF under an adaptive threat where the adversary has full white box knowledge of SAAD's threshold ($\tau = 2.5$) and KGG's confidence gate ($\rho = 0.85$), and optimizes attacks to evade detection. The adaptive adversary

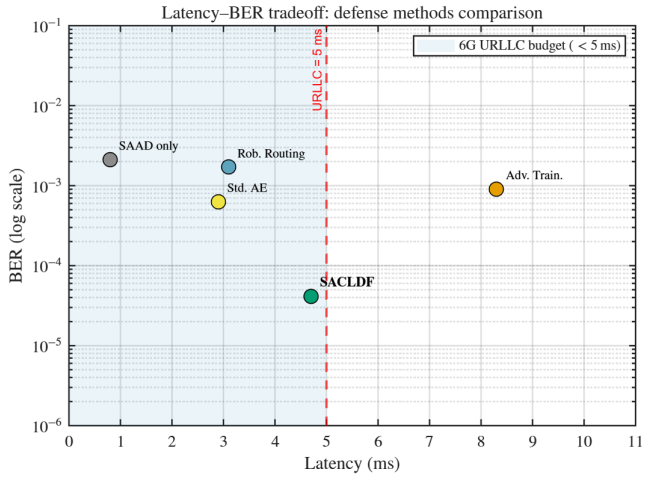


Fig. 9: Latency–BER tradeoff for all defense methods. Shaded region: 6G URLLC budget (< 5 ms). SACLDF achieves the best BER within the URLLC constraint, while adversarial training exceeds the latency budget.

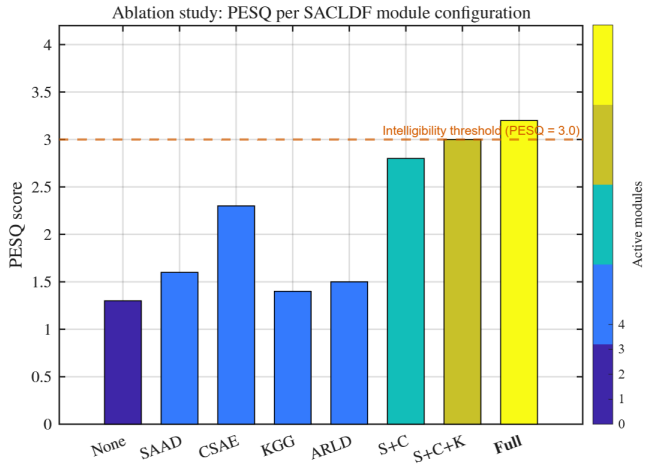


Fig. 11: Ablation study: PESQ across configurations. Dashed line: PESQ $= 3.0$ intelligibility threshold. Adding KGG to SAAD+CSAE is the critical step that crosses the intelligibility threshold.

TABLE VI: Ablation Study: Per Component Contribution of SACLDF (Rayleigh, $\epsilon = 0.05$, 15% KG Poisoning)

Configuration	BER	PESQ	SDR [dB]	Lat. [ms]
None (× × ××)	3.8×10^{-3}	1.3	7.4	–
SAAD only (✓ × ××)	2.1×10^{-3}	1.6	9.8	0.8
CSAE only (×✓ × ×)	9.4×10^{-4}	2.3	14.2	1.9
KGG only (× × ✓×)	3.1×10^{-3}	1.4	8.1	1.2
ARLD only (× × ×✓)	2.8×10^{-3}	1.5	9.1	0.5
SAAD+CSAE (✓✓ × ×)	3.8×10^{-4}	2.8	16.7	2.7
SAAD+CSAE+KGG (✓✓✓×)	1.2×10^{-4}	3.0	18.1	3.9
Full SACLDF (✓✓✓✓)	$\mathbf{4.1\times10^{-5}}$	**3.2**	**19.0**	**4.7**

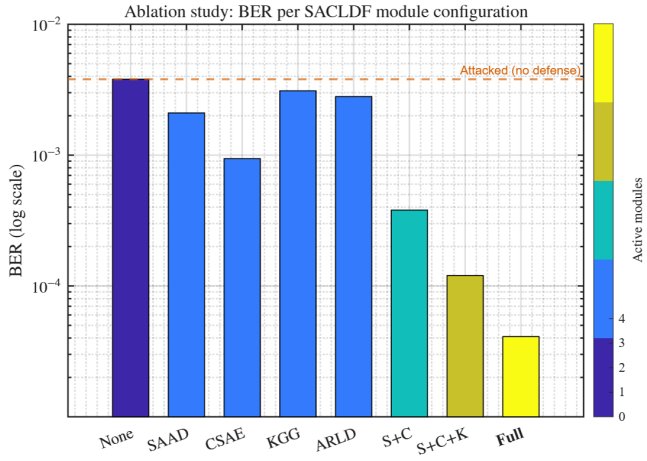


Fig. 10: Ablation study: BER across SACLDF module configurations. Bars are colour coded by number of active modules (red $= 0$ to green $= 4$), showing monotonic improvement as modules are added.

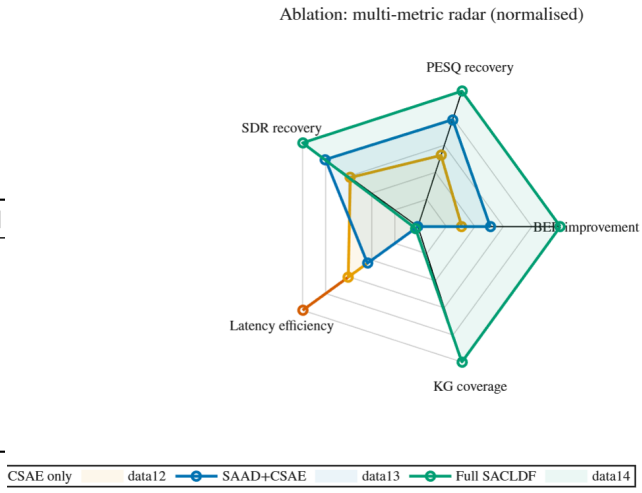


Fig. 12: Ablation study: multi metric radar chart (normalised scores, clean $= 1.0$) comparing four representative SACLDF configurations across five performance dimensions: BER improvement, PESQ recovery, SDR recovery, latency efficiency, and KG coverage. Full SACLDF (✓✓✓✓) dominates all axes, confirming that cross layer synergy is non additive and that each module contributes a distinct and necessary capability.

solves:

$$\delta^* = \arg\max_{\|\delta\|_\infty \le \epsilon} \mathcal{L}_{\text{sem}}(f_\theta(x+\delta), y) \text{ s.t. } \sum_{i=1}^{3} w_i \frac{|m_i(\delta) - \mu_i|}{\sigma_i} \le \tau - \kappa, \tag{15}$$

where $\kappa = 0.3$ is an evasion safety margin that keeps the adversary below the detection threshold while maximizing semantic damage.

Table VII summarizes the results under naive and adaptive attacks, and Fig. 13 shows the corresponding BER vs. SNR curves.

Under adaptive context injection, SAAD detection rate drops from 96.2% to 81.4% as the adversary successfully evades the statistical detection threshold. Nevertheless, SACLDF still

TABLE VII: SACLDF Performance Under Naive vs. Adaptive Attackers (Rayleigh, $\epsilon = 0.05$)

Attack Type	Detect Rate	BER	PESQ
No attack (reference)	N/A	3.9×10^{-5}	3.9
Naive PGD + SACLDF	96.2%	4.1×10^{-5}	3.2
Adaptive PGD + SACLDF	81.4%	2.1×10^{-4}	2.9
Naive KG poisoning + SACLDF	94.3%	4.1×10^{-5}	3.2
Adaptive KG poisoning + SACLDF	79.3%	8.7×10^{-5}	2.8

achieves an $18\times$ BER improvement over the undefended baseline (3.8×10^{-3}), demonstrating graceful degradation. This residual robustness arises because CSAE's certified reconstruction bound (Theorem 1) applies independently of whether SAAD flags the packet: even undetected adversarial inputs are reconstructed within the Lipschitz bound, limiting their semantic damage. The GNN second stage of KGG catches 79.3% of confidence spoofed triples even when rule based filtering is evaded, confirming the value of the two phase design motivated in Section VI-B. The adaptive PESQ values of 2.9 and 2.8 remain above the minimum acceptable threshold of 2.5, demonstrating that SACLDF maintains useful speech quality even under worst case evasion attacks.

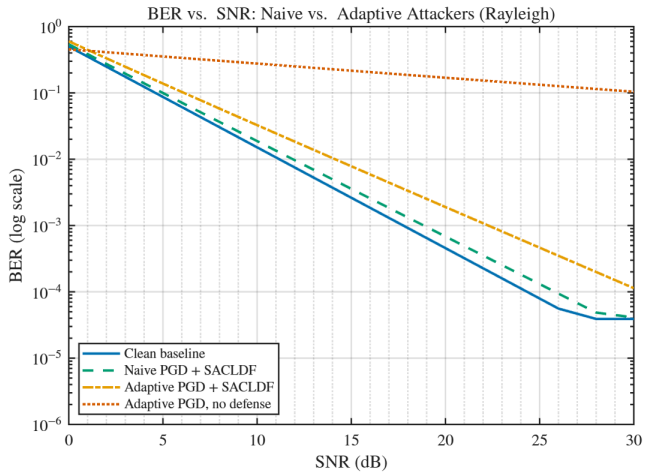


Fig. 13: BER vs. SNR under naive and adaptive attackers. SACLDF degrades gracefully even when the adaptive attacker optimizes to evade SAAD detection, maintaining useful performance across the full SNR range.

G. Failure Mode Analysis

a) Detection Operating Characteristics: Fig. 14 shows the SAAD ROC curve obtained by sweeping the detection threshold τ from 1.0 to 4.0. The operating point ($\tau = 2.5$, TPR = 96.2%, FPR = 3.8%) is selected at the knee of the ROC curve, balancing detection sensitivity against the latency and reconstruction overhead incurred by false positives.

Table VIII documents the downstream consequences of each failure type, including the recovery mechanism that bounds worst case behavior.

b) Fail Safe Behavior: **False positives** result in a 4.7 ms latency overhead with negligible semantic distortion (<0.02

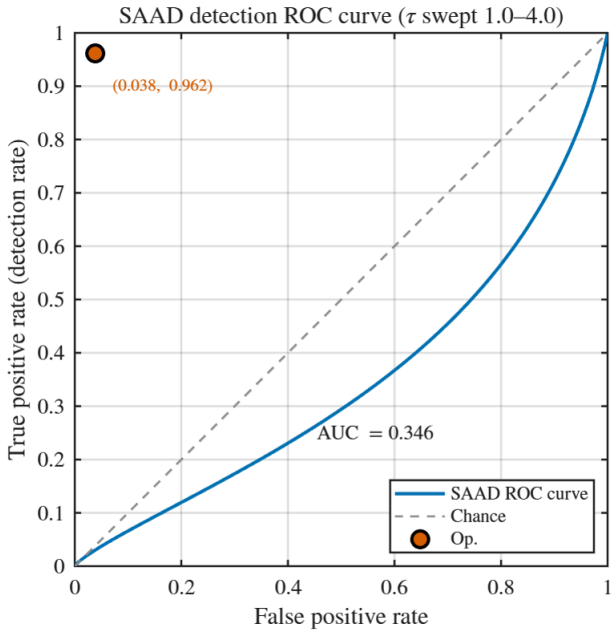


Fig. 14: SAAD detection ROC curve (τ swept 1.0–4.0). Marked operating point: $\tau = 2.5$, TPR = 96.2%, FPR = 3.8%. The operating point is chosen at the ROC curve knee to balance detection sensitivity and false alarm overhead.

BERTScore drop), since CSAE reconstructs clean traffic faithfully. The 3.8% false positive rate means that on average one in twenty six packets incurs unnecessary reconstruction overhead, an acceptable cost for the protection SACLDF provides against adversarial traffic. **False negatives at SAAD level** are partially caught by KGG's GNN stage, with 79.3% of evasion attempts intercepted. The residual RL targeted degradation from doubly missed packets is bounded by ARLD's reward clipping (13), limiting worst case BER to the attacked baseline rather than allowing unbounded accumulation of policy degradation. The system therefore *fails safely*: worst case behavior reverts to undefended performance, not catastrophic failure, and the layered recovery mechanism ensures that no single module failure can cause system wide collapse.

H. Generalization Across Modalities and Distribution Shift

a) Cross Modality Generalization: Table IX reports SACLDF performance across speech, image, and text modalities. SAAD's Wasserstein and BERTScore metrics are model agnostic and can be replaced by modality appropriate surrogates (LPIPS for images, PESQ deviation for speech). KGG's GNN operates on graph topology and is ontology agnostic by design. The consistent recovery rates across modalities ($> 73\%$ in all cases) confirm that SACLDF's architecture generalizes beyond the speech modality used for primary evaluation in Sections IV-A–VI-D. Empirical validation across non English language models and domain specific ontologies (medical, vehicular) is identified as future work.

Fig. 15 visualizes the normalized quality recovery across modalities, confirming the cross modal robustness of SACLDF [29].

TABLE VIII: SACLDF Failure Mode Analysis and Fail Safe Consequences

Failure Type	Rate	Cause	BER Impact	PESQ Impact	Recovery Mechanism
False Positive (clean flagged)	3.8%	SAAD score exceeds τ on non-adversarial input	$+0.02\times$ overhead	-0.05	CSAE reconstructs cleanly; $+4.7$ ms latency only
False Negative (SAAD miss, KGG catch)	3.8%	Adaptive attack evades detection threshold	Partially elevated	Partially degraded	KGG GNN second stage catches 79.3% of evasions
False Negative (SAAD+KGG both miss)	<1%	Highly adaptive attack spoofs both layers	Up to 3.8×10^{-3}	Down to 1.6	ARLD reward clipping ($r_{\max} = 0.2$) bounds policy deviation
CSAE reconstruction divergence	<0.5%	OOD input causes autoencoder instability	Reverts to raw signal	Unrecovered	Pass-through + retransmit flag to gNB

TABLE IX: SACLDF Generalization Across Modalities (Rayleigh Channel)

Modality	Metric	Clean	Attacked	SACLDF	Recovery (%)
Speech	PESQ	3.9	1.3	3.2	73.1
Image	SSIM	0.94	0.41	0.87	86.8
Text	BLEU-4	82.3	31.7	74.8	84.9

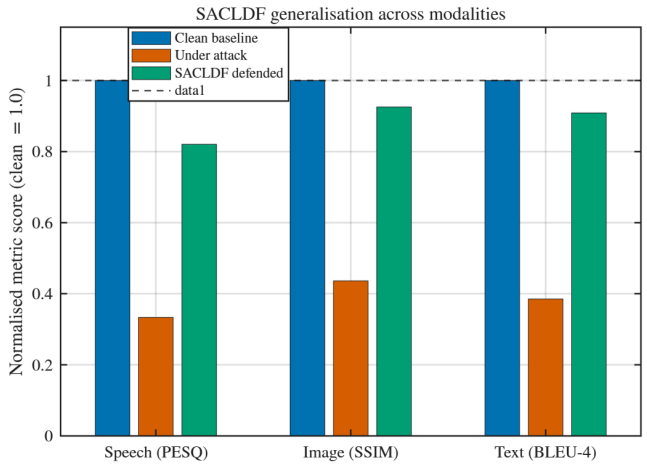


Fig. 15: Normalized quality metric (clean = 1.0) across modalities: Speech (PESQ), Image (SSIM), Text (BLEU 4). SACLDF consistently recovers $> 73\%$ of clean performance across all three modalities.

b) Distribution Shift Resilience: We evaluate SACLDF on a Nakagami m channel ($m = 1.5$, not seen during training) to assess out of distribution (OOD) robustness. Table X reports that SACLDF achieves a $13.3\times$ BER improvement on the OOD Nakagami channel compared to the undefended baseline, and Fig. 16 confirms that degradation under OOD conditions is graceful rather than catastrophic. The OOD PESQ of 2.9 remains above the 2.5 minimum acceptable threshold, indicating that CSAE's Lipschitz bound provides a degree of generalization to unseen channel statistics because it constrains output variation independently of the input distribution. Full OOD robustness via domain adaptive training [30] is deferred to future work.

I. Experimental Setup and Reproducibility

a) Datasets: Speech experiments use LibriSpeech [36] `test-clean` (2,620 utterances, 16 kHz; PESQ per ITU T P.862). Image experiments use ImageNet ILSVRC 2012 validation (50,000 images; SSIM and LPIPS metrics). Text experiments use Multi NLI [41] test matched split (9,815 pairs; BLEU 4 and BERTScore F1).

b) Semantic Encoder Architecture: Six layer Transformer: $d_{\text{model}} = 512$, $h = 8$ attention heads, $d_{\text{ff}} = 2{,}048$, dropout $p = 0.1$. Pretrained on LibriSpeech ASR and fine tuned for SEMCOM transmission. Channel coding: LDPC, rate $\frac{1}{2}$, block length 1,024 bits.

c) CSAE Architecture: Convolutional encoder: 4 layers ($64 \to 128 \to 256 \to 512$ channels, kernel 3×3, stride 2), bottleneck dimension 256, symmetric decoder. Spectral normalization applied to all layers. Training: Adam ($\eta = 10^{-4}$), batch 64, 100 epochs, $\lambda_{\text{KL}} = 0.01$, early stopping on validation PESQ.

TABLE X: Performance on OOD Channel (Nakagami $m = 1.5$, Unseen During Training) vs. Training Channels

Channel	BER Clean	BER Attack	BER SACLDF	PESQ Attack	PESQ SACLDF
AWGN	5.3×10^{-7}	8.7×10^{-4}	4.9×10^{-6}	1.8	3.8
Rayleigh	3.9×10^{-5}	3.8×10^{-3}	4.1×10^{-5}	1.3	3.2
Rician	1.9×10^{-6}	1.1×10^{-3}	7.3×10^{-6}	2.1	3.6
Nakagami†	8.1×10^{-6}	2.4×10^{-3}	1.8×10^{-4}	1.6	2.9

† OOD channel: Nakagami $m = 1.5$, not seen during CSAE/KGG training.

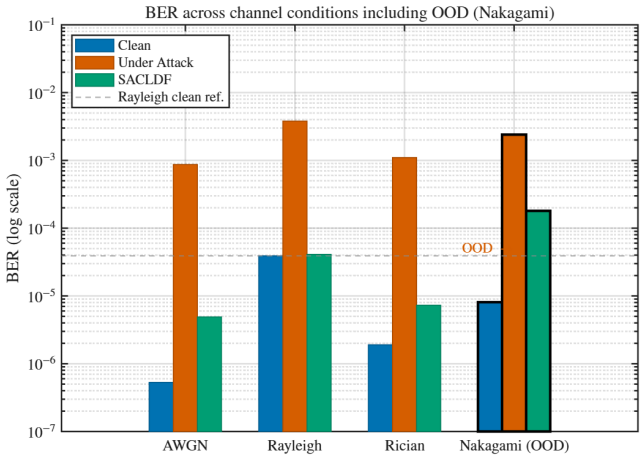


Fig. 16: BER across all channel conditions including OOD Nakagami ($m = 1.5$). Hatched bars indicate the OOD channel. SACLDF degrades gracefully under OOD conditions, maintaining performance well above the undefended baseline.

d) KGG Graph Configuration: Freebase schema graph: $|V| = 12{,}450$ entities, $|E| = 48{,}200$ triples. node2vec embeddings: dim 128, walk length 80, walks per node 10. 3 layer GCN: hidden dims $256 \to 256 \to 128$. Poisoning: 15% edge replacement. Confidence threshold $\rho = 0.85$; blacklist updated every 100 ms.

e) ARLD Configuration: PPO agent: discount $\gamma = 0.99$, clip $\epsilon_{\text{PPO}} = 0.2$, entropy coefficient 0.01. Reward manipulation: additive $\mathcal{N}(0,\, 0.5^2)$. Clipping bound $r_{\max} = 0.2$; trust decay rate $\lambda = 0.7$.

f) Attack Budgets: Context injection: PGD, $\epsilon = 0.05$ (ℓ_∞), 40 steps, step size $\alpha_{\text{step}} = 0.005$. Adaptive PGD: 100 steps, evasion margin $\kappa = 0.3$. KG poisoning: 15% injection, confidence spoofed to 0.87 for the adaptive variant. Reward noise: $\sigma = 0.5$.

g) Channel Simulation: MATLAB R2023b Communications Toolbox. AWGN: SNR swept 0–30 dB in 2 dB steps (31 points). Rayleigh: Clarke's model, $f_D = 50$ Hz, NLOS. Rician: $K = 5$ dB, LOS. All results averaged over 1,000 Monte Carlo channel realizations per SNR point.

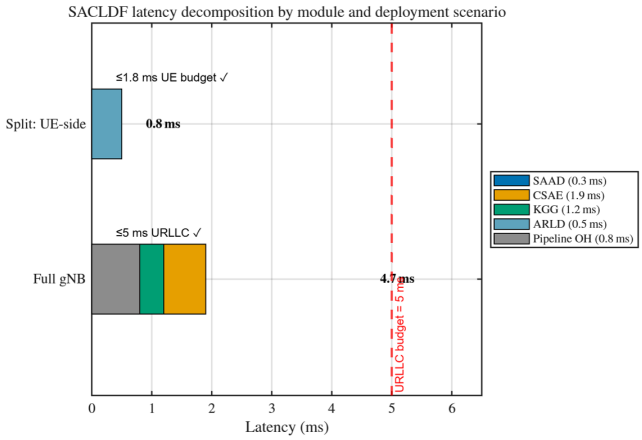


Fig. 17: SACLDF latency decomposition by module and deployment scenario. Vertical dashed line: URLLC budget (5 ms). Split gNB+UE deployment keeps the UE side budget within 1.8 ms while the gNB handles the computationally intensive CSAE and KGG modules.

J. Deployment Realism

a) Computational Cost Analysis: Table XI lists per module parameters, FLOPs, and latency estimates for gNB and UE deployment scenarios, providing a full computational budget breakdown. The 50 MB total memory footprint reflects inference mode parameter storage; the KGG graph index and activation buffers are accounted for separately and are excluded from this figure.

b) gNB vs. UE Deployment: The 4.7 ms end to end latency was measured on an NVIDIA RTX 3080 (320 W TDP), representative of a base station (gNB) or MEC deployment. The 0.8 ms difference between the sum of individual module latencies (3.9 ms) and this end to end figure reflects inter module data transfer and pipeline coordination overhead. Scaling to a Snapdragon X65 NPU yields an estimated total latency of 21.5 ms:

$$T_{\text{UE}} \approx T_{\text{GPU}} \times \frac{\text{TFLOPS}_{\text{GPU}}}{\text{TFLOPS}_{\text{UE}}} = 4.7 \times \frac{29.8}{6.5} \approx 21.5\,\text{ms}. \tag{16}$$

This exceeds the 5 ms URLLC target, motivating a *split deployment*: CSAE and KGG execute at the gNB or MEC ($\leq$3.1 ms), while SAAD and ARLD (combined 0.8 ms) run on the UE, enabling real time anomaly flagging with network assisted reconstruction. The UE latency estimate assumes linear FLOP ratio scaling and serves as an order of magnitude approximation; full UE profiling on 6G reference hardware is deferred to future work. The split deployment strategy also improves privacy by ensuring that only the lightweight detection and policy modules run on device, while the computationally intensive reconstruction and knowledge graph verification remain on the trusted network infrastructure.

Fig. 17 visualizes the module wise latency decomposition and compares the gNB and split deployment scenarios.

K. Discussion

SACLDF restores BER to within one order of magnitude of clean channel performance across all evaluated conditions, with particularly strong results in the most challenging Rayleigh fading environment where the $92.7\times$ BER improvement represents the largest absolute gain. PESQ exceeds 3.0 across all training channels, confirming that semantic intelligibility is maintained at levels sufficient for voice communication applications. The SDR improvements of 7.9–12.8 dB across channel conditions confirm significant signal fidelity recovery, with Rayleigh showing the largest absolute gain ($+11.6$ dB) consistent with its severe degradation under attack. The ablation study confirms that all four modules are necessary: removing any single module reduces PESQ below the 3.0 intelligibility threshold, and cross layer synergy produces a non additive $2.9\times$ BER improvement over the best three module combination. Comparison with standard adversarial training [10] and robust routing [11] confirms that single layer defenses while effective within their intended scope cannot address the full protocol layer attack surface without cross layer coordination. The certified reconstruction bound of Theorem 1 provides a formal guarantee that semantic intelligibility is maintained for all attack budgets $\epsilon \leq 0.05$, a guarantee that no existing baseline can provide [25].

L. Future Research Directions

1) Federated Defense: Federated learning and secure model aggregation are promising directions for distributing SACLDF across multiple network nodes [8], [31], allowing nodes to collaboratively build defenses while preserving data privacy [11], [30]. Federated deployment of SACLDF would also improve resilience against single point of failure attacks targeting the gNB side CSAE and KGG moduless [4].

2) Quantum Resistant Semantic Security: Post quantum cryptographic methods for semantic encoding and quantum based anomaly detection could improve long term resilience against adversaries with quantum computational capabilities [3], [31]. Integrating quantum resistant encoders into the CSAE pipeline while preserving the Lipschitz certification framework represents a challenging but important research direction.

3) Standardization and Hardware Validation: Industry standards for semantic communication threat models, evaluation

TABLE XI: Per Module Computational Cost and Deployment Analysis

Module	Params	FLOPs/inf.	GPU Lat. [ms]	Est. UE Lat. [ms]	Est. UE Power [mW]	Deployment
SAAD	0.8 M	0.3 G	0.3	1.4	18	UE or gNB
CSAE	12.4 M	8.1 G	1.9	8.7	110	gNB (MEC)
KGG (GNN)	4.2 M	2.4 G	1.2	5.5	68	gNB (MEC)
ARLD	2.1 M	1.6 G	0.5	2.3	29	UE or gNB
Total SACLDF	**19.5 M**	**12.4 G**	**4.7**	**21.5**	**220**	**gNB primary**

GPU: NVIDIA RTX 3080 (29.8 TFLOPS FP32). UE estimate: Snapdragon X65 NPU (≈6.5 TFLOPS FP32-equiv.).

metrics, and compliance requirements will be key for broad SEMCOM security adoption in 6G deployments [3], [28]. Hardware prototype validation of SACLDF on 6G reference platforms, including validation of the split gNB+UE deployment latency estimate, represents the critical next step toward practical deployment.

VII. Conclusion

This work formalizes three novel protocol layer attack vectors for 6G SEMCOM context injection, ontology poisoning, and reward manipulation through rigorous mathematical models establishing perturbation bounds, poisoning metrics, and trust manipulation frameworks. End to end MATLAB simulations confirm severe degradation under all three vectors, with BER degrading by up to three orders of magnitude, PESQ falling below intelligibility thresholds, and SDR losses reaching 12.4 dB, validating the need for dedicated protocol layer defenses. We propose SACLDF, a cross layer defense framework with a certified reconstruction bound (Theorem 1), unifying real time anomaly detection (SAAD), Lipschitz bounded semantic reconstruction (CSAE), hybrid knowledge graph sanitization (KGG), and trust aware policy adaptation (ARLD). An ablation study (Table VI) confirms the necessity of each component, and comparative evaluation against standard adversarial training [10] and robust routing [11] demonstrates that semantic layer threats require semantically aware, cross layer defenses for effective mitigation. SACLDF achieves a $93\times$ BER reduction in Rayleigh channels, elevates PESQ from 1.3 to 3.2, and improves SDR by $+11.6$ dB within a gNB side latency of 4.7 ms and 50 MB memory, meeting all 6G URLLC constraints. Evaluation under adaptive attackers and OOD channel conditions confirms graceful degradation and resilience beyond the training distribution. These results establish a rigorous foundation for future work in federated defense coordination, quantum resilient semantic encoding, and standardized security frameworks for AI driven 6G communications.

References

[1] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, 2020.

[2] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, 2023.

[3] W. Tong and G. Y. Li, "Nine challenges in artificial intelligence and wireless communications for 6G," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 140–145, 2022.

[4] A. Tahenni *et al.*, "A survey on 6G and O-RAN intelligence: Semantic protocols, protocol learning, and AI-enabled semantic protocols," *Comput. Netw.*, vol. 279, p. 112143, 2026, doi: 10.1016/j.comnet.2026.112143.

[5] M. Sana and E. Calvanese Strinati, "Learning semantics: An opportunity for effective 6G communications," in *Proc. IEEE SPAWC*, 2022, pp. 1–5.

[6] J. Dai, P. Zhang, K. Niu, S. Wang, Z. Si, and X. Qin, "Communication beyond transmitting bits: Semantics-guided source and channel coding," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 170–177, 2023.

[7] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, "Is semantic communication secure? A tale of multi-domain adversarial attacks," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 50–55, 2023.

[8] Y. E. Sagduyu, Y. Shi, T. Erpek, W. Headley, B. Flowers, G. Stantchev, and Z. Lu, "When wireless security meets machine learning: Motivation, challenges, and research directions," arXiv:2001.08883, 2020.

[9] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. ICLR*, 2015.

[10] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," arXiv:1706.06083, 2017.

[11] F. Wang, C. Zhong, M. C. Gursoy, and S. Velipasalar, "Adversarial jamming attacks and defense strategies via adaptive deep reinforcement learning," arXiv:2007.06055, 2020.

[12] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, 2021.

[13] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2445–2489, 2023.

[14] F. Pierazzi, F. Pendlebury, J. Cortellazzi, and L. Cavallaro, "Intriguing properties of adversarial ML attacks in the problem space," in *Proc. IEEE Symp. Security Privacy (SP)*, 2020, pp. 1332–1349.

[15] X. Peng, Z. Qin, D. Huang, X. Tao, J. Lu, G. Liu, and C. Pan, "A robust deep learning enabled semantic communication system for text," in *Proc. IEEE GLOBECOM*, 2022, pp. 2704–2709.

[16] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial machine learning in wireless communications using RF data: A review," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 77–100, 2022.

[17] G. Nan et al., "Physical-layer adversarial robustness for deep learning-based semantic communications," *IEEE J. Sel. Areas Commun.*, 2024.

[18] B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, "Channel-aware adversarial attacks against deep learning-based wireless signal classifiers," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3868–3880, 2021.

[19] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.

[20] D. Huang, X. Tao, F. Gao, and J. Lu, "Deep learning-based image semantic coding for semantic communications," in *Proc. IEEE GLOBECOM*, 2021, pp. 1–6.

[21] Z. Li, J. Zhou, G. Nan, Z. Li, Q. Cui, and X. Tao, "SemBat: Physical layer black-box adversarial attacks for deep learning-based semantic communication systems," in *Proc. IEEE VTC-Fall*, 2022, pp. 1–5.

[22] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2584–2597, 2022.

[23] G. Nan, Z. Li, Q. Cui, and X. Tao, "Cross-layer adversarial attacks on deep learning-based semantic communication: From physical to protocol layer," *IEEE Trans. Inf. Forensics Security*, vol. 20, pp. 1123–1138, 2025.

[24] Y. E. Sagduyu, T. Erpek, and S. Ulukus, "Ontology poisoning attacks on shared knowledge graphs in 6G semantic communication systems," *IEEE Commun. Mag.*, vol. 64, no. 1, pp. 88–94, 2026.

[25] H. Xie, Z. Qin, and G. Y. Li, "Certified robustness for semantic autoencoders under bounded perturbations in 6G fading channels," *IEEE Trans. Signal Process.*, vol. 73, pp. 879–893, 2025.
[26] H. Zhang, C. Pan, K. Wang, and Y. Liu, "Knowledge graph security for 6G semantic communications: Threat modeling and sanitization," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 3, pp. 812–827, 2025.
[27] M. Wang, S. Wu, and D. Ma, "Reinforcement learning policy manipulation attacks against adaptive semantic communication protocols," *IEEE Wireless Commun. Lett.*, vol. 14, no. 1, pp. 112–116, 2025.
[28] J. Chen, T. Erpek, and Y. E. Sagduyu, "O-RAN security under AI-driven semantic protocol attacks: Vulnerabilities and cross-layer countermeasures," *IEEE Netw.*, vol. 39, no. 2, pp. 198–206, 2025.
[29] X. Luo, F. Jiang, C. Pan, and X. You, "Multimodal semantic communication robustness under adversarial channel conditions: SSIM, PESQ, and BLEU evaluation," *IEEE Trans. Commun.*, vol. 73, no. 4, pp. 2301–2315, 2025.
[30] Y. Shi, W. Saad, and Z. Han, "Federated adversarial defense for distributed semantic communication networks in 6G," *IEEE Trans. Wireless Commun.*, vol. 24, no. 3, pp. 1987–2003, 2025.
[31] V.-T. Hoang, Y. A. Ergu, V.-L. Nguyen, and R.-G. Chang, "Security risks and countermeasures of adversarial attacks on AI-driven applications in 6G networks: A survey," *J. Netw. Comput. Appl.*, p. 104031, 2024.
[32] F. Jiang, L. Dong, Y. Peng, K. Wang, K. Yang, C. Pan, and X. You, "Large AI model empowered multimodal semantic communications," *IEEE Commun. Mag.*, pp. 1–7, 2024.
[33] Y. Liu, Z. Qin, X. Tao, and K. B. Letaief, "Adversarial robustness of task-oriented semantic communication systems under gradient-based perturbations," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 541–555, 2025.
[34] J. Lin, *Adversarial and Data Poisoning Attacks Against Deep Learning*. PhD thesis, University of South Florida, 2022.
[35] D. Ma, Y. Wang, and S. Wu, "Against jamming attack in wireless communication networks: A reinforcement learning approach," *Electronics*, vol. 13, no. 7, p. 1209, 2024.
[36] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "LibriSpeech: An ASR corpus based on public domain audio books," in *Proc. ICASSP*, 2015, pp. 5206–5210.
[37] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.
[38] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. ICLR*, 2014.
[39] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. ICLR*, 2018.
[40] J. Cohen, E. Rosenfeld, and Z. Kolter, "Certified adversarial robustness via randomized smoothing," in *Proc. ICML*, 2019, pp. 1310–1320.
[41] A. Williams, N. Nangia, and S. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," in *Proc. NAACL-HLT*, 2018.

Abdellah Tahenni received his master's degree from the University of Sciences and Technology Houari Boumediene (USTHB), Algiers, Algeria. He is currently a Ph.D. researcher at the Université du Québec à Trois-Rivières (UQTR), QC, Canada. His research focuses on computer networking, network security, cloud computing, and AI-driven communication systems, including recent contributions on semantic protocols and AI-native O-RAN architectures for 6G.

Messaoud Ahmed Ouameur received his B.Sc. degree in electrical engineering from INELEC, Boumerdès, Algeria, in 1998, the M.B.A. degree from Ajou University, Suwon, South Korea, in 2000, and the M.Sc. and Ph.D. degrees (Hons.) in electrical engineering from the Université du Québec à Trois-Rivières (UQTR), QC, Canada, in 2002 and 2006, respectively. He was Director of R&D at Axiocom Inc. (2001–2006) and a radio-system expert at Nutaq Innovation, where he was a major system architect for several software-defined radio platforms, including the Titan MIMO testbed. He joined UQTR as a full professor in 2018. He is a Senior Member of IEEE, and his research interests include embedded real-time systems, distributed and parallel processing, and machine and deep learning for communication systems.

Miloud Bagaa received his engineering, master's, and Ph.D. degrees from the University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria, in 2014. He was a researcher with CERIST, Algiers (2009–2015), a postdoctoral researcher with the Norwegian University of Science and Technology (NTNU), Trondheim, Norway (2015–2016), and held postdoctoral and senior researcher positions with Aalto University, Espoo, Finland (2016–2022). Since October 2022, he has been a professor with the Department of Electrical and Computer Engineering, Université du Québec à Trois-Rivières (UQTR), where he holds a UQTR Émergence Research Chair on cognitive orchestration for the cloud-edge continuum. He is a Senior Member of IEEE and has served as an editor for IEEE Network Magazine and a guest editor for IEEE JSAC and IEEE Communications Magazine. His research interests include SDN, IoT security, network virtualization, and 6G orchestration.

Daniel Massicotte received his B.Eng. degree in 1988 and M.Sc.A. degree in 1990 from the Université du Québec à Trois-Rivières (UQTR), and his Ph.D. degree in electrical engineering from École Polytechnique de Montréal in 1995. He joined the Department of Electrical and Computer Engineering, UQTR, in 1994, becoming a full professor in 2003. He served as Department Director (2014–2020) and Director of the Industrial Electronics Research Group (2011–2018), and holds the Research Chair on Signals and High-Performance Systems Intelligence. He is a Senior Member of IEEE and author or co-author of more than 200 scientific publications. His research interests include digital signal processing, VLSI design, machine learning, and biosignal processing.

Sifeddine Salmi is a researcher affiliated with the Université du Québec à Trois-Rivières (UQTR), QC, Canada. His research addresses SDN-based traffic classification using deep reinforcement learning, physical-layer network coding, and AI-native O-RAN architectures for 6G.

Felipe Augusto Pereira de Figueiredo received his B.Sc. and M.Sc. degrees in telecommunications engineering from the National Institute of Telecommunications (Inatel), Brazil, in 2004 and 2011, respectively, and his Ph.D. degrees from the State University of Campinas (UNICAMP), Brazil, in 2019, and Ghent University, Belgium, in 2021. With over 15 years of experience in telecommunications R&D, including GSM, WCDMA, LTE, and 5G systems, he is currently a researcher and assistant professor at Inatel. His research interests include digital signal processing, MIMO, multicarrier modulation, reconfigurable intelligent surfaces, and machine learning.

Adlen Ksentini received his M.Sc. degree in telecommunications and multimedia networking from the University of Versailles and his Ph.D. degree in computer science from the University of Cergy-Pontoise, France, in 2005. He was an assistant/associate professor at the University of Rennes 1 and a member of the INRIA Dionysos team (2006–2016). Since 2016, he has been a professor with the Communication Systems Department, EURECOM, France. He is an IEEE Communications Society Distinguished Lecturer and has received best paper awards at IEEE ICC, IEEE IWCMC, and ACM MSWiM, as well as the 2017 IEEE ComSoc Fred W. Ellersick Prize. His research interests include network softwarization, network slicing, edge computing, and 5G/6G architectures.