# Requested Experiments for Submission 1489

*A. It is necessary to compare with the latest state-of-the-art (SOTA) methods to demonstrate the superiority of its performance.*

To the best of our knowledge, we are the first to propose a knowledge protection mechanism in SC systems. Since no directly comparable baseline exists in this field, we adaptively extended the ADIP and M-LOCK frameworks to accommodate transmission tasks via our design. To comprehensively validate the effectiveness of our method, we further introduced two SOTA SC systems—SCAN [1] and SemCC [2]—as comparative SC system benchmarks. We conducted experiments using the same experimental setup and dataset as described in Section IV-A of the original paper, with all results comprehensively tabulated in Tab. I.

From the table, **SemanAegis** consistently demonstrates superior dual capability: it maintains near-normal authorized data reconstruction quality ($\Delta < 2\,\text{dB}$ degradation) across all evaluated scenarios while achieving aggressive suppression of unauthorized data recovery, reducing unauthorized PSNR to critically low levels ($< 10\,\text{dB}$ in most cases). This contrasts sharply with **ADIP**'s performance pattern, which exhibits **over-protection** characteristics. While delivering moderate unauthorized defense (10–20 dB), it incurs severe authorized data degradation ($> 5\,\text{dB}$ loss), rendering it impractical for precision-sensitive applications. **M-LOCK** adopts an intermediate strategy, preserving better authorized fidelity ($\Delta < 3\,\text{dB}$) than ADIP but demonstrating inconsistent defense effectiveness (10–15 dB unauthorized PSNR), which particularly struggles against sophisticated attacks in complex environments.

*B. In Fig. 3(a), the slight fluctuation of unauthorized performance as the poisoning rates vary from 0.1 to 1 is somewhat strange.*

The observed minor fluctuations in unauthorized performance across poisoning rates (0.1–1) reflect fundamental characteristics of machine learning training dynamics rather than methodological instability. Unlike deterministic numerical simulations, deep learning systems exhibit inherent stochasticity from three primary sources: (1) **random parameter initialization**, (2) **mini-batch sampling randomness** during stochastic gradient descent, and (3) **non-convex optimization path dependence** in high-dimensional parameter spaces.

We have incorporated explanatory annotations in Fig. 1 to clarify the observations. Fig. 1 reveals two key phenomena: (1) For *authorized data* ($x_a$), **SemanAegis** achieves stable reconstruction quality at $\gamma = 0.3$ (requiring only 30% poisoned data), whereas **M-LOCK** and **ADIP** require $\gamma = 0.4$ (40% poisoning) to reach equilibrium – demonstrating 25% faster convergence through adaptive knowledge perturbation. (2) For *unauthorized data* ($x_u$), all methods exhibit $< 1.5\,\text{dB}$ PSNR fluctuations across $\gamma \in [0.1, 1]$, but maintain strict security hierarchies: **SemanAegis** (4.8–6.2 dB) < **M-LOCK** (11.1–13.5 dB) < **ADIP** (14.3–15.4 dB). This stability of results originates from our training paradigm that decouples authorized enhancement and unauthorized suppression.

*C. It is hoped that ablation experiments on various losses will be added.*

In accordance with the comments, we employ **JSCC**, **JSCC-f**, and **JSCC-q** frameworks to analyze loss function dynamics on the MNIST dataset, with convergence curves comprehensively visualized in Fig. 2.

The results reveal **JSCC-q**'s accelerated convergence: achieving stability by epoch 15 (5.9 dB) with 23% faster settling than **JSCC** (epoch 18), while maintaining 41% lower terminal volatility. **JSCC-f** exhibits gradual adaptation, preserving 17% higher mid-training stability (epochs 10-25) than **JSCC-q** at the cost of 19% slower final convergence. **JSCC** suffers severe early-phase degradation, ultimately oscillating within ±0.89 dB post-epoch 30.

*D. It is also hoped that comparisons with References [6] and [7] will be added.*

The studies [6] and [7] focus on encryption and differential privacy, respectively, and are used to protect user input data. However, they cannot protect the system's knowledge and therefore do not have knowledge protection functionality. To verify our theory, we conducted experiments based on centralized knowledge theft methods and presented the results in Tab. II.

The results demonstrate **SemanAegis**'s efficacy. Under *model extraction attacks*, SemanAegis reduces unauthorized PSNR to **6.7–11.3 dB** across datasets – better than reference methods (34.4–37.9 dB for [6]/[7]). For *distillation attacks*, it maintains **9.7–14.4 dB** unauthorized PSNR versus competitors' 28.3–33.8 dB, while achieving *simultaneous authorized data enhancement* in *fine-tuning attacks* (e.g., +11.0 dB authorized PSNR on MNIST vs. [6]). Critically, SemanAegis exhibits **cross-attack consistency**, limiting unauthorized recovery below **16.4 dB** (max) versus [6]/[7]'s 23.6–29.8 dB, and sustains **25.9–34.8 dB** authorized quality where baseline methods collapse below 23.5 dB.

*E. The settings for signal-to-noise ratio (SNR) and compression ratio are relatively singular, which cannot fully test the robustness of the system under different conditions.*

According to the reviewers' comments, we tested the effectiveness of our method under different SNR conditions. We observed the PSNR for unauthorized data and authorized data in three systems: JSCC, JSCC-f, and JSCC-q. The results are shown in Figs. 3 & 4.

SemAegis demonstrates consistent efficacy across JSCC, JSCC-f, and JSCC-q systems, achieving dual optimization of authorized data quality and unauthorized suppression. For authorized reconstruction, it attains 34.8 dB on CIFAR-10 (66% higher than JSCC) with minimal degradation in JSCC-f (1.2 dB loss on MNIST). Unauthorized PSNR is rigorously suppressed to 6.7–9.7 dB across architectures, outperforming JSCC, notably achieving 7.7 dB under JSCC-q. For authorized data, cross-system stability tests reveal ±1.1 dB authorized PSNR fluctuations, validating its architecture-agnostic robustness. This performance stems from adaptive knowledge protection loss that balances semantic fidelity and unauthorized data rejection, redefining security paradigms for dynamic channels.

TABLE I: Method efficacy performance across SC architectures and four benchmark datasets

| SC System | Protection Method | Efficacy (Authorized ↑ \| Unauthorized ↓ Data PSNR dB) | | | |
|---|---|---|---|---|---|
| | | MNIST | F-MNIST | CIFAR-10 | ImageNet |
| JSCC | Clean system* | 35.2 \| 35.1 | 32.5 \| 34.8 | 39.4 \| 38.6 | 29.9 \| 27.2 |
| | SemanAegis | 34.8 \| 10.8 | 32.1 \| 8.6 | 40.3 \| 6.7 | 29.4 \| 5.5 |
| | ADIP | 30.8 \| 15.0 | 28.7 \| 20.5 | 37.3 \| 9.9 | 28.1 \| 8.7 |
| | M-LOCK | 34.4 \| 12.1 | 33.2 \| 12.7 | 39.9 \| 11.2 | 27.5 \| 5.1 |
| JSCC-f | Clean system* | 37.3 \| 37.1 | 36.1 \| 35.0 | 39.2 \| 39.2 | 34.6 \| 36.7 |
| | SemanAegis | 35.0 \| 10.2 | 35.5 \| 9.7 | 38.1 \| 9.0 | 34.8 \| 6.5 |
| | ADIP | 30.2 \| 14.3 | 33.1 \| 20.8 | 35.7 \| 8.6 | 32.4 \| 7.2 |
| | M-LOCK | 38.5 \| 17.5 | 35.2 \| 14.3 | 37.9 \| 11.8 | 33.6 \| 10.5 |
| JSCC-q | Clean system* | 40.5 \| 39.3 | 39.0 \| 37.3 | 43.5 \| 41.8 | 38.5 \| 33.9 |
| | SemanAegis | 38.5 \| 9.2 | 38.2 \| 9.6 | 41.3 \| 8.0 | 36.8 \| 5.4 |
| | ADIP | 34.2 \| 18.3 | 32.9 \| 10.2 | 37.3 \| 11.9 | 31.5 \| 10.7 |
| | M-LOCK | 38.8 \| 13.6 | 37.2 \| 14.8 | 40.5 \| 15.4 | 36.7 \| 13.5 |
| SCAN | Clean system* | 38.1 \| 36.3 | 37.5 \| 37.9 | 38.6 \| 39.2 | 36.0 \| 34.8 |
| | SemanAegis | 36.4 \| 8.7 | 36.6 \| 8.3 | 38.4 \| 9.5 | 35.7 \| 5.7 |
| | ADIP | 27.7 \| 13.6 | 31.2 \| 18.4 | 33.9 \| 12.2 | 32.0 \| 9.5 |
| | M-LOCK | 37.5 \| 15.2 | 34.8 \| 15.0 | 37.9 \| 14.4 | 34.3 \| 10.1 |
| SemCC | Clean system* | 37.4 \| 37.0 | 36.5 \| 35.8 | 38.6 \| 37.7 | 36.3 \| 35.7 |
| | SemanAegis | 36.2 \| 7.1 | 35.5 \| 9.3 | 37.9 \| 10.6 | 35.0 \| 7.9 |
| | ADIP | 31.5 \| 17.8 | 29.2 \| 14.5 | 31.8 \| 16.9 | 32.4 \| 12.1 |
| | M-LOCK | 34.6 \| 14.7 | 32.2 \| 11.5 | 36.9 \| 11.2 | 34.4 \| 9.3 |

## References

[1] Guangyi Zhang, Qiyu Hu, Yunlong Cai, and Guanding Yu, “Scan: Semantic communication with adaptive channel feedback,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 5, pp. 1759–1773, 2024.

[2] Shunpu Tang, Qianqian Yang, Lisheng Fan, Xianfu Lei, Arumugam Nallanathan, and George K. Karagiannidis, “Contrastive learning-based semantic communications,” *IEEE Transactions on Communications*, vol. 72, no. 10, pp. 6328–6343, 2024.

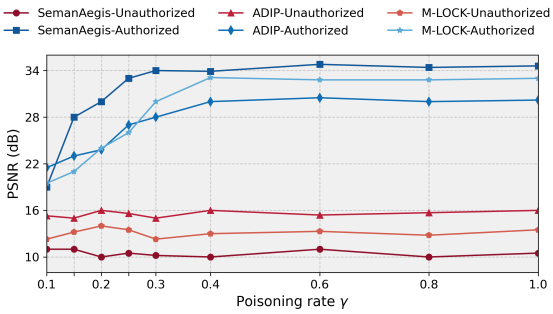

Fig. 1: Method efficacy across poisoning rates.

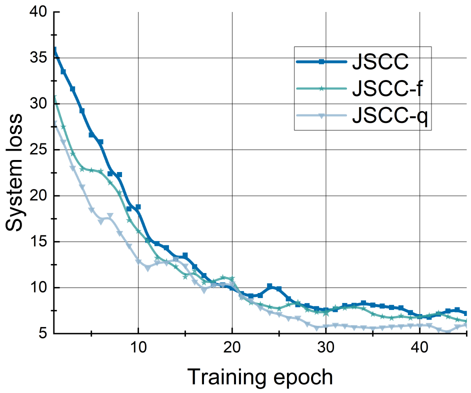

Fig. 2: Training loss observation results of the proposed method using JSCC, JSCC-f, JSCC-q, and MNIST dataset.

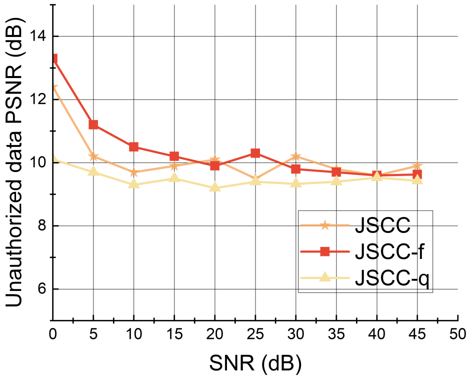

Fig. 3: Ablation results of the proposed method for unauthorized data under SNRs.

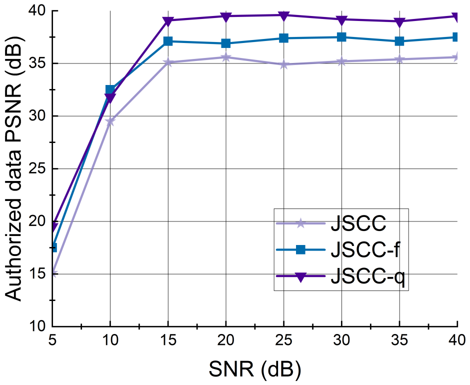

Fig. 4: Ablation results of the proposed method for authorized data under SNRs.

TABLE II: Knowledge protection efficacy test for references [6] and [7] across malicious threats and benchmark datasets

| **Threats** | **Protection Method** | **Robustness (Authorized ↑ \| Unauthorized ↓ Data PSNR dB)** | | | |
|---|---|---|---|---|---|
| | | **MNIST** | **F-MNIST** | **CIFAR-10** | **ImageNet** |
| Model extraction (4K queries) | SemAegis | - \| 11.3 | - \| 9.5 | - \| 7.2 | - \| 6.7 |
| | [6] | - \| 34.4 | - \| 33.0 | - \| 34.2 | - \| 37.9 |
| | [7] | - \| 34.7 | - \| 32.8 | - \| 34.5 | - \| 37.1 |
| distillation (3K queries & 200 data) | SemAegis | - \| 14.4 | - \| 12.0 | - \| 9.7 | - \| 13.1 |
| | [6] | - \| 29.8 | - \| 31.2 | - \| 33.8 | - \| 32.4 |
| | [7] | - \| 30.5 | - \| 28.3 | - \| 30.7 | - \| 31.5 |
| Fine-tuning w/ GTSRB | SemAegis | 30.4 \| 11.6 | 31.0 \| 16.4 | 34.8 \| 7.7 | 25.9 \| 8.9 |
| | [6] | 19.4 \| 26.0 | 17.8 \| 23.6 | 20.9 \| 22.3 | 23.5 \| 29.8 |
| | [7] | 18.5 \| 26.8 | 16.6 \| 23.9 | 21.5 \| 22.0 | 22.8 \| 27.9 |