Semantic Communication for Distributed Spectrum Monitoring over Unreliable Links

Samer Lahoud
Faculty of Computer Science
Dalhousie University
Halifax, NS, B3H 4R2 Canada
sml@dal.ca

Kinda Khawam
ROCS, LISN
Paris-Saclay University
91190 Gif-sur-Yvette, France
kinda.khawam@universite-paris-saclay.fr

***Abstract*—Distributed spectrum monitoring infers the state of the radio environment from spatially separated receivers. A central task is to determine how many emitters are active and where they are located. Sensors typically report over shared unreliable links, where contention, fading, or energy limits erase whole reports, so forwarding raw I/Q samples to a fusion node becomes impractical. We formulate distributed emitter counting and localization as a semantic communication problem over such links. Each receiver maps its local time–frequency observation to a compact latent and transmits it. The fusion node pairs each surviving latent with the corresponding receiver position and operates on the resulting unordered set, which keeps fusion compatible with variable membership and receiver-count changes. Encoder and decoder are trained end-to-end with report erasure, and optionally finite-rate quantization, inside the task objective. Experiments on synthetic multi-emitter scenes show that channel-aware training improves robustness under erasure, that set fusion outperforms fixed-order concatenation, and that one trained model operates across receiver counts. A few bits per latent component suffice.**

***Index Terms*—Semantic communication, task-oriented communication, spectrum monitoring, emitter localization, permutation-invariant fusion, erasure channel.**

I. Introduction

Semantic and task-oriented communication shift the design objective from reproducing data to delivering the information needed for a downstream task [1]–[3]. Distributed spectrum monitoring is a natural setting for this shift. A monitoring system infers how many emitters are active and where they are located. This supports dynamic spectrum access, interference management, anomaly detection, and security in beyond-5G networks. Such systems increasingly rely on low-cost distributed receivers. Forwarding raw in-phase/quadrature (I/Q) streams from all of them costs bandwidth, energy, and access-channel resources, so each receiver should transmit a compact task-relevant report instead.

The link that carries these reports is also part of the problem. Monitoring receivers often use shared, contention-based, or grant-free links, where a collision, a deep fade, or an energy limitation erases a whole report [4]. The impairment is therefore packet-level. A model trained on complete report sets can then degrade sharply when the fusion node sees a subset of receivers. The reporting set also changes from one observation window to the next.

We study emitter counting and localization as a concrete monitoring task, and we align the learning architecture with the reporting channel. Each receiver applies a shared encoder to its local time–frequency observation and sends the resulting compact latent. The fusion node pairs each survivor with its known receiver position and processes the reports as an unordered set. Our contributions are threefold. First, we cast distributed counting and localization over unreliable reporting links as a channel-aware semantic communication problem, optimizing the representation for task distortion under erasure and rate constraints. Second, we introduce a permutation-invariant fusion architecture over the surviving position-tagged reports, which tolerates variable membership and enables cross-count deployment without retraining. Third, we quantify the effects of channel-aware training, fusion architecture, receiver-count variation, and latent quantization.

Related work. Task-oriented communication replaces bit-level reconstruction by task utility [1], [2], from single-device edge inference [3] to cooperative inference through the distributed information bottleneck [5]. Deep joint source–channel coding maps a source directly to channel symbols [6]. We keep a digital reporting interface instead, and we model the erasure of whole reports. Learned RF sensing and classical cooperative sensing generally assume that features or quantized decisions reach the fusion point reliably. A closely related recent work studies task-oriented compression for multi-emitter localization with spectral overlap over a clean link and a fixed array [7]. Our fusion rule builds on permutation-invariant set models [8], [9], and here the channel dictates that choice: the fusion node receives whichever subset of reports survives.

II. System Model and Method

A. Monitoring Scenario and Task

A two-dimensional region $\mathcal{A} \subset \mathbb{R}^2$ is observed by M sensing nodes. Node i has known position $r_i \in \mathcal{A}$, which reaches the decoder as side information. Positions are redrawn for every scene, so the model handles an arbitrary receiver layout.

The region holds K simultaneously active emitters, with $1 \le K \le K_{\max}$. Emitter k has location p_k and nuisance parameters $\xi_k = (\nu_k, B_k, P_k, g_k)$: center-frequency offset,

979-8-3195-0167-7/26/$31.00 ©2026 IEEE

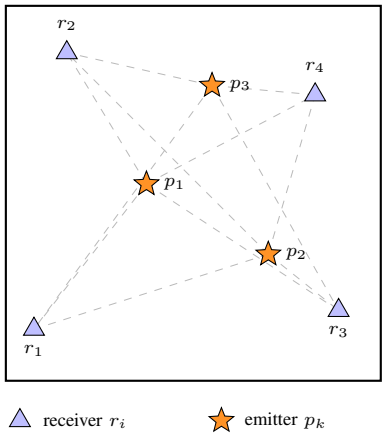


Fig. 1. Monitoring geometry: a 1000×1000 m region with emitters at unknown locations (stars) and M receivers at known, scene-dependent positions (triangles); dashed lines denote propagation.

occupied bandwidth, transmit power, and waveform family. These parameters shape the received signals and often create spectral overlap. The monitoring output is the emitter set $\Theta = \{p_k\}_{k=1}^{K}$ with its cardinality K, and ξ_k remains a nuisance variable.

Counting and localization both draw on spatial diversity, so every lost report removes one spatial view of the scene. The task therefore probes reporting-link reliability directly. Since Θ is unordered with unknown cardinality, a permutation-invariant set distortion $D_{\text{set}}(\Theta, \hat{\Theta})$ scores $\hat{\Theta}$ after the best assignment and charges both cardinality and localization errors (Section II-D). Fig. 1 shows the geometry.

B. Distributed Semantic Encoding and Rate

The fusion node recovers just enough information to count and locate the emitters, which frees the reports from full reconstruction. Receiver i observes a window of N complex baseband samples $x_i \in \mathbb{C}^N$, computes a time–frequency feature $X_i = \mathcal{T}(x_i)$, and maps it to a compact latent

$$z_i = \text{Enc}_\phi(X_i) \in \mathbb{R}^d, \qquad u_i = (z_i, r_i), \tag{1}$$

where u_i is the report token assembled at the fusion node. Here $\mathcal{T}$ is a short-time Fourier transform whose real and imaginary parts are stacked as two channels. The encoder Enc_ϕ is a residual convolutional network with global average pooling and a linear projection. Group normalization keeps the encoder independent of batch statistics, which matters because erasures are sampled during training. All receivers share the same ϕ, since positions vary across scenes and every receiver plays the same role. The position tag is essential: the same latent carries a different geometric meaning depending on where the observation was taken. Only z_i crosses the reporting link.

A finite-rate link quantizes each component of z_i to b bits by per-dimension affine min–max calibration. This gives a per-receiver rate $R_i = bd$ and a total offered reporting load $R_{\text{tot}} = Mbd$ bits per window. Receiver positions arrive as side information, so R_i counts the semantic payload alone. Experiments that isolate erasure and fusion architecture use full-precision latents, which keeps quantization out of the comparison. In Section III-E, b is sampled during training so that one model covers several reporting rates.

C. Report Erasure and Set Fusion

Reports reach the fusion node over shared links accessed in a contention-based or grant-free manner [4]. The salient impairment there is the loss of a whole report. We model this as whole-report erasure: receiver i joins the surviving set $\mathcal{S} \subseteq \{1, \dots, M\}$ with probability $1 - \epsilon$ and drops out otherwise, and the pattern is resampled each window. As an access-level reading, a slotted-ALOHA link of offered load G gives $\epsilon = 1 - e^{-G}$. The training range $\epsilon \in [0, 0.5]$ then corresponds to $G \lesssim 0.69$, below the throughput peak, while larger values serve as stress tests.

The decoder therefore receives a variable-size collection whose membership the channel determines. We model it as a set,

$$\hat{\Theta} = \text{Dec}_\psi\big(\{u_i : i \in \mathcal{S}\}\big), \tag{2}$$

with Dec_ψ permutation-invariant [8], [9]. This is the central architectural choice: the unordered set sits on the reporting side as well as the output side, so a lost report simply leaves the input set. The decoder then stays well defined for any $\mathcal{S}$ and applies at receiver counts beyond the training one. A fixed-order decoder ties its input dimension to a prescribed receiver count and index ordering, so it applies at its own training count alone.

A shared multilayer perceptron embeds each surviving report as $h_i = \rho([z_i, r_i])$, with r_i normalized. Attention pooling then aggregates the embeddings using a learned query q and keys k_i,

$$s = \sum_{i \in \mathcal{S}} \alpha_i h_i, \qquad \alpha_i = \frac{\exp(q^\top k_i / \sqrt{d_h})}{\sum_{j \in \mathcal{S}} \exp(q^\top k_j / \sqrt{d_h})}, \tag{3}$$

using the multi-head Set Transformer pooling block [9]. Sum and softmax range over $\mathcal{S}$ alone, which makes (3) permutation-invariant and confines the summary to the reports that arrive. When $\mathcal{S} \neq \emptyset$, erased latents are zeroed and excluded from the attention weights. For $\mathcal{S} = \emptyset$, the implementation assigns uniform weights to the M inputs with $z_i = \mathbf{0}$. Since the receiver positions remain available, s then depends on receiver geometry alone. The normalized survivor count $|\mathcal{S}|/M$ is appended to s, so the decoder knows how much spatial evidence it holds. A final multilayer perceptron emits $K_{\max}$ slots $\hat{y}_j = (\hat{e}_j, \hat{p}_{x,j}, \hat{p}_{y,j})$, where $\hat{e}_j$ is an existence logit. Fig. 2 summarizes the pipeline.

D. Set Distortion and Channel-Aware Training

The true emitter set is unordered with variable cardinality, so slots are matched before the loss is computed. Ground truth is padded to $K_{\max}$ slots with existence labels $e_k \in \{0, 1\}$. The cost between predicted slot j and target slot k is

$$C(j, k) = \ell_{\text{BCE}}(\hat{e}_j, e_k) + e_k \sum_{c \in \{x,y\}} \ell_{\text{reg}}(\hat{p}_{c,j}, p_{c,k}), \tag{4}$$

Here ℓ_{BCE} is the binary cross-entropy and ℓ_{reg} the smooth-L_1 loss. The gate e_k activates the localization term for real emitters, so padded targets carry an existence cost alone.

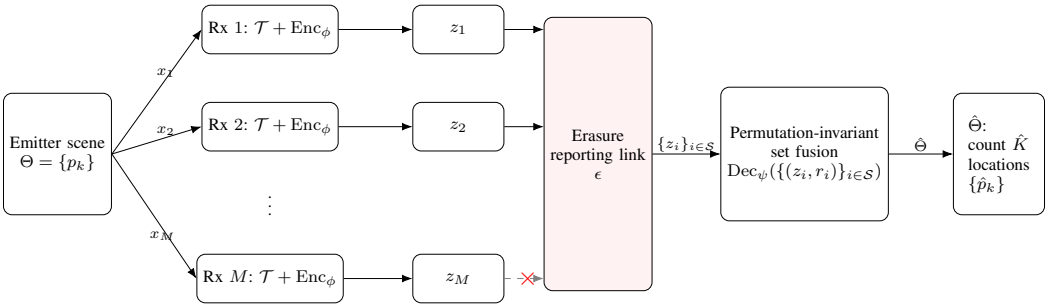


Fig. 2. Distributed semantic reporting and set fusion. Each receiver transmits a compact latent z_i over a link that may erase whole reports with probability ϵ ($\times$). At the fusion node, each surviving latent is paired with its known receiver position r_i to form $u_i = (z_i, r_i)$, and the set decoder returns an emitter count and locations.

The distortion is $D_{\text{set}} = \sum_k C(\pi^\star(k), k)$ under the optimal permutation $\pi^\star$ of the $K_{\max}$ slots, found by exhaustive search. The same matching defines the reported metrics. Encoder and decoder minimize the expected distortion over random scenes, geometries, and channel realizations,

$$\min_{\phi,\,\psi} \mathbb{E}_{\Theta,\,\mathcal{H}}\left[D_{\text{set}}(\Theta, \hat{\Theta})\right] \quad \text{s.t.} \quad R_{\text{tot}} \le R_{\text{budget}}, \tag{5}$$

where $\mathcal{H}$ is the reporting-link realization: the surviving set $\mathcal{S}$ in the erasure experiments, plus the sampled bit depth in the rate study. Placing $\mathcal{H}$ inside the expectation is what makes the system channel-aware. Erasure is applied between encoder and decoder at every minibatch with $\epsilon \sim \mathcal{U}[0, 0.5]$, and a straight-through estimator carries gradients through the quantizer. The channel-naive baseline occupies the clean-link corner of (5). The fixed-order baseline shares the encoder and also receives receiver positions, yet it consumes the M reports in a fixed order. An erased latent becomes a zero vector there, which holds the input dimension fixed.

III. Experiments and Results

A. Setup

Scenes are generated synthetically per Section II. Unless stated otherwise, $M = 4$ receivers are placed uniformly at random in a 1000×1000 m region. Each scene contains $K \in \{1, 2, 3\}$ emitters drawn uniformly, with random locations and nuisance parameters: bandwidth $B_k \in [1, 16]$ MHz, center-frequency offset ν_k uniform subject to the occupied band lying within the ± 20 MHz baseband, power $P_k \in [-10, 10]$ dBm, and waveform family either OFDM or single-carrier. Spectral overlap is therefore common. Propagation follows a power-law path loss with exponent $\eta = 2.7$, and all receivers share a -90 dBm noise power.

Each receiver observes a 1 ms window sampled at 40 MHz, giving $N = 4 \times 10^4$ samples. The short-time Fourier feature uses a Hann window of length 256 and hop 160, which yields $N_t = 249$ frames. The shared encoder produces a $d = 16$ latent per receiver, and the decoder outputs $K_{\max} = 3$ slots. The dataset holds 40,000 training, 4,000 validation, and 4,000 test scenes. All models train for 50 epochs with AdamW at learning rate 10^{-3} and batch size 64, under a cosine schedule. The best-validation checkpoint is tested, and each operating point is averaged over 30 channel realizations with fixed seeds. After set matching we report cardinality accuracy, the probability that $\hat{K} = K$, and localization RMSE over matched active emitters.

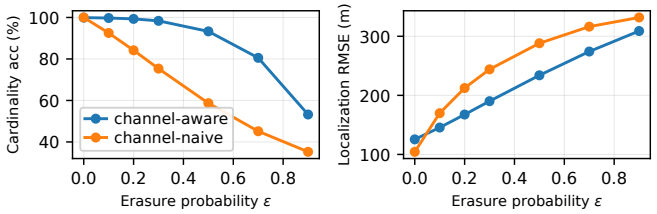


Fig. 3. Channel-aware versus channel-naive set-fusion models under report erasure: cardinality accuracy (left) and localization RMSE (right).

B. Channel-Aware Co-Design

Fig. 3 compares two models that share the encoder and set decoder and differ in the training channel alone. The channel-aware model samples $\epsilon \sim \mathcal{U}[0, 0.5]$, while the channel-naive one trains on clean links. Both perform well on clean links, and the naive model even holds a slight localization advantage there, since it is optimized for the full-report regime. As ϵ grows, its cardinality accuracy drops sharply, which reveals a decoder that relies on the joint availability of several reports. The channel-aware model instead preserves counting accuracy over a much wider range, and localization shows the same crossover. Placing the channel inside the training expectation therefore reshapes the representation, making each report useful on its own. The gain appears as robustness where it matters.

C. Set Fusion versus Fixed-Order Concatenation

We next isolate the fusion rule. Both models share the encoder, receive receiver positions, and train with channel-aware erasure under the same budget, so the representation of the report collection is the single difference. Fig. 4 shows the permutation-invariant decoder to be substantially more robust. Both count accurately on clean links, yet the set decoder already localizes better, and the cardinality gap widens sharply as erasure increases. In the set decoder, pooling runs over the reports that arrive, so an erased report simply leaves the set. The fixed-order decoder instead retains a fixed-dimensional, ordered input, so missing reports perturb a representation built for constant membership. Set fusion also localizes better across the sweep, most clearly at low to moderate erasure. At very

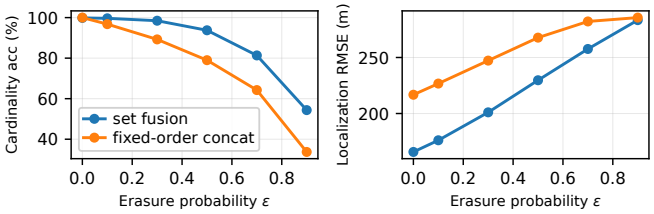


Fig. 4. Permutation-invariant set fusion versus fixed-order concatenation under report erasure. Both are trained with channel-aware erasure and both receive receiver-position information.

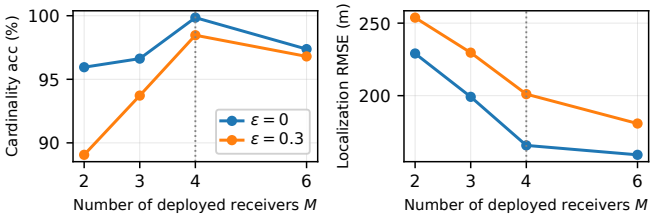


Fig. 5. Cross-count deployment of one channel-aware set-fusion model trained at $M = 4$ receivers (dotted line) and evaluated without retraining at $M \in \{2, 3, 4, 6\}$.

high erasure the curves converge, since the few surviving reports limit both models.

D. Deployment Across Receiver Counts

We then test operation beyond the training receiver count. The channel-aware model is trained with $M = 4$ receivers and evaluated without retraining at $M \in \{2, 3, 4, 6\}$. This sweep suits the permutation-invariant formulation, since a fixed-order decoder fixes its input dimension at the training count and therefore accepts $M = 4$ alone. Fig. 5 shows the same model remains effective across counts. Localization improves with more receivers, on clean links and under erasure alike, as spatial diversity better constrains the inverse problem. The gain is clearest from two to four receivers, with diminishing returns beyond the training value. Cardinality accuracy stays high and peaks near the training configuration. The drop at $M = 2$ reflects limited spatial evidence, and the mild drop at $M = 6$ shows that extrapolating beyond the training count remains possible at a small cost. Performance therefore tracks the receiver count, while one architecture and one checkpoint cover the whole range.

E. Rate–Reliability

Finally we study how the semantic rate interacts with report loss. Each latent component is quantized to b bits, so the load is $R_{\text{tot}} = Mbd = 64b$ bits per window at $M = 4$ and $d = 16$. Fig. 6 plots both metrics against the total offered reporting load R_{tot} for several erasure probabilities. A few bits per component suffice: performance improves rapidly at low rates and saturates by $b = 4$, that is, $R_{\text{tot}} \approx 256$ bits. Erasure shifts the performance level while the saturation point holds. Higher ϵ lowers accuracy and raises RMSE as fewer reports arrive, yet the knee stays in roughly the same rate range. Quantization and report loss therefore act as partly separable impairments.

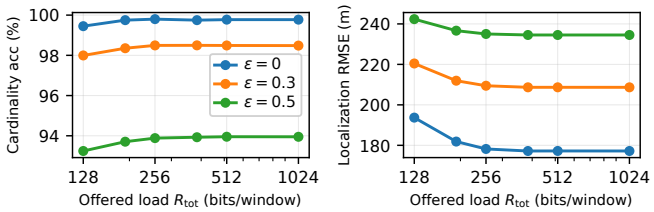


Fig. 6. Cardinality accuracy (left) and localization RMSE (right) versus total offered reporting load $R_{\text{tot}} = Mbd$ for $\epsilon \in \{0, 0.3, 0.5\}$.

Quantization limits the information carried by each surviving report, and erasure limits how much spatial evidence reaches the fusion node, so extra bit depth beyond the knee leaves the missing observations unrecovered. The load stays small compared with raw forwarding. At $b = 4$ each receiver sends 64 bits, or 256 bits in total, against 1.28 Mbit per receiver and 5.12 Mbit in total for the raw window with 16-bit I/Q samples, a reduction of about 2×10^4.

IV. Conclusion

We presented a semantic communication scheme for distributed spectrum monitoring over unreliable links. Receivers send compact latent reports, and a permutation-invariant decoder pairs each survivor with its receiver position to estimate the emitter count and locations. Training with report erasure inside the task objective improves robustness. Set fusion outperforms fixed-order concatenation and supports deployment across receiver counts, and four bits per latent component suffice, or 256 bits per window at $M = 4$. Future work will extend the study to denser scenes, more realistic propagation, correlated and geometry-dependent erasures, and estimation of the spectral nuisance parameters.

References

[1] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, 2023.

[2] E. Calvanese Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, p. 107930, 2021.

[3] J. Shao, Y. Mao, and J. Zhang, "Learning task-oriented communication for edge inference: An information bottleneck approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 197–211, 2022.

[4] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 2020.

[5] J. Shao, Y. Mao, and J. Zhang, "Task-oriented communication for multi-device cooperative edge inference," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 73–87, 2023.

[6] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, 2019.

[7] H. N. Bicer and J. N. Laneman, "Spatially distributed task-oriented compression for multi-emitter localization and characterization with spectral overlap," arXiv:2606.01446 [eess.SP], 2026.

[8] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, "Deep sets," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 3391–3401.

[9] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer: A framework for attention-based permutation-invariant neural networks," in *Proc. Int. Conf. Machine Learning (ICML)*, 2019, pp. 3744–3753.