Participation Games: Collaborative Strategies for Robust Inference in Energy Harvesting Wireless Sensor Networks

Anonymous Authors[1]

Abstract

Energy-harvesting wireless sensor networks (EH-WSNs) offer sustainable solutions for large-scale IoT deployments but face challenges due to the unreliability and intermittent availability of individual sensors. We propose a comprehensive framework that integrates a game-theoretic participation strategy with a federated learning approach tailored for EH-WSNs. Our game-theoretic model enables sensors to make optimal participation decisions based on energy levels, data quality, and collective inference impact, fostering cooperative behavior while managing individual energy constraints. The federated learning framework accommodates intermittent participation and variable data quality, ensuring robust model training despite sensor unreliability. Simulation results demonstrate that our integrated approach significantly enhances inference accuracy and energy efficiency compared to traditional participation strategies.

1. Introduction

The rapid proliferation of the Internet of Things (IoT) has sparked a tremendous growth in the scale and diversity of sensor deployments, from smart homes to expansive industrial and environmental monitoring systems. As these networks continue to expand, sustaining continuous operation in the face of finite power sources becomes a paramount concern. To address this, energy harvesting (EH) technologies have emerged as a viable solution, enabling sensors to convert ambient energy (e.g., solar, thermal, or vibration) into electrical power. This approach promises perpetual, maintenance-free operation, significantly reducing environmental impact and long-term operational costs.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Despite these advantages, large-scale EH Wireless Sensor Networks (EH-WSNs) remain inherently uncertain. Ambient energy availability varies over time and space, leading to fluctuating sensor activity levels and intermittent participation in both training and inference tasks. Some sensors may frequently become inactive or produce low-quality data due to energy scarcity or environmental noise. Consequently, the mere presence of numerous EH sensors does not guarantee robust and reliable performance for complex tasks such as image recognition, acoustic surveillance, or precision agriculture monitoring.

Achieving accurate inference in these complex scenarios depends on effective coordination. Multiple sensors observing the same phenomenon from different angles can collectively provide more comprehensive and reliable insights than any single sensor could. However, requiring all sensors to participate at all times is impractical, as it drains energy reserves too quickly. Conversely, simplistic policies—such as selecting only the highest-energy sensors—ignore factors like data relevance, sensor quality, and the strategic implications of current participation on future network states.

This challenge motivates the need for intelligent, context-aware participation strategies that dynamically determine which sensors should engage during both the *training* phase—where global model parameters are periodically fine-tuned or updated—and the *inference* phase—where newly observed data are aggregated to produce predictions. Sensors must carefully balance immediate accuracy gains against conserving energy for future tasks, while also anticipating the behavior of other sensors that may be collaborating or competing.

To address these interdependent decisions, we employ a game-theoretic framework. Unlike simple heuristic methods that ignore future resource allocation or complex approaches like reinforcement learning that may be too costly to implement, game theory provides equilibrium guarantees. By modeling each sensor as a rational player aiming to optimize its own long-term utility, we achieve stable, cooperative equilibria where no sensor can improve its outcome through unilateral deviation. This strategic equilibrium underpins both training and inference participation decisions, ensuring that the sensors most likely to improve the global model—given their energy, data quality, and network conditions—are the ones that engage.

To refine the global model parameters without incurring continuous on-edge training costs, we adopt a federated learning paradigm adapted to EH-WSNs. Rather than relying on persistent, centralized updates or continuous federated aggregation, we perform *periodic* or *equilibrium-driven* fine-tuning rounds. These updates occur only when sensors have sufficient energy to participate meaningfully, guided by the game-theoretic equilibrium strategy. By integrating the game-theoretic approach with federated learning principles, we reduce communication overhead and ensure that contributions to model updates come from sensors best positioned to improve accuracy under energy constraints and uncertain availability. Our key contributions are as follows:

• **Game-Theoretic Participation Strategy:** We develop a novel game-theoretic model for EH-WSNs that applies to both training and inference phases. This model balances anticipated energy availability, local data quality, and global benefit to establish stable and cooperative equilibria, optimizing the energy-accuracy trade-offs.

• **Federated Learning Integration:** We introduce a federated learning-based framework tailored for intermittent participation and heterogeneous data quality. Unlike continuous on-edge training, we employ periodic or triggered fine-tuning sessions aligned with equilibrium strategies, ensuring robust and progressively improving global models.

• **Joint Optimization of Training and Inference:** Our unified solution aligns training participation decisions with inference needs. Sensors strategically decide when to expend energy on local model updates and when to engage in inference tasks, ultimately maximizing their long-term contribution to the network's performance.

• **Demonstrated Performance Gains:** Through simulations (add simulation details later), we show that our integrated framework outperforms baseline approaches—such as always-on participation or simplistic energy-based selection—by achieving higher inference accuracy, lower energy consumption, and more sustainable long-term operation in EH-WSNs.

By tackling the dual challenges of sensor unreliability and energy scarcity through a rigorous game-theoretic and federated learning lens, our work addresses a critical gap in the design of sustainable, intelligent EH-WSNs. This integrated framework is theoretically grounded, yet practical for a wide range of IoT applications, from remote wildlife monitoring to large-scale industrial status tracking and precision agriculture.

The remainder of this paper is organized as follows. In Section 3, we present the system model, detailing the EH-WSN setup and data capture process. In Section 4, we introduce the game-theoretic model of sensor participation, motivating our approach against simpler heuristics and discussing why equilibrium solutions are desirable. Section 5 outlines the training and fine-tuning framework that integrates the equilibrium strategies into a federated learning paradigm. Finally, Section **??** presents simulation results and Section **??** concludes with a discussion of limitations and future work.

2. Background and Related Work

Very basic, just the outline, compact and make robust

2.1. Energy Harvesting Wireless Sensor Networks

Energy harvesting wireless sensor networks (EH-WSNs) have emerged as a sustainable solution for long-term environmental monitoring, infrastructure surveillance, and IoT applications (**?**). By harnessing ambient energy sources such as solar, thermal, or kinetic energy, EH sensors can operate indefinitely without the need for battery replacement or external power supplies. However, the intermittent and unpredictable nature of harvested energy introduces significant challenges in maintaining reliable and consistent network performance (**?**).

The unreliability of individual EH sensors, due to fluctuations in energy availability, necessitates the deployment of a large number of inexpensive and potentially unreliable devices to ensure network robustness. This redundancy allows for continuous operation despite individual sensor failures or downtime. However, it also introduces complexities in coordinating sensor activities, managing energy resources, and ensuring efficient data collection and processing (**?**).

2.2. Participation Strategies in EH-WSNs

Efficient participation strategies are critical in EH-WSNs to optimize network performance while conserving limited energy resources. Traditional approaches often assume continuous participation of all sensors, which is impractical in energy-constrained environments (**?**). Some methods propose selecting a subset of sensors based on energy levels or predefined schedules (**?**), but these can lead to suboptimal performance by not considering the sensors' data quality or potential future contributions.

Several works have explored adaptive participation strategies that consider energy harvesting rates, energy consumption patterns, and application-specific requirements (**?**). These strategies aim to balance energy expenditure with the need for timely and accurate data, often using heuristic or optimization-based approaches. However, they may not fully exploit the potential for collaboration among sensors or account for the strategic interactions inherent in decentralized networks.

2.3. Game-Theoretic Models in Sensor Networks

Game theory provides a powerful framework for modeling and analyzing strategic interactions in distributed systems, including sensor networks (**?**). In the context of EH-WSNs, game-theoretic models have been employed to design distributed algorithms for resource allocation, power control, and cooperative communication (**?**).

Cooperative game theory has been used to encourage collaboration among sensors to enhance network performance (**?**). Non-cooperative game models allow sensors to make autonomous decisions while considering the potential actions of others, leading to equilibria that balance individual utility with collective goals (**?**). However, integrating game-theoretic participation strategies with machine learning tasks, particularly in energy-harvesting environments, remains an area with limited exploration.

2.4. Federated Learning in Resource-Constrained Environments

Federated learning enables multiple devices to collaboratively train a global model without sharing raw data, preserving privacy and reducing communication overhead (**?**). While federated learning has gained significant attention in mobile and IoT devices, applying it to EH-WSNs presents unique challenges due to intermittent participation, limited computational capabilities, and variable data quality (**?**).

Recent studies have begun to address federated learning in resource-constrained and unreliable networks. Strategies include adaptive aggregation methods, energy-aware training schedules, and robustness to device dropouts (**?**). However, these approaches often assume some level of reliability or do not fully integrate energy harvesting dynamics into the learning process.

2.5. Multi-View Learning and Collaborative Inference

Multi-view learning leverages multiple sources or perspectives to improve learning performance (**?**). In EH-WSNs, sensors providing different views of the same scene can enhance inference accuracy through collaborative processing. Techniques such as co-training, consensus learning, and ensemble methods have been explored to combine information from multiple sensors (**?**).

Collaborative inference in sensor networks involves combining local inferences to achieve a global understanding of the environment (**?**). Challenges include aligning heterogeneous data, managing communication costs, and dealing with unreliable or missing inputs. Existing methods may not account for the energy constraints and participation variability inherent in EH-WSNs.

3. System Model

We consider a network of N EH sensors $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ deployed to monitor a common scene. Each sensor observes the environment from a distinct vantage point. Time is slotted and indexed by $t \in \mathbb{N}$. In each time slot, the network may perform an inference event, during which sensors have the opportunity to contribute data that enhances the accuracy of a global inference task, such as object detection or environmental classification.

Each sensor s_i harvests energy from ambient sources, such as solar or vibrational energy, resulting in a stochastically varying energy supply. We denote by $E_i(t)$ the energy harvested by sensor s_i during slot t. The sensor maintains an energy buffer whose state evolves as

$$B_i(t+1) = B_i(t) + E_i(t) - e_i(t),$$

where $B_i(t)$ is the energy available at the beginning of slot t, and $e_i(t)$ is the energy expended during that slot. Predicting future energy intake is challenging, so each sensor employs an estimator $\hat{E}_i(t+1)$ to anticipate its upcoming energy resources. Incorporating uncertainty-aware models or robust estimation techniques is beyond the scope of this paper.

Prior to deployment, a global inference model f_θ is trained offline on representative data and distributed to each sensor. This model maps sensor observations to inference outputs. Although parameters θ can theoretically be updated through on-edge training, we assume that frequent retraining in situ is prohibitively expensive given energy constraints. Thus, θ remains largely static post-deployment. Sensors focus on inference using their local copies of f_θ. However, the subsequent training framework, discussed in Section 5, allows for occasional fine-tuning of θ based on equilibrium-driven participation, thereby refining the model to better suit the operational dynamics of the network.

In each inference event, sensors decide whether to participate. If sensor s_i participates at time t, it must capture data at a chosen Signal-to-Noise Ratio (SNR), process the data using f_θ, and transmit the result to a designated *lead sensor*. High-SNR data capture improves the sensor's contribution to global accuracy but consumes more energy. Let $e_{\text{cap}}(\text{SNR})$ denote the energy required for capture at a given SNR level. We assume a monotonic relationship: higher SNR increases both the capture cost and the expected accuracy contribution. This assumption simplifies the model by ensuring that better data quality unequivocally enhances inference performance, while also making the energy expenditure predictable. In addition to capture costs, participation incurs inference computation cost e_{inf} and communication cost e_{comm}. Thus, if sensor s_i participates with SNR $\text{SNR}_i(t)$, its total energy expenditure is

$$e_i(t) = e_{\text{cap}}(\text{SNR}_i(t)) + e_{\text{inf}} + e_{\text{comm}}$$

The improvement in global inference accuracy due to sensor s_i is denoted by $\Delta A_i(t)$. This quantity depends on $\text{SNR}_i(t)$ and on the data contributed by other participating sensors, as their combined perspectives shape the overall result. While $\Delta A_i(t)$ may not be known precisely, we assume that each sensor can estimate its expected contribution based on historical observations and current conditions. Every inference event presents a binary decision for sensor s_i: $a_i(t) \in \{\text{Participate (P)}, \text{Not Participate (NP)}\}$. Choosing P involves selecting an SNR level, incurring energy costs, and aiming to improve global accuracy. Choosing NP conserves energy but forfeits any contribution or associated reward. Because sensors have limited energy and the network may operate for extended periods, each sensor must consider the future implications of its current actions. The interplay of multiple sensors making similar decisions under uncertainty and energy constraints naturally suggests a game-theoretic framework for modeling their interactions.

4. Game-Theoretic Modeling

4.1. Motivation for Game Theory over Simpler Methods

While heuristic methods—such as always selecting the top-k sensors based on current energy levels—or greedy algorithms that maximize immediate utility might offer straightforward solutions, they fall short in addressing the strategic and long-term dynamics inherent in EH-WSNs. These simplistic approaches ignore the interdependencies among sensor decisions and fail to account for future resource allocation, potentially leading to suboptimal performance over time. For instance, always selecting the highest-energy sensors can rapidly deplete their energy reserves, reducing the network's resilience during critical future events.

Alternatively, reinforcement learning or Markov Decision Process-based approaches could adaptively learn participation policies that consider both immediate rewards and future states. However, these methods often require extensive training data, significant computational resources, and complex communication protocols, which may be impractical for resource-constrained sensor nodes.

In contrast, a game-theoretic framework provides equilibrium guarantees, ensuring stable and cooperative participation strategies. By modeling each sensor as a rational player optimizing its own utility, we can derive participation patterns that are robust against unilateral deviations. This stability is crucial for maintaining long-term network performance without necessitating continuous recalibration or extensive communication overhead.

4.2. Utility Function Definition

We define a utility function $U_i(t)$ for each sensor s_i that encapsulates the trade-off between accuracy gains and energy expenditures, as well as future opportunities. The utility function is designed to reflect both immediate rewards and long-term sustainability.

Immediate Rewards and Penalties: Let $\gamma > 0$ be a scaling factor that translates accuracy gains into utility rewards. When sensor s_i participates ($a_i(t) = \text{P}$) and contributes correctly to the inference task, it receives a reward proportional to the improvement in global accuracy, denoted by $\Delta A_i(t)$:

$$R_i(t) = \begin{cases} \gamma \cdot \Delta A_i(t), & \text{if } a_i(t) = \text{P and correct inference,} \\ -\delta, & \text{if } a_i(t) = \text{P and incorrect inference,} \\ -\eta, & \text{if } a_i(t) = \text{NP.} \end{cases}$$

Here, $\delta > 0$ penalizes incorrect participation, discouraging sensors from submitting low-quality data, while $\eta > 0$ penalizes non-participation to prevent perpetual abstention. Importantly, we set $\eta > \delta$, ensuring that consistently opting out is more detrimental than occasionally providing inaccurate data.

Energy Costs and Future Utility: Participation incurs energy costs, reducing the sensor's capacity for future tasks. Additionally, sensors must consider the discounted value of future utility. Let $C_i(t)$ represent the cost component:

$$C_i(t) = e_i(t) + \beta V_i(t+1),$$

where $e_i(t)$ is the total energy expenditure for participation, encompassing data capture, inference computation, and communication:

$$e_i(t) = e_{\text{cap}}(\text{SNR}_i(t)) + e_{\text{inf}} + e_{\text{comm}}.$$

The discount factor $\beta \in [0, 1)$ captures how sensors value future utility, with $V_i(t+1)$ representing the expected future utility given current decisions and predicted energy availability $\hat{E}_i(t+1)$.

Overall Utility Function: Combining immediate rewards and costs, the overall utility function for sensor s_i at time t is:

$$U_i(t) = R_i(t) - C_i(t).$$

This utility function effectively balances the benefits of participation against the associated costs and future opportunities, guiding sensors to make strategic decisions that optimize their long-term contributions to the network.

Nash Equilibrium and Stability: A Nash equilibrium (NE) represents a stable action profile $\mathbf{a}^*(t)$ where no sensor can unilaterally improve its utility by deviating from its current strategy:

$$U_i(a_i^*(t), \mathbf{a}_{-i}^*(t)) \geq U_i(a_i(t), \mathbf{a}_{-i}^*(t)) \quad \forall a_i(t), \forall i.$$

Achieving an NE ensures that sensor participation patterns are stable; once equilibrium is reached, no single sensor benefits from changing its participation decision independently. This stability is critical for maintaining consistent network performance and energy sustainability over time.

Distributed Best-Response Algorithm: To realize the NE, we propose a distributed best-response algorithm where each sensor iteratively adjusts its action based on the current state and the expected actions of others. The algorithm operates as follows:

Algorithm 1 Distributed Best-Response Participation Algorithm

---

1: **Input:** Current energies $B_i(t)$, predicted harvest $\hat{E}_i(t+1)$, parameters $\gamma, \delta, \eta, \beta$, and energy costs $e_{\text{cap}}(\cdot), e_{\text{inf}}, e_{\text{comm}}$.

2: **At each inference event:**

3: Each sensor s_i receives a solicitation from the lead sensor and forms an estimate of $\Delta A_i(t)$ given potential SNR choices and expected actions of others.

4: For each action candidate $a_i(t) \in \{\text{P}, \text{NP}\}$, the sensor computes the expected utility:

$$U_i^{a_i(t)} = \mathbb{E}[R_i(t)] - \mathbb{E}[e_i(t)] - \beta\mathbb{E}[V_i(t+1)],$$

where the expectations are taken over uncertainties in correctness, SNR impact, and future energy.

5: If $U_i^{\text{P}} \geq U_i^{\text{NP}}$ and $B_i(t) \geq e_{\text{cap}}(\text{SNR}_i(t)) + e_{\text{inf}} + e_{\text{comm}}$, then s_i chooses P. Otherwise, it chooses NP.

6: After all sensors decide, the action profile $\mathbf{a}(t)$ is realized, and energies are updated:

$$B_i(t+1) = B_i(t) + E_i(t) - e_i(t).$$

7: Sensors iterate this process at each inference event, refining their estimates and converging to stable action patterns.

Sensors employ this best-response mechanism, continuously updating their participation decisions based on the evolving network state and the actions of other sensors. Over repeated iterations, under suitable conditions, this process converges to a Nash equilibrium where participation strategies are mutually optimal.

Existence and Convergence of Equilibrium:

Theorem 4.1. *Suppose that each utility function $U_i(t)$ is non-decreasing in $\Delta A_i(t)$, that energy constraints and discounting ensure diminishing marginal returns for repeated deviations, and that sensors have consistent estimation of $\Delta A_i(t)$ and $\hat{E}_i(t+1)$. Then, the iterative best-response updates described in Algorithm 1 converge to a Nash equilibrium action profile $\mathbf{a}^*(t)$.*

Proof of Theorem 4.1: The proof constructs a potential function $\Phi(\mathbf{a}(t)) = \sum_{i=1}^{N} U_i(a_i(t), \mathbf{a}_{-i}(t))$ that strictly increases whenever a sensor makes a profitable unilateral deviation. Since utilities are bounded (due to finite energy and limited accuracy gains) and returns diminish over time, no infinite sequence of profitable deviations is possible. Hence, the best-response dynamics must terminate at a profile where no sensor can improve its utility alone, i.e., a Nash equilibrium. A complete formal proof, including all technical conditions, is provided in Appendix B.

Guidelines for Hyperparameter Selection: The parameters γ, δ, and η critically influence sensor behavior by dictating the trade-offs between participation rewards, penalties for incorrect submissions, and deterrents against non-participation. Detailed guidelines for selecting these hyperparameters are provided in Appendix A. Briefly, these parameters should be chosen to ensure that: **(1)** γ sufficiently incentivizes correct participation without leading to excessive energy expenditure. **(2)** δ appropriately penalizes incorrect inferences, discouraging low-quality data contributions. **(3)** $\eta > \delta$ to prevent sensors from consistently abstaining, thereby promoting overall network engagement. These guidelines help in balancing immediate utility gains with long-term energy sustainability, ensuring that the game-theoretic model drives desirable participation behaviors.

5. Training and Aggregation Framework

Having established the equilibrium participation strategies and the underlying reward-based utility functions, we now consider the training process that fine-tunes the global inference model $\theta \in \mathbb{R}^d$ within this EH, multi-sensor environment. Initially, θ is pre-trained offline and deployed to all sensors, enabling them to perform basic inference tasks. However, this initial model may not be optimally adapted to the complex operational reality of the network, where sensors strategically choose SNR levels, participate intermittently according to equilibrium strategies, and generate data distributions that deviate from the original training set.

The goal of the training process is to adjust θ to these conditions, effectively *fine-tuning* the model to the nonstationary data distribution $\mathcal{D}$ induced by the sensors' equilibrium behaviors. At equilibrium, sensors strike a balance between accurate data contribution and energy conservation, resulting in a stable pattern of participation and SNR choices. Over time, this induces a stationary, albeit non-trivial, effective data distribution $\mathcal{D}$.

Learning Approach: Our approach diverges from classical Learning paradigms. We adopt a hybrid strategy where periodic or event-triggered updates refine the model parameters based on equilibrium-driven data collection. This hybrid approach mitigates the high communication overhead and energy consumption, making it more suitable for resource-constrained EH-WSNs. While traditional methods require persistent communication between sensors and the aggregator, our framework leverages the established equilibrium strategies to determine optimal times for model updates. By aligning update events with periods when sensors are most likely to participate meaningfully, we ensure that the global model is refined efficiently without imposing excessive energy demands on the sensors.

Training Objective and Regularization: To enhance robustness and efficiency, we incorporate regularizers that penalize undesirable model properties. Specifically, we introduce two regularizers: **(1)** $\Omega_{\text{SNR}}(\theta)$: Encourages the model to maintain performance across varying SNR levels, preventing over-reliance on high-SNR data. **(2)** $\Omega_{\text{complexity}}(\theta)$: Controls model complexity, reducing computational and communication overheads by discouraging overly intricate models. The full training objective is formulated as:

$$J(\theta) = L(\theta) + \lambda_1 \Omega_{\text{SNR}}(\theta) + \lambda_2 \Omega_{\text{complexity}}(\theta),$$

where

$$L(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_\theta(x), y)],$$

and $\lambda_1, \lambda_2 \geq 0$ are hyperparameters that balance accuracy, robustness, and efficiency.

Gradient Computation and Backpropagation: Integrating regularizers into the training process is straightforward due to their known closed-form gradients. During backpropagation, each sensor computes the gradient of the loss function $\nabla\ell(f_\theta(x), y)$ with respect to θ based on locally available samples from $\mathcal{D}$. Additionally, the gradients of the regularizers, $\nabla\Omega_{\text{SNR}}(\theta)$ and $\nabla\Omega_{\text{complexity}}(\theta)$, are analytically derived and added to the local gradient estimates. Since both regularizers are convex and smooth, their inclusion ensures that the overall objective $J(\theta)$ maintains desirable convexity and smoothness properties, facilitating the convergence of stochastic gradient descent (SGD).

Periodic Equilibrium-Aware Training: Model updates are performed periodically at an aggregator node that collects gradient estimates from participating sensors. Participation during training follows the same equilibrium model: sensors decide whether to compute and send gradients based on their current energy states, predicted future utilities, and the established reward structure. By aggregating these gradient updates over multiple training rounds, the aggregator approximates the gradient $\nabla J(\theta)$ and performs an SGD step. Our training framework is encapsulated in Algorithm 2, which outlines the periodic equilibrium-aware training process. This training framework is intrinsically linked to the game-theoretic participation strategies. Sensors participate in training rounds based on their equilibrium-driven decisions, ensuring that gradient updates are contributed by those sensors most capable and willing to improve the global model. This alignment minimizes unnecessary energy expenditure and maximizes the efficacy of each training round.

Algorithm 2 Periodic Equilibrium-Aware Training Algorithm

1: **Initialization:** Initialize θ_0. Broadcast θ_0 to all sensors. Set a diminishing step-size schedule $\{\alpha_k\}_{k\geq 0}$.
2: **for** each training round $k = 0, 1, 2, \ldots$ **do**
3: The aggregator signals that a training update round is imminent.
4: Sensors decide on participation. Participation involves:
   1. Determining if they have enough energy and incentive (based on the established equilibrium strategy and reward parameters γ, δ, η).
   2. If participating: capturing data at their chosen SNR, performing inference, and computing local gradients $\nabla\ell(f_{\theta_k}(x), y)$ on their locally available samples drawn from $\mathcal{D}$.
   3. Adding regularizer gradients $\lambda_1\nabla\Omega_{\text{SNR}}(\theta_k)$ and $\lambda_2\nabla\Omega_{\text{complexity}}(\theta_k)$.
5: A subset of sensors, determined by the equilibrium, send their gradient estimates to the aggregator.
6: The aggregator forms an unbiased estimate of the full gradient:
$$\widehat{\nabla} J(\theta_k) = \widehat{\nabla} L(\theta_k) + \lambda_1 \nabla\Omega_{\text{SNR}}(\theta_k)$$
$$+\lambda_2 \nabla\Omega_{\text{complexity}}(\theta_k).$$
7: Update model parameters:
$$\theta_{k+1} = \theta_k - \alpha_k \widehat{\nabla} J(\theta_k).$$
8: Broadcast θ_{k+1} to all sensors.
9: **end for**

Regularizers and SGD Convergence: The chosen regularizers $\Omega_{\text{SNR}}(\theta)$ and $\Omega_{\text{complexity}}(\theta)$ are both convex and smooth, with known closed-form gradients. This property guarantees that the inclusion of regularizers does not compromise the convexity or smoothness of the overall objective $J(\theta)$. Consequently, the stochastic gradient descent (SGD) updates retain their convergence properties, ensuring that the training process reliably optimizes $J(\theta)$.

6. Implementation and Evaluation

6.1. Discussions and Limitations

7. Conclusion

This should finish at 8 pages.

Impact Statement

Authors are **required** to include a statement of the potential broader impact of their work, including its ethical aspects and future societal consequences. This statement should be in an unnumbered section at the end of the paper (co-located with Acknowledgments – the two may appear in either order, but both must be before References), and does not count toward the paper page limit. In many cases, where the ethical impacts and expected societal implications are those that are well established when advancing the field of Machine Learning, substantial discussion is not required, and a simple statement such as the following will suffice:

"This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here."

The above statement can be used verbatim in such cases, but we encourage authors to think about whether there is content which does warrant further discussion, as this statement will be apparent if the paper is later flagged for ethics review.

References

A. Guidelines for Hyperparameter Selection and Bounds on Reward Parameters

The parameters γ, δ, and η govern the reward structure of the proposed framework, influencing whether sensors participate consistently, over-participate and waste energy, or abstain altogether. This appendix provides a systematic approach to selecting these parameters, including formal bounds, practical heuristics, and an algorithmic procedure to explore suitable values.

Conceptual Role of Parameters

The scalar $\gamma > 0$ represents the reward scaling for correct participation. If γ is too low, sensors will not have sufficient incentive to expend energy on high-SNR captures. If γ is too high, sensors may waste energy attempting difficult inferences. The parameter $\delta > 0$ penalizes incorrect inferences, discouraging reckless submissions of low-quality data. The parameter $\eta > 0$ penalizes non-participation, ensuring that sensors do not remain idle indefinitely. As discussed, maintaining $\eta > \delta$ encourages sensors to at least attempt participation rather than always remain offline.

Formal Bounds and Conditions

To ensure balanced behavior, it is helpful to relate γ, δ, and η to typical values of accuracy improvement and energy costs.

Accuracy Gains and Costs: Let $\Delta A_{\min}$ and $\Delta A_{\max}$ denote the minimum and maximum expected accuracy improvements from any sensor's participation. Let $e_{\text{total}}^{\max} = e_{\text{cap}}^{\max} + e_{\text{inf}} + e_{\text{comm}}$ represent the maximum energy cost (for a chosen SNR mode). A baseline condition that ensures correct participation can overcome occasional penalties is:

$$\gamma \cdot \Delta A_{\min} > \delta + e_{\text{total}}^{\max}.$$

This inequality implies that even in a worst-case scenario for accuracy gain, the net expected benefit of correct participation surpasses the sum of potential incorrect penalties and energy costs. Without this condition, sensors might find participation systematically unprofitable.

Non-Participation and Equilibrium: Since $\eta > \delta$, we ensure that sensors prefer risking occasional incorrect inferences over consistently abstaining. A suitable gap might be chosen so that:

$$\delta < \eta \leq \delta + c,$$

for some small $c > 0$. Choosing c relative to typical gains, say $c \approx 0.1 \cdot \gamma \cdot \Delta A_{\max}$, helps maintain a moderate deterrent against non-participation without forcing sensors to always participate.

Energy Preservation: If γ is too large, sensors might not value future energy at all. To prevent myopic strategies, one can limit γ such that continuously investing in high-SNR captures does not dominate long-term considerations. For example:

$$\gamma \cdot \Delta A_{\max} < \eta + \text{margin},$$

where margin accounts for future opportunities and energy savings. A small margin ensures sensors do not always expend maximal energy for short-term gains.

Practical Hyperparameter Tuning Strategies

1. *Baseline Ratios:* Start with ratios that link γ to typical accuracy gains and set δ, η based on fractions or multiples of $\gamma \cdot \Delta A_{\min}$ or $\gamma \cdot \Delta A_{\max}$.

2. *Iterative Refinement:* Use simulation or small-scale experimental runs to refine parameters. If sensors rarely participate, increase γ or decrease η. If sensors over-exert themselves, reduce γ or increase δ, η.

3. *Adaptive Tuning:* If conditions change over time, adjust γ, δ, and η dynamically based on observed participation rates, accuracy levels, and energy depletion patterns.

Exploration Algorithm

Algorithm 3 outlines a systematic approach to exploring suitable hyperparameter values. It combines theoretical bounds with empirical evaluation, guiding the search toward stable and efficient equilibria. The above guidelines and the exploration algorithm provide a structured approach to selecting and refining γ, δ, and η. By starting from theoretically informed baseline conditions and iteratively refining through simulation-based feedback, it is possible to reach a stable set of parameters that promotes balanced participation, discourages perpetual abstention, and prevents excessive energy expenditure. Regular re-tuning may be warranted as operating conditions, energy harvesting patterns, or accuracy requirements evolve over the network's lifetime.

Algorithm 3 Hyperparameter Exploration for Reward Parameters

```
Inputs: Estimates ΔA_min, ΔA_max, energy costs e_cap^max, e_inf, e_comm, initial guesses γ_0, δ_0, η_0, and tuning increments Δ_γ, Δ_δ, Δ_η.
Compute e_total^max = e_cap^max + e_inf + e_comm.
Ensure baseline feasibility: If γ_0 · ΔA_min ≤ δ_0 + e_total^max, increase γ_0 until this condition is met.
Set η_0 > δ_0. Start with η_0 = δ_0 + c, where c is a small positive number. If preliminary tests show insufficient participation, slightly increase η_0. If participation is overly aggressive, reduce γ_0 or increase δ_0.
Simulation-Refinement Loop:
for k = 1, 2, ..., K (number of refinement iterations) do
    Run a simulation or small-scale test deployment using the current γ_k, δ_k, η_k.
    Measure key indicators: participation rate, average energy depletion rate, frequency of incorrect inferences, and overall inference accuracy.
    if participation is too low (e.g., < p_min) or sensors remain idle too often then
        Increase γ_k ← γ_k + Δ_γ or decrease η_k ← η_k − Δ_η.
    else if participation is too high, leading to frequent energy depletion then
        Decrease γ_k ← γ_k − Δ_γ or increase δ_k ← δ_k + Δ_δ to discourage high-risk attempts.
    else if incorrect inferences are prevalent then
        Increase δ_k ← δ_k + Δ_δ to penalize low-quality submissions more strongly.
    end if
    Check feasibility conditions again to ensure no violation of baseline inequalities.
    If performance metrics (accuracy, sustainability) are satisfactory, terminate. Otherwise, continue refinement.
end for
```

B. Equilibrium Existence and Convergence with Reward-Based Utility

In this appendix, we provide a detailed and formal proof that the best-response dynamics, incorporating the newly defined reward-based utility functions, converge to a Nash equilibrium (NE). We first restate the key assumptions and the utility model. We then show that the iterative best-response updates cannot lead to infinite improvement cycles, implying the existence of an NE. Finally, we prove that the equilibrium is reached under the given assumptions.

Restatement of the Utility Function and Assumptions

Recall that at each inference event t, each sensor s_i chooses an action $a_i(t) \in \{\text{P}, \text{NP}\}$. The chosen action profile is $\mathbf{a}(t) = (a_1(t), \dots, a_N(t))$.

The immediate reward for sensor s_i is defined as:

$$R_i(t) = \begin{cases} \gamma \cdot \Delta A_i(t), & \text{if } a_i(t) = \text{P and inference is correct}, \\ -\delta, & \text{if } a_i(t) = \text{P and inference is incorrect}, \\ -\eta, & \text{if } a_i(t) = \text{NP}. \end{cases}$$

Here, $\gamma > 0$ scales the reward for correct participation, $\delta > 0$ penalizes incorrect inference, and $\eta > 0$ penalizes non-participation, with $\eta > \delta$ ensuring that remaining idle is more penalizing than at least attempting participation.

The cost incorporates energy consumption and future opportunities. Let $e_i(t)$ be the energy expenditure for sensor s_i if it participates at time t, accounting for capture, inference, and communication costs. Introduce a discount factor $\beta \in [0, 1)$, and let $V_i(t+1)$ represent the expected future utility of sensor s_i given its current decisions and predicted energy availability. The cost is:

$$C_i(t) = e_i(t) + \beta V_i(t+1).$$

The overall utility is:

$$U_i(t) = R_i(t) - C_i(t).$$

We assume that $\Delta A_i(t)$ is non-decreasing in the quality of sensor s_i's data (e.g., higher SNR yields higher $\Delta A_i(t)$). We also assume that energy resources, accuracy gains, and reward/penalty parameters are finite and bounded, and that sensors have consistent estimation mechanisms for $\Delta A_i(t)$ and $\hat{E}_i(t+1)$.

Potential Function Construction

To prove convergence, we define a potential function that reflects the collective utility of the sensor network:

$$\Phi(\mathbf{a}(t)) = \sum_{i=1}^{N} U_i(a_i(t), \mathbf{a}_{-i}(t)).$$

Since $U_i(t) = R_i(t) - C_i(t)$, we have:

$$\Phi(\mathbf{a}(t)) = \sum_{i=1}^{N} [R_i(t) - C_i(t)].$$

The terms $R_i(t)$ depend on the chosen actions and correctness of inferences. Due to bounded γ, δ, and η, and the fact that $\Delta A_i(t)$ and energy costs are bounded, each $U_i(t)$ is finite. Thus, $\Phi(\mathbf{a}(t))$ is also finite for all feasible action profiles.

Monotonicity of the Potential Function

Consider a unilateral deviation by a single sensor s_j from an action $a_j(t)$ to a different action $a'_j(t)$. Such a deviation affects only $U_j(t)$, not the utilities of other sensors directly in a one-step change. If this deviation is profitable for sensor s_j, we have:

$$U_j(a'_j(t), \mathbf{a}_{-j}(t)) > U_j(a_j(t), \mathbf{a}_{-j}(t)).$$

Because the other sensors' utilities do not change instantaneously by s_j's unilateral action, the increment in $U_j(t)$ results in:

$$\Phi(\mathbf{a}_{-j}(t), a'_j(t)) - \Phi(\mathbf{a}(t)) =$$

$$U_j(a'_j(t), \mathbf{a}_{-j}(t)) - U_j(a_j(t), \mathbf{a}_{-j}(t)) > 0.$$

Thus, any unilateral profitable deviation increases $\Phi(\mathbf{a}(t))$.

Boundedness and Impossibility of Infinite Improvement Sequences

Since all utilities are bounded (due to finite γ, δ, η, bounded $\Delta A_i(t)$, and bounded energy resources), there exists a finite upper bound $\Phi_{\max}$ such that:

$$\Phi(\mathbf{a}(t)) \leq \Phi_{\max} \quad \forall \mathbf{a}(t).$$

Suppose, for contradiction, that there exists an infinite sequence of unilateral profitable deviations. Each such deviation strictly increases $\Phi(\mathbf{a}(t))$. Because Φ is bounded above by $\Phi_{\max}$, only a finite number of increments can occur before no further improvements are possible. This contradiction shows that no infinite improvement sequence can occur.

Existence of a Nash Equilibrium

Since no infinite sequence of profitable unilateral deviations can occur, the best-response dynamics must terminate in a state where no sensor can unilaterally improve its utility. By definition, this state is a Nash equilibrium $\mathbf{a}^*(t)$:

$$U_i(a_i^*(t), \mathbf{a}_{-i}^*(t)) \geq U_i(a_i(t), \mathbf{a}_{-i}^*(t)) \quad \forall a_i(t), \forall i.$$

Thus, the existence of a Nash equilibrium follows directly from the finiteness of utilities, the monotonicity of Φ, and the impossibility of infinite improvement sequences.

Convergence to the Nash Equilibrium

The final step is to show that the iterative best-response dynamics indeed converge to the NE identified above. Since each sensor's best-response update seeks to maximize its own utility, sensors will continue to deviate as long as profitable deviations exist. Our argument shows that profitable deviations must terminate. Under the assumptions that $\Delta A_i(t)$ is non-decreasing and that sensors have consistent energy and accuracy estimates, no cyclical behavior can persist. A cycle would imply an infinite sequence of improvements or a return to a previously visited state without improvement, which cannot occur since profitable deviations strictly increase $\Phi(\mathbf{a}(t))$.

The presence of the discount factor β further stabilizes the process. With $\beta \in [0, 1)$, sensors value future utility less than immediate utility. This discounting ensures diminishing returns for postponing beneficial participation or indefinitely waiting for ideal conditions. As a result, sensors do not continually defer improvements, preventing complex long-term cycles.

Because the best-response process eliminates profitable deviations step by step and cannot cycle indefinitely, the action profile sequence generated by iterative best responses converges to the NE.

We have shown that, with the reward-based utility function that includes correct participation rewards ($\gamma \cdot \Delta A_i(t)$), penalties for incorrect inferences (δ), and penalties for non-participation (η), the best-response dynamics lead to a Nash equilibrium. The proof relies on constructing a potential function Φ that is strictly increased by unilateral profitable deviations and bounded above. The impossibility of infinite improvement sequences guarantees the existence of an NE, and the assumptions on monotonicity, boundedness, and discounting ensure that the iterative best-response process converges to this equilibrium.

□

C. Proof of Convergence for the Equilibrium-Aware Training Process

In this appendix, we provide a comprehensive and detailed proof of the convergence theorem stated in the main text. We also elaborate on how the new loss function, the introduced regularizers, and their gradients integrate into the backpropagation and stochastic gradient descent (SGD) steps. Additionally, we discuss bounds on the newly introduced hyperparameters and provide guidelines for selecting them.

Problem Setting and Notation

We consider a global inference model $f_\theta : \mathcal{X} \to \mathcal{Y}$ parameterized by $\theta \in \mathbb{R}^d$. The model's performance is measured by a loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$ that is convex in θ for any fixed input-label pair (x, y). The model operates in an energy-harvesting wireless sensor network (EH-WSN) environment where sensors participate strategically in inference tasks based on a game-theoretic equilibrium.

Let $\mathcal{D}$ denote the effective data distribution induced by the equilibrium strategies of the sensors. Under equilibrium conditions, the distribution $\mathcal{D}$ is stationary or at least stationary over sufficiently large timescales. The expected loss is

$$L(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_\theta(x), y)].$$

To enhance robustness and efficiency, we introduce two regularizers:

$$\Omega_{\text{SNR}}(\theta) \quad \text{and} \quad \Omega_{\text{complexity}}(\theta).$$

$\Omega_{\text{SNR}}(\theta)$ encourages the model to perform reasonably well across varying SNR levels, while $\Omega_{\text{complexity}}(\theta)$ penalizes overly complex models that might demand excessive energy or communication costs. Both are assumed convex and have bounded gradients.

The final training objective is:

$$J(\theta) = L(\theta) + \lambda_1 \Omega_{\text{SNR}}(\theta) + \lambda_2 \Omega_{\text{complexity}}(\theta),$$

where $\lambda_1, \lambda_2 \geq 0$ are hyperparameters controlling the influence of the regularizers.

Our goal is to show that by running a diminishing step-size SGD on $J(\theta)$, using unbiased gradient estimates from the equilibrium distribution $\mathcal{D}$, the parameters $\{\theta_k\}$ converge in expectation to a stationary point θ^* of $J(\theta)$.

Key Assumptions and Conditions

1. **Convexity of ℓ.** The loss $\ell(f_\theta(x), y)$ is convex in θ. Consequently, the expected loss $L(\theta)$ is also convex.

2. **L-smoothness of ℓ.** There exists a constant $L > 0$ such that for all θ, θ',

   $$\|\nabla L(\theta) - \nabla L(\theta')\| \leq L\|\theta - \theta'\|.$$

   This ensures that $L(\theta)$ is Lipschitz-smooth.

3. **Convexity and boundedness of regularizers.** The regularizers Ω_{SNR} and $\Omega_{\text{complexity}}$ are convex in θ, and their gradients are bounded. Let

   $$\|\nabla\Omega_{\text{SNR}}(\theta)\| \leq G_1, \quad \|\nabla\Omega_{\text{complexity}}(\theta)\| \leq G_2 \quad \forall\theta.$$

4. **Stationary distribution $\mathcal{D}$.** The equilibrium participation strategies induce a stationary effective distribution $\mathcal{D}$. Over sufficiently large timescales, the system does not drift away from this equilibrium, and samples (x, y) can be considered drawn i.i.d. from $\mathcal{D}$.

5. **Unbiased gradient estimates.** When the aggregator requests a training update, a subset of sensors, determined by equilibrium conditions, provide local gradients. Although not all sensors participate every time, the equilibrium ensures a stable pattern of participation. Averaged over multiple rounds, the collected gradients form an unbiased estimator $\widehat{\nabla} L(\theta)$ of $\nabla L(\theta)$:

   $$\mathbb{E}[\widehat{\nabla} L(\theta)] = \nabla L(\theta).$$

   Since the regularizers are deterministic, their gradients $\nabla\Omega_{\text{SNR}}(\theta)$ and $\nabla\Omega_{\text{complexity}}(\theta)$ do not introduce bias.

Integration of Regularizers in Backpropagation and SGD

During the training iteration k:

1. **Forward pass:** Each participating sensor collects data (x, y) and evaluates $\ell(f_{\theta_k}(x), y)$.

2. **Backward pass:** The sensor computes $\nabla_\theta \ell(f_{\theta_k}(x), y)$ via standard backpropagation. To incorporate regularizers, the sensor (or the aggregator after collecting updates) adds $\lambda_1 \nabla \Omega_{\text{SNR}}(\theta_k)$ and $\lambda_2 \nabla \Omega_{\text{complexity}}(\theta_k)$. These gradients are computed analytically since the regularizers are explicit, differentiable functions of θ.

3. **Aggregation:** The aggregator averages the received gradients:

   $$\widehat{\nabla} J(\theta_k) = \widehat{\nabla} L(\theta_k) + \lambda_1 \nabla \Omega_{\text{SNR}}(\theta_k) + \lambda_2 \nabla \Omega_{\text{complexity}}(\theta_k).$$

   Since $\mathbb{E}[\widehat{\nabla} L(\theta_k)] = \nabla L(\theta_k)$, we also have $\mathbb{E}[\widehat{\nabla} J(\theta_k)] = \nabla J(\theta_k)$.

4. **Update step:** With a chosen step size α_k,

   $$\theta_{k+1} = \theta_k - \alpha_k \widehat{\nabla} J(\theta_k).$$

Diminishing Step-Size and Convergence Results

Classical convex optimization theory (see Bottou et al. (2018) or Nemirovski et al. (2009)) states that for convex, Lipschitz-smooth objectives and unbiased gradient oracles, SGD converges to a stationary point if the step sizes $\{\alpha_k\}$ decrease at an appropriate rate. A common choice is $\alpha_k = 1/\sqrt{k}$, but any diminishing sequence with $\sum_k \alpha_k = \infty$ and $\sum_k \alpha_k^2 < \infty$ works.

Under these conditions, we have:

$$\lim_{k\to\infty} \mathbb{E}[J(\theta_k)] = J(\theta^*) \quad \text{and} \quad \lim_{k\to\infty} \mathbb{E}[\|\nabla J(\theta_k)\|] = 0.$$

This implies θ_k converges in expectation to a stationary point θ^* of $J(\theta)$.

Equilibrium Stability and Impact on Stationarity

The key subtlety is that $\mathcal{D}$ depends on equilibrium strategies. However, the equilibrium ensures a stable operating regime where sensor behaviors—and thus $\mathcal{D}$—do not change drastically over time. This stability allows us to treat $\mathcal{D}$ as effectively fixed for the purpose of the asymptotic analysis. If $\mathcal{D}$ were to drift significantly, standard SGD results would not directly apply. The equilibrium prevents such non-stationary behavior in the long run.

Furthermore, since the regularizers are deterministic and have bounded gradients, they do not add pathological conditions to the optimization landscape. They may alter the shape of $J(\theta)$, encouraging certain regions of parameter space, but they do not prevent convergence. On the contrary, they may help by smoothing out undesirable minima or limiting model complexity.

Bounding and Selecting Hyperparameters λ_1, λ_2

The choice of λ_1 and λ_2 affects the curvature of $J(\theta)$ and can influence convergence speed and the location of θ^*. Some guidelines include:

1. Start with small values of λ_1 and λ_2 to avoid overwhelming the primary loss $L(\theta)$. Gradually increase them if the model relies too heavily on high-SNR data or becomes too complex.

2. Ensure $\lambda_1 \le \frac{c_1}{G_1}$ and $\lambda_2 \le \frac{c_2}{G_2}$ for some constants $c_1, c_2 > 0$, to prevent excessively large gradients due to the regularizers.

3. Tune λ_1, λ_2 based on validation performance. If the model overfits high-SNR data, increase λ_1. If it becomes too large and slow to run, increase λ_2.

By keeping λ_1, λ_2 within reasonable bounds, we ensure that the modified gradient $\widehat{\nabla} J(\theta)$ remains well-behaved, preserving the conditions for SGD convergence.

We have shown that under the stated assumptions—convexity and smoothness of ℓ, convexity and bounded gradients of Ω_{SNR} and $\Omega_{\text{complexity}}$, stationarity of $\mathcal{D}$ induced by equilibrium strategies, and unbiased gradient estimates—the diminishing step-size SGD applied to $J(\theta)$ converges in expectation to a stationary point θ^*.

The equilibrium ensures $\mathcal{D}$ remains stable, allowing classical stochastic optimization theory to hold. The regularizers, being convex and with bounded gradients, integrate seamlessly into the backpropagation and SGD updates, shaping the optimization landscape but not invalidating convergence properties. Proper selection and tuning of λ_1, λ_2 help maintain stable and robust training dynamics.

Thus, the proposed training process achieves a harmonious balance: it respects the strategic, energy-constrained environment (through equilibrium and game-theoretic considerations), while leveraging well-established convex optimization guarantees to ensure convergence of the global model parameters.