

Game Theoretic Analysis of Age of Federated Information for Participatory Data Ecosystems

Alessandro Buratto *Graduate Student Member, IEEE*, Alessio Mora *Member, IEEE*,
 Armir Bujari *Member, IEEE*, and Leonardo Badia *Senior Member, IEEE*

Abstract—We investigate a scenario where multiple sources independently and voluntarily contribute status reports, which are then aggregated through a federated process. To address the challenge of partial participation in distributed systems, we introduce the age of federated information (AoFI), a novel metric that quantifies data freshness. This metric is specifically designed to bridge the gap between classical age of information (AoI), which is unsuitable for collaborative tasks, and the often impractical age of correlated information (AoCI), which requires full participation. To model distributed optimization across multiple independent sources, we adopt a game-theoretic framework. In this framework, users strategically minimize their individual penalty, computed as a global-local combination of the overall AoFI on the common receiver's side and their individual energy expenditure. We derive the worst-case Nash equilibrium of this game and compare its efficiency with the centralized optimization optimum. Our efficiency analysis reveals a critical design trade-off for practical IIoT deployments: while decentralized coordination is highly efficient in high-participation regimes, performance in low-participation regimes is paradoxically optimized by actively restricting the number of sources to prevent strategic inefficiencies.

Index Terms—Age of federated information; Participatory sensing; Federated learning; Mobile crowdsensing; Game theory.

I. INTRODUCTION

Many modern communication networks are tightly integrated with the internet of things (IoT), a paradigm of interconnected physical devices that sense their environment, process data, and communicate over the Internet [1]. These devices are often resource-constrained, making the design of lightweight and distributed communication protocols essential to ensure efficiency and scalability.

At the same time, participatory and federated paradigms offer promising ways to leverage decentralized data sources

This work was supported by the Italian PRIN project 2022PNRR “DIGIT4CIRCLE”, project code P2022788KK and by the European Union under the Italian National Recovery and Resilience Plan (NRRP) Mission 4, Component 2, Investment 1.3, CUP C93C22005250001, partnership on “Telecommunications of the Future” (PE00000001 - program “RESTART”)

Alessandro Buratto and Leonardo Badia are with the Dept. of Information Engineering (DEI), University of Padova, Italy (email: {alessandro.buratto.1@phd., leonardo.badia@unipd.it}).

Alessio Mora and Armir Bujari are with the Dept. of Computer Science and Engineering, University of Bologna, Italy (email: {alessio.mora, armir.bujari}@unibo.it)

for distributed analytics and intelligent decision making [2], [3]. In participatory sensing, autonomous nodes independently collect and disseminate environmental data, thereby facilitating crowd-sourced real-time intelligence. However, data generated by uncoordinated sources pose a fundamental challenge, namely the need to aggregate information efficiently while minimizing both redundancy and resource consumption.

Aggregating data from disparate sources presents challenges that are relevant for various new applications. Take, for example, a predictive maintenance scenario in a smart factory, where multiple battery-operated vibration sensors supervise a crucial production line. A single sensor's alert may represent a false positive, yet a significant maintenance alert arises when several sensors together detect anomalies in the same timeframe. The certainty of this diagnosis improves as more sensors support the anomaly. Maintaining up-to-date data is essential for averting expensive breakdowns, but each sensor's energy limitations create a strategic balance between frequent transmissions and power conservation. Similarly, in participatory environmental sensing, building an accurate, localized air pollution map requires a critical mass of current reports from a specific region to create a reliable, immediate alert. In these cases, traditional AoI (with its single-update criterion) fails, prompting us to introduce a quorum-based approach to data freshness.

Federated learning addresses this by distributing the learning process across decentralized nodes [4], [5]. Although this improves diversity and resilience, it also introduces inefficiencies due to limited coordination. Traditional metrics such as throughput or delay fail in capturing the timeliness of data, as a high throughput may consist of only old updates and a short delay may occur even with infrequent updates. A suitable performance metric for these applications is the age of information (AoI), which measures the freshness of received data [6], [7].

Despite growing interest in AoI for distributed sensing, there remains a gap in understanding how participatory and federated systems behave when nodes act independently and selfishly. Specifically, existing work does not sufficiently address how coordination (or lack thereof) affects the AoI in decentralized setups, nor how to incentivize optimal behavior. Game theory has recently been explored to model this challenge [8], [9], especially in contexts where devices cannot be centrally controlled.

To this end, we investigate a game-theoretic model of

N independent sources monitoring the same physical phenomenon. These nodes decide, over repeated rounds, whether to transmit updates, balancing a global AoI objective with individual transmission costs [10]. While AoI is highly effective for systems where updates from any single source are interchangeable, this model does not capture the nature of federated and collaborative tasks, where data from a single node is often insufficient. For example, in federated learning, a global model update requires aggregating models from multiple clients. At the other extreme lies the age of *correlated* information (AoCI) [11], which defines a successful update as a complete “snapshot” of data from *all* participating nodes. This requirement, while useful for perfectly synchronized systems, is often too brittle for practical wireless deployments where full participation cannot be guaranteed.

Our work addresses the space between these two extremes. We introduce *Age of Federated Information* (AoFI), a metric designed for systems with partial participation requirements. AoFI measures the time elapsed since the receiver successfully aggregated updates from a sufficient, but not necessarily complete, subset of sources. It is this ability to model a *quorum-based* success condition that distinguishes AoFI from its predecessors and makes it an essential tool for analyzing the freshness of participatory data ecosystems, such as federated sensing, crowdsourcing, and distributed analytics.

This setup reveals a tradeoff: with minimal participant requirements, selfish nodes may depend on others for transmission, deteriorating AoFI. As participation demands increase, AoFI initially increases due to node selfishness [12], then decreases when almost complete collaboration is required, aligning with AoCI findings [11]. AoFI is a valuable tool for planning and assessing data freshness with partial participation, supporting informed decisions in systems, such as industrial IoT, predictive maintenance, or real-time monitoring.

Finally, we compute the price of anarchy (PoA) and show that the Nash equilibrium (NE) can be highly inefficient compared to centralized solutions, except in cases requiring full participation. Thus, we argue that even when full collaboration is not strictly necessary, enforcing it can significantly improve system performance.

The main contributions of this article are summarized as follows:

- We introduce the AoFI metric to bridge the gap in the literature between *anycast*-based status updates and fully collaborative network updates, as required by the AoCI.
- We prove the existence of multiple Nash equilibria resulting from the interaction between sensors, and we focus our analysis on the least performing one on a system-wide scale to derive lower bounds on the potential inefficiencies of a partially collaborative scenario, as measured by the PoA.
- We validate our system model using a real-world dataset of IoT nodes and demonstrate that the proposed modeling assumptions are consistent with practical deployment scenarios.

The remainder of this article is organized as follows. Sect. II discusses the relevant literature and compares our contribution to previous studies. Sect. III presents the mathematical model

| Reference | AoI | Game Theoretic | Collaborative Setup | Efficiency of Distributed Solution |
|--------------|-----|----------------|---------------------|------------------------------------|
| [3] | ✓ | | ✓ | |
| [7] | ✓ | | ✓ | |
| [9] | ✓ | ✓ | | |
| [11] | ✓ | | ✓ | ✓ |
| [12] | ✓ | ✓ | ✓ | |
| [13] | | ✓ | ✓ | |
| [14] | ✓ | ✓ | | |
| [15] | | ✓ | ✓ | ✓ |
| [16] | | ✓ | ✓ | |
| [17] | | ✓ | ✓ | ✓ |
| [18] | | ✓ | ✓ | |
| This article | ✓ | ✓ | ✓ | ✓ |

TABLE I
COMPARISON BETWEEN RELATED WORKS.

of the targeted scenario, defines the structure of the game theoretic problem, and derives the NE solution. Sect. IV gives numerical results. Finally, Sect. V draws the conclusions.

II. RELATED WORK

This work lies at a crucial intersection of information freshness metrics and game-theoretic models for distributed systems. Game-theoretic approaches have long provided a foundational framework for analyzing network scenarios involving self-controlling agents that compete over shared resources. Influential work in this area has built models about medium access control and resource allocation and has shown how far strategic interaction shapes system effectiveness [19]. We adapt this classic paradigm to solve the specific modern problems associated with keeping data fresh in participatory systems.

Although the foundational research [20] introduced the concept of AoI as a critical notion, only a handful of works approach it through a game-theoretic lens, where the objectives of the players are directly influenced by goals related to AoI. The prevailing body of literature can be classified into two main categories. A significant category emphasizes adversarial settings, in which a malicious participant intends to exacerbate the AoI of others [14], [21]. In contrast, another category examines resources-limited games, where various sources compete to provide updates, and each player tries to reduce their own AoI [1], [22], [23].

In comparison to these purely adversarial or competitive models, our situation is semi-cooperative: all individuals are driven by a shared objective to decrease some global measure of freshness, yet are also aware of minimizing their own costs. This dynamic is characteristic of participatory platforms, from the mobile crowd sensing systems detailed in [24] to the neighborhood-scale energy management investigated in [17], where game-theoretic frameworks are applied to facilitate efficient collaborative outcomes. In this domain, incentive mechanisms are commonly formulated through bargaining or auction schemes to ensure desirable interactions [13], [16]. Although important, these works often do not address the difficulty of ineffective updates in the event that the quota for participants is not met. Consequently, other studies introduce AoI to federated learning frameworks as part of some incentive or scheduling mechanism to promote cooperation between

nodes [7] or use a weighted measure of AoI to optimize coverage quality [3], rather than analyzing general strategic inefficiencies that arise without such schemes.

The formulation of relevant models for private costs is essential. These costs may encompass physical quantities such as energy use [25] or include intangible considerations such as privacy risks, security, and trust in data [26], [27]. Our model facilitates these multifaceted issues through a parametric model. The parametric form of our utility functions enables us to incorporate these considerations in a linear combination similar to [9], aligning the semantic objective of minimizing AoI with these cost-related considerations. In addition, we study a complete information game. Although most models utilize incomplete knowledge to model private user information, our assumption enables an analytically tractable description of the worst-case inefficiency and serves as a fundamental baseline from which to understand system limits on achievable performance. Complementing the direct application of such analytical models is an emerging methodological trend that seeks to address the inherent complexity of their formulation; recent work, for instance, proposes leveraging Generative AI frameworks to automate the construction and solution of game-theoretic models for networking, with the goal of making these powerful tools more accessible to system designers [18].

Although globally cooperative and centralized architectures can achieve a lower peak AoI [11], such approaches are impractical for the voluntary and decentralized systems that are the focus of our work. The work most similar to our scenario is [12], which considers multiple sources with the aim of collectively obtaining the minimum value of AoI. However, the work lacks rigorous proof regarding various equilibria and does not consider the typical inefficiencies in a distributed system. Our work directly addresses these issues. In this work we introduce the concept of AoFI, bridging the gap between AoI and AoCI. Another metric that addresses collaboration of nodes stemming from the AoI is the age of collection (AoC) [28], and in our scenario it is identical to the AoCI. The idea of *partial collaboration* can be naturally extended to other freshness metrics. For instance, similar quorum-based mechanisms could be formulated for the age of incorrect information (AoII), which measures the timeliness of accurate rather than merely recent updates [29]; or the version age of information (VAoI), which captures the staleness of versioned updates in distributed systems [30]. A formal treatment of these extensions is beyond the scope of this article but represents a promising direction for future research. Table I outlines these distinctions, emphasizing the novelties of our contribution.

III. SYSTEM MODEL

A. System Overview

Our system involves a set of N nodes, which are potential participants in a given task, and a centralized aggregator server R , also referred to as the *receiver*. We consider a discrete-time axis divided into slots. The receiver is interested in keeping fresh information of the process monitored by the nodes,

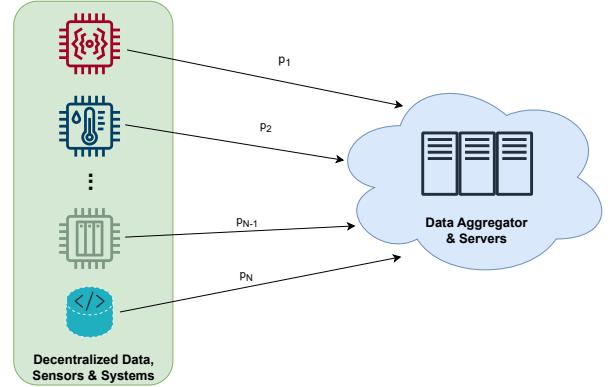


Fig. 1. High-level description of the considered scenario. Nodes, representing sensors or systems, participate in a federated process, voluntarily sourcing data related to the environment or a monitored phenomenon with a probability p_i .

which is tracked by the instantaneous AoI, denoted as $\delta(t)$ and defined as the current time instant t minus the timestamp of the last successful update t_k [6]

$$\delta(t) = t - t_k. \quad (1)$$

The nodes can choose, independently of each other, their transmission probability p_i for a specific slot. We further assume that nodes always have data to transmit (generate-at-will model) and we neglect the transmission delay, as commonly done in the literature [6], [9]. This consideration holds true not only in an analytical scenario, but also in the presence of low power networks such as LoRaWAN and energy harvesting sensors for which the interupdate times are much larger than the transmission delays [31], [32]. See Fig. 1 for a graphical representation of the scenario.

In each slot, the nodes may independently decide to participate in the task. The literature offers some similar analysis in the context of *medium access control*, where the nodes belong to the same collision domain and the goal is to coordinate their transmission in a distributed fashion [9], [20].

Here, we look instead at the sensing or federated process from the perspective of the upper layers, and the physical location of the nodes is irrelevant, as long as they are all able to collect information about the process being monitored. In this context, we are interested in describing the strategic interaction of the players when making the crowdsensing decision. A complete summary of the notation adopted in this work is reported in Tab. II.

B. Success Probability Model

Centralized optimization techniques can minimize peak AoI with fully correlated information sources [11]. However, if fewer than a threshold m nodes participate in a round, the receiver may not fully understand the measured phenomenon. To address this, we define a concave function that reduces the success probability when some nodes do not transmit, even if at least m are collaborating. This accounts for data noise or the need for sufficient collaboration to understand certain characteristics of the phenomenon.

| Notation | Description |
|-------------------|---|
| N | number of nodes |
| $\delta(t)$ | instantaneous AoI at time t |
| C_i | individual node cost term |
| \tilde{c} | normalized cost factor |
| p_i | transmission probability of node i |
| P_{succ} | probability of successful transmission |
| $Q(\cdot)$ | survival rate |
| m | minimal number of nodes required to participate |
| α | exponent of the success rate |
| θ_{\min} | minimal acceptable ML model accuracy |
| \mathcal{P}_i | penalty of node i |

TABLE II
MATHEMATICAL NOTATIONS USED IN THE ARTICLE.

In our analysis, we focus on a generalized case, where we want to model the success probability with a soft gradient along the lines of [12]. Furthermore, we want to account both for the requirement of a minimum number of participants and that this is not sufficient to guarantee the task's success. In particular, the conditional probability of a successful update as a function of the number of transmitting nodes is taken as a strictly increasing concave function, regulated by tunable parameters, such that it increases if more than m nodes transmit and is equal to 1 only when $m = N$. This design criterion results in choosing

$$P[\text{succ}|x] = \begin{cases} \left(\sqrt{1 - \frac{(x-N)^2}{(N-m+1)^2}}\right)^\alpha & \text{if } m \leq x \leq N \\ 0 & \text{if } 0 < x < m \end{cases} \quad (2)$$

where the exponent $\alpha \geq 0$ is used to adjust the steepness of the growth of the success rate.

We focus on two versions of the model: (i) hard threshold (HT), where $\alpha = 0$, i.e., the communication attempt is successful with probability 1 if *at least* m nodes collaborate and (ii) soft threshold (ST), where we choose $\alpha = 3$; in the latter, the success probability is not 1 unless *all* nodes collaborate in the same round.

To validate this analytical formulation in a real scenario, we focus on the accuracy reached by an ML model trained with the Edge-IIoTset dataset [33]. This dataset contains traffic logs from 10 different IoT nodes in both a normal operation and a malicious scenario. We train a random forest classifier to convergence, varying the number of sensors contributing to the data collection. After averaging the obtained accuracies over all possible sensor combinations, we obtain the accuracy values displayed in Fig. 2. We also introduce a parameter θ_{\min} which indicates the minimal accuracy we want the ML model to achieve. If the number of collaborating nodes is not enough to achieve an accuracy above this threshold, we set $P[\text{succ} | x] = 0$. Otherwise, this conditional probability is proportional to its position in the interval between the maximum achievable accuracy and the threshold. With this setup, choosing θ_{\min} maps directly to choosing m , i.e., deciding how many nodes are needed to collaborate. See Fig. 3 for a graphical representation of (2) and the real data case.

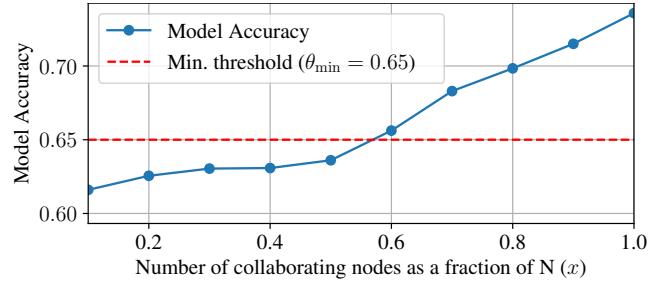


Fig. 2. Mean accuracy achieved by the ML model as a function of the number of collaborating nodes.

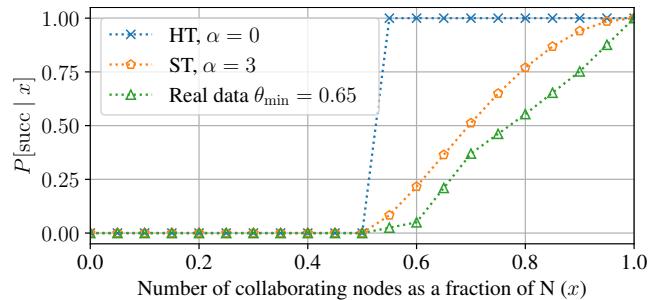


Fig. 3. Conditional success probability function for $m = 0.55N$ in HT, ST and real data with $\theta_{\min} = 0.65$ cases.

C. Expected AoFI and Success Rate

The expected AoI for a receiver getting independent updates that may or may not be successful is computed as [6]

$$\mathbb{E}[\delta] = \frac{1}{P_{\text{succ}}} - 1, \quad (3)$$

where P_{succ} is the probability of a successful update to occur. In turn, this event depends on whether enough nodes transmit and, since they are independent sources, P_{succ} is the survival rate of a Poisson-binomial distribution X .

If every transmission attempt with at least m simultaneous transmissions always results in a successful update, we can write:

$$P_{\text{succ}} = Q(m) = \sum_{t=m}^N P[X = t], \quad (4)$$

Applying our thresholding functions, the law of total probability gives:

$$P_{\text{succ}} = \sum_{t=m}^N P[X = t] \cdot P[\text{succ} | X = t]. \quad (5)$$

Expanding further using [34] for the probability mass function (PMF) of the Poisson-binomial distribution, we derive:

$$P_{\text{succ}} = \sum_{n=0}^N \left\{ \left[\sum_{t=m}^N P[\text{succ} | X=t] \cdot \exp\left(-\frac{2\pi j n t}{N+1}\right) \right] \cdot \prod_{\ell=1}^N \left(p_\ell \left(\exp\left(\frac{2\pi j n}{N+1}\right) - 1 \right) + 1 \right) \right\} / (N+1) \quad (6)$$

where j is the imaginary unit.

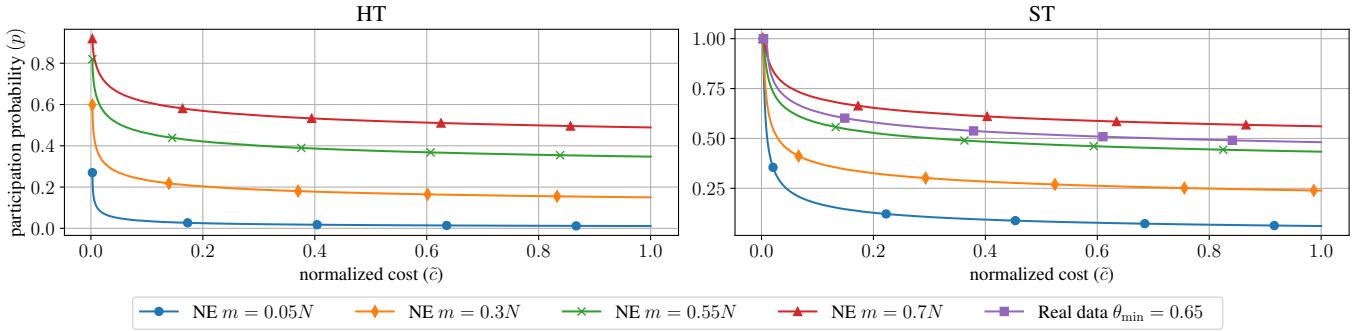


Fig. 4. Transmission probability with the NE for different threshold profiles, LR.

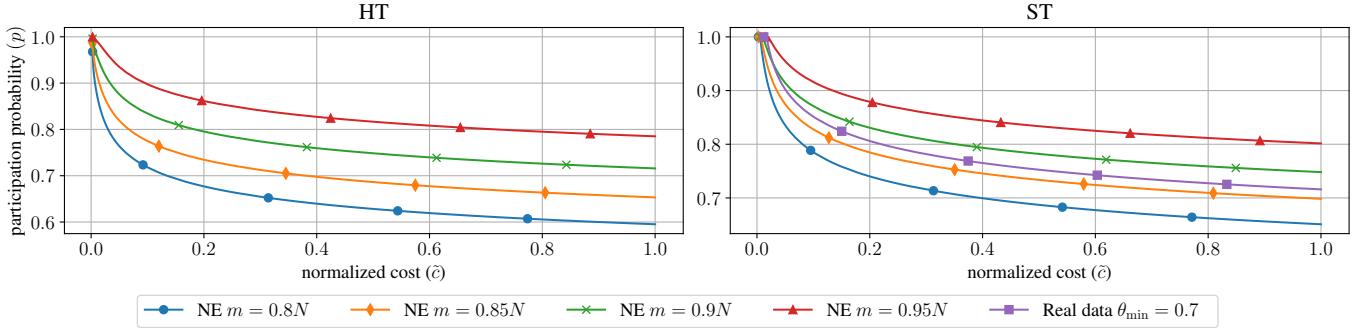


Fig. 5. Transmission probability with the NE for different threshold profiles, HR.

D. Penalty Function and Node Cost

Each node i has a penalty defined as [9]

$$\mathcal{P}_i = \mathbb{E}[\delta] + \mathcal{C}_i = \frac{1}{P_{\text{succ}}} - 1 + cp_i, \quad (7)$$

where the cost term is:

$$\mathcal{C}_i = cp_i. \quad (8)$$

Note the concordant sign for both terms in (7), since the players in the game seek to minimize both AoFI and cost.

E. Game Theoretic Setup

From a game theory perspective, the interactions among the nodes can be effectively represented as a *static game of complete information*, denoted by $\mathcal{G} = (\mathcal{S}, \mathcal{A}, \mathcal{U})$. The decision to use this model is driven by the characteristics of such games. Indeed, the complete information assumption¹ facilitates a thorough examination of the role of uniform sensors, focusing uncertainty exclusively on their actions rather than on the sensor types. This approach also enables an analytically manageable description of NEs and PoA. In this framework, $\mathcal{S} = \{S_1, S_2, \dots, S_N\}$ is the set of players, where $|\mathcal{S}| = N$ as the receiver is treated as a passive entity, meaning that it does not actively participate in decision making. The set of possible actions \mathcal{A} represents the transmission probabilities $p_i \in [0, 1]$ chosen by each player $i \in \mathcal{S}$. The utilities \mathcal{U} describe the trade-offs faced by the players, reflecting penalties or rewards

¹The game theoretic terminology of *complete information* only refers to the data related to the game, that is, it implies that each player is aware of the existence of other nodes in the federated task, and also that these are rational agents; it does not imply any exchange of data among the federated nodes, which keep their content as private.

based on their actions and the resulting network performance. The utility function for each player is formally defined in (7).

The NE of \mathcal{G} is obtained through a one-sided optimization of the utility, i.e. each player looks for a *best response* to the unchanged actions of the other players. Each player's choice of transmission probability not only affects their own utility but also influences the utilities of other players, which must be considered in the optimization. Without loss of generality, we will focus on player 1 as the solution for the NE of the others is symmetric as proven by the following theorem.

Theorem 1. All NE in mixed strategies of game \mathcal{G} are symmetric, i.e. all nodes transmit with the same probability $p_1 = \dots = p_N$. ■

Proof: See Appendix II.

An NE must satisfy the condition

$$\frac{\partial \mathcal{P}_1}{\partial p_1} = -\frac{\partial P_{\text{succ}}}{\partial p_1} \cdot \frac{1}{(P_{\text{succ}})^2} + c = 0, \quad (9)$$

that can be computed by applying the chain rule on (7). The derivative of P_{succ} with respect to p_1 can be obtained in closed-form from (6) with simple derivation rules, or alternatively, as in Appendix I.

Theorem 2. Game \mathcal{G} with HT admits multiple feasible NEs within 3 categories depending on the relationship between the number of active participating nodes μ and the minimum number of participating nodes m : if $\mu = m$ all active nodes participate with the same probability $p^{(m)} = 1$. If $m < \mu \leq N$ there is an NE where μ nodes are active but $p^{(\mu)} < p^{(m)}$; additionally, there are $\binom{\mu}{\zeta}$ pure and mixed strategies NEs, with $\zeta = 1, \dots, \kappa$ and $\kappa = \mu - m$, where $N - \mu + \zeta$ nodes are silent.

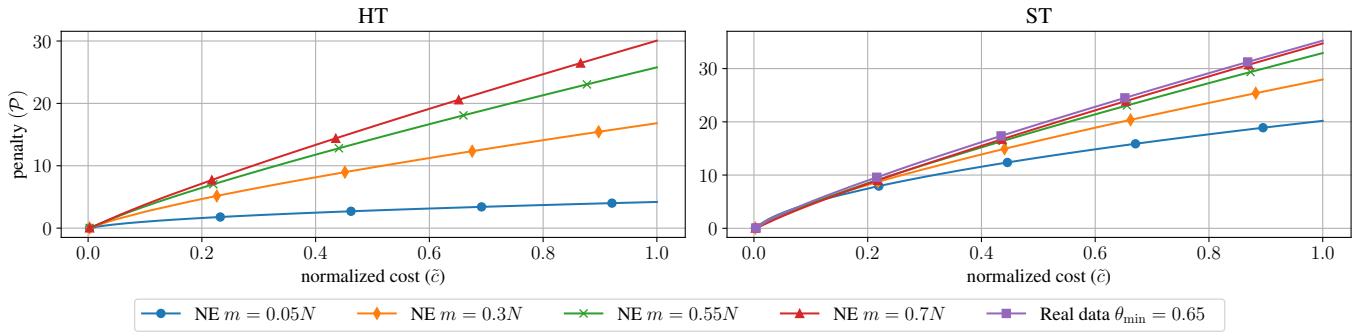


Fig. 6. Penalty with the NE for different threshold profiles, LR.

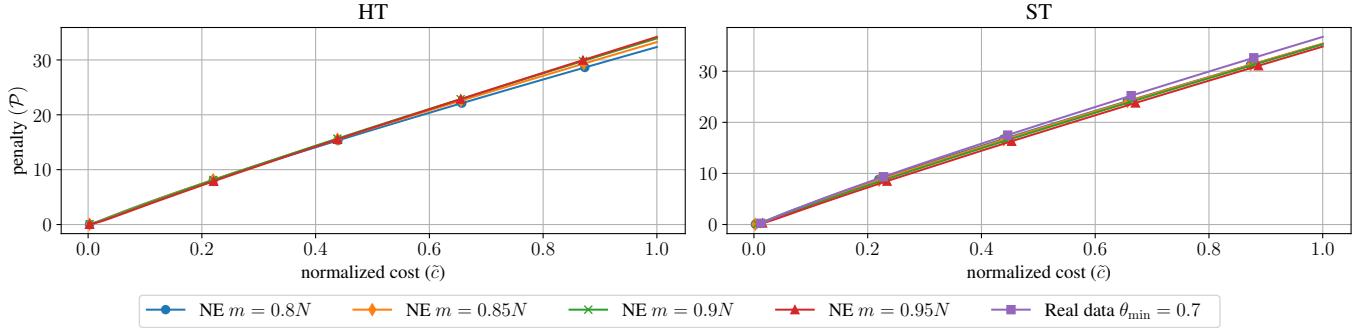


Fig. 7. Penalty with the NE for different threshold profiles, HR.

Proof: See Appendix III. \blacksquare

Theorem 3. *Game \mathcal{G} with ST admits fewer NEs compared to HT, as there exists an $m^* \geq m$ that indicates the emergence of NE. For $\mu < m^*$ there exists only a catastrophic equilibrium in which all the nodes decide to stay silent.*

Proof: See Appendix IV. \blacksquare

Theorem 4. *In game \mathcal{G} with HT, $m \rightarrow 1$ is the best Pareto efficient working point for the system.*

Proof: See Appendix V. \blacksquare

We further calculate the PoA to evaluate how much a decentralized solution deteriorates the optimal centralized one's performance. The PoA is calculated as [35]

$$\text{PoA} = \frac{\max_{s \in \text{NE}} \sum_i \mathcal{P}_i^s}{\min_{s \in \mathcal{T}} \sum_i \mathcal{P}_i^s} \quad (10)$$

where, in the numerator, we take strategy s from the set of NEs that maximize the penalty, and at the denominator, the optimal centralized strategy that gets the lowest social penalty.

IV. RESULTS

In this section, we analyze the equations for the threshold functions that were previously defined, considering various values for the number of collaborating nodes m , while keeping the total node count N constant. For analytical results, we set $N = 20$, and used $N = 10$ for simulations with real data [33]. The generalizability of our findings is introduced by a scale-invariant model design based on two key normalizations. First, the strategic behavior of the nodes is governed by the relative participation threshold m/N , which defines partial collaboration independent of the absolute system size N . Equally

important is the normalized cost, $\tilde{c} = c/N$, which is critical to preserving the strategic trade-off at any scale. Without this normalization, as N grows, the impact of a fixed individual cost would become negligible compared to the global AoFI penalty, trivializing the game-theoretic decision. By scaling the cost, we ensure the tension between individual sacrifice and collective benefit remains important for nodes' policies. Consequently, the system's equilibria depend on these relative parameters rather than absolute node counts, which ensures the scalability of our conclusions. We further differentiate the plots for two utilization regimes: Low utilization Regime (LR) when the participation threshold is $m \leq 0.7N$ and High utilization Regime (HR). This subdivision is motivated by the different behavior of the system when subjected to different loads. The results for the standard AoI metric is reported for $m \rightarrow 0.05N$ and HT, similarly AoCI is found for $m \rightarrow N$.

Fig. 4 illustrates the NE transmission probabilities within an LR setting. As anticipated, higher m values necessitate increased transmission probabilities, since overly reducing a node's transmission probability is not advantageous. Such a reduction would lead to an undesirable increase in AoFI. Interestingly, as $m \rightarrow N$, not all nodes transmit with probability 1 when \tilde{c} is high (Fig. 5). This reflects selfish behavior, as nodes reduce transmission to avoid communication costs if the expected AoFI impact is minimal.

Fig. 6 shows the penalty values obtained by the nodes in the NE solution. For HT, penalties increase almost linearly as costs increase. In contrast, for ST, the curves for lower utilization factors become more concave and, for certain normalized cost values, they result in higher penalties compared to those with higher utilization factors. This plot additionally confirms Theorem 4, demonstrating that the smallest penalties

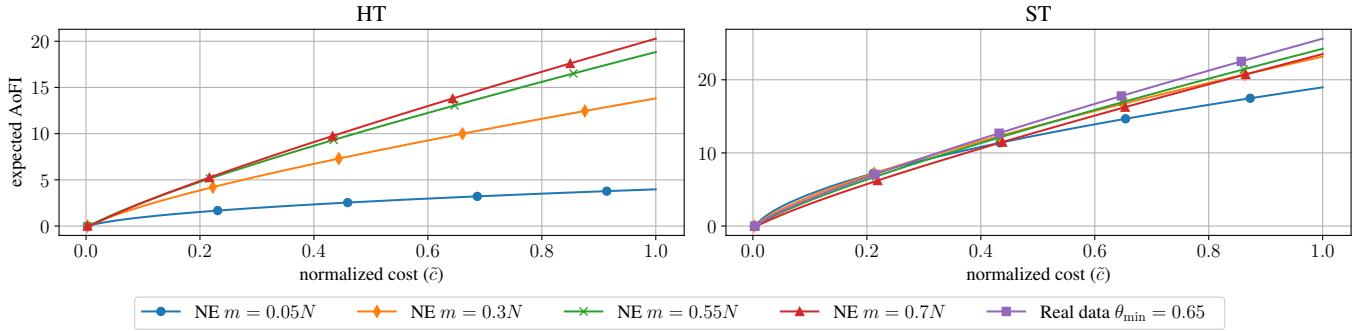


Fig. 8. Expected AoFI with the NE for different threshold profiles, LR.

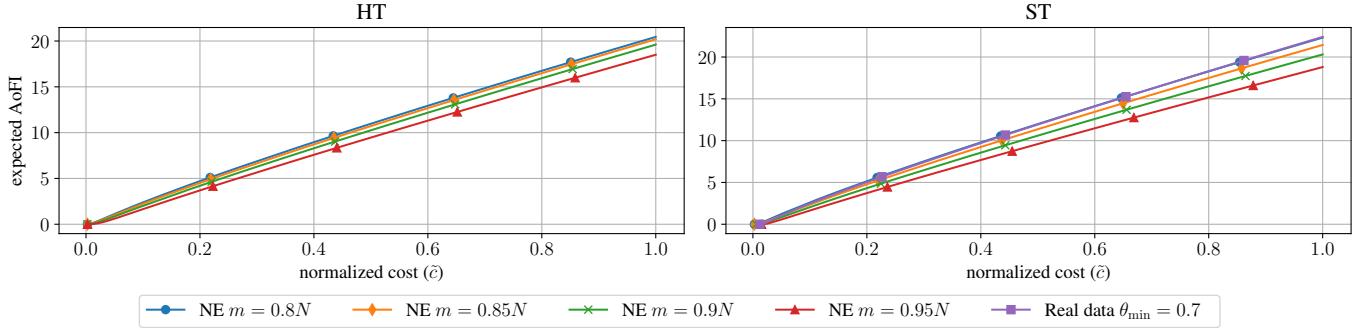


Fig. 9. Expected AoFI with the NE for different threshold profiles, HR.

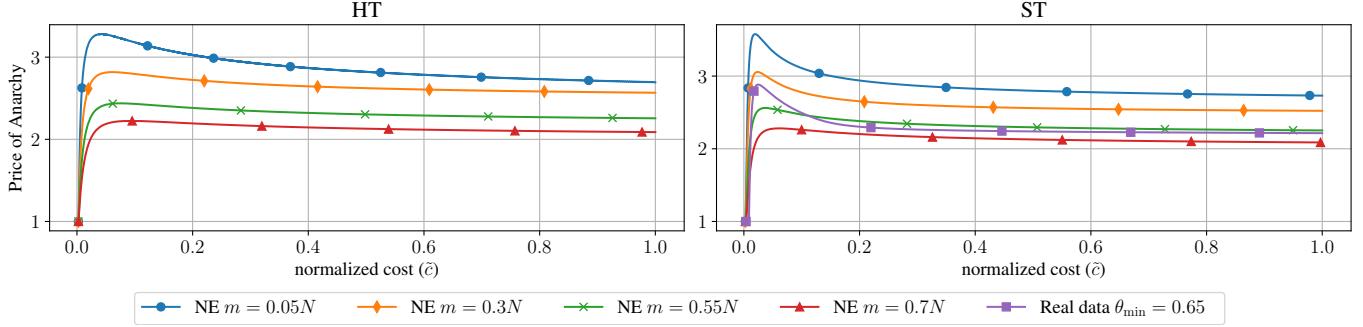


Fig. 10. PoA for different threshold profiles, LR.

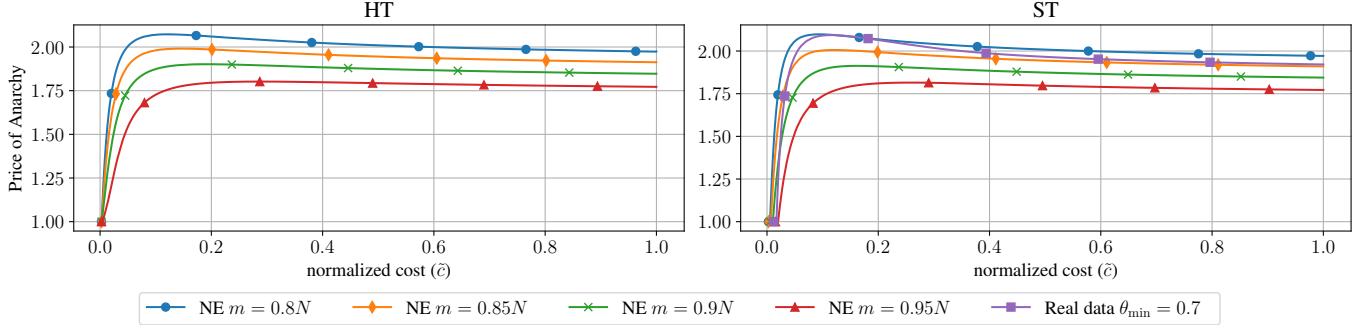


Fig. 11. PoA for different threshold profiles, HR.

for HT are achieved with the smallest m . In Fig. 7, both the HT and ST curves generally follow the same trend, but for ST, lower utilization factors underperform across nearly all cost values. This occurs because the nodes have a higher transmission probability with ST than with HT. With real data, the penalty remains highest, attributed to the decrease in

success probability compared to the analytical model, leading to more frequent transmissions to prevent AoFI divergence.

Fig. 8 shows the expected AoFI obtained in the NE for the LR case. Similarly to the penalty discussed in the previous paragraph, the ST increases the concavity of the curves as the normalized cost increases. An important remark is that

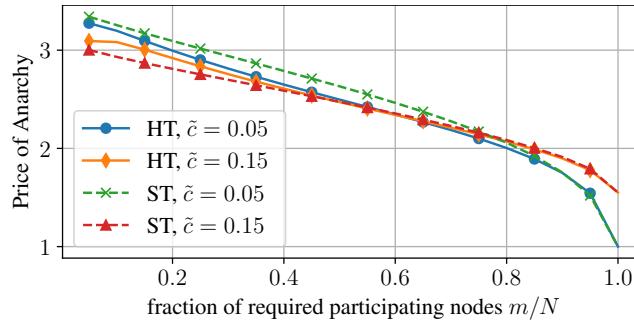


Fig. 12. PoA obtained for different required number of participating nodes for selected values of parameters α and c .

for HT there is a clear advantage in lowering AoFI when m is close to $m = 0.1N$. This situation drastically changes when ST is applied. Fig. 9 shows the expected AoFI in the HR case. Requiring more nodes to work together clearly enhances efficiency in reducing AoFI. Additionally, this graph demonstrates how resilient participation regimes are to the type of threshold needed for a successful update. As shown in Fig. 5, when $m \approx N$, the participation probability does not see a notable increase between HT and ST. This indicates that the expected number of participating nodes per communication round is already sufficient for success. An interesting trend in Fig. 9 is that the expected AoFI $\mathbb{E}[\delta]$ does not consistently decrease with m due to the strategic behavior of the players. As m rises, more nodes engage, reducing AoFI even if the task success probability drops. Yet, higher costs dampen participation, raising AoFI and limiting success likelihood. This highlights the significant role of cost: low costs motivate engagement in difficult tasks, while high costs deter it. The curves for the real dataset follow similar patterns, which reinforces that a lower success probability escalates AoFI.

Fig. 10 shows the PoA, computed as in (10), as a function of the cost factor c in LR. The NE solution is always less efficient than the centralized one, with a PoA above 2 for all \tilde{c} and m values examined, highlighting the selfish behavior of the nodes. In LR, nodes lower their participation probability for individual gain, causing inefficient resource use since tasks succeed with limited participation, a phenomenon known as the *tragedy of the commons* [36]. In the top graphs, the threshold and strategic interactions affect curve concavity, with partial participation decreasing success probability, thereby reducing PoA and slightly mitigating inefficiency. However, as shown in Fig. 9, even for $m = 0.95N$, the NE solution remains far from optimal. Small costs notably increase inefficiency, emphasizing their effect on the NE solution.

Fig. 12 shows the PoA as a function of the fraction of participating nodes m/N for some values of the normalized cost \tilde{c} and threshold profiles. As shown in Fig. 10 and Fig. 11, all curves decrease steadily as more nodes participate, with higher m curves consistently below curves associated with a lower m . The impact of applying ST on increasing PoA is only apparent when $m \ll N$, diminishing with more participating nodes. Conversely, c sets the efficiency limit when full node collaboration is required. Interestingly, when $m \ll N$, a higher cost results in a more efficient outcome using both

| Metric | Low Utilization (LR) | High Utilization (HR) |
|-------------------------------|--|--|
| Tx Prob. (p) | Increases with m ; lower for high \tilde{c} ; not all nodes transmit at $m \rightarrow N$ (selfish behavior) | High even at $m \rightarrow N$; HT and ST curves nearly overlap |
| Penalty (\mathcal{P}) | HT: grows linearly; ST: more concave and higher for small m ; real data has highest penalty | Similar trends; ST performs worse at low m |
| AoFI ($\mathbb{E}[\delta]$) | HT: lowest at small m ; ST: less efficient, especially for low m | Higher m improves AoFI; cost \tilde{c} significantly impacts participation |
| PoA | Always > 2 ; worst at low m and low \tilde{c} ; <i>tragedy of the commons</i> visible | Decreases with m ; ST increases PoA mainly when $m \ll N$ |

TABLE III
SUMMARY OF MAIN RESULTS ACROSS UTILIZATION REGIMES

threshold profiles, this is especially noticeable with ST.

A summary of the results is reported in Tab. III.

V. CONCLUSIONS

This study introduces the AoFI, a new metric that measures information freshness in partially collaborative contexts, bridging the gap between the standard AoI and AoCI. Our model considers updates to be successful if a sufficient number of nodes transmit at the same time. This condition was modeled with a monotonically increasing function that decreased the likelihood of success when too few nodes participated. Explicit equations for the average AoFI revealed a symmetric NE, where nodes used the same transmission probability.

In systems with high demand, stricter requirements for successful updates can decrease the AoFI at the receiver compared to lenient conditions, as shown through PoA analysis. Controlling transmission cost, it is advantageous to have all nodes cooperate, ensuring the lowest AoFI and minimal PoA. Conversely, in low-demand systems, it is advantageous to involve only the essential number of nodes required for success. This insight is further supported by experiments with real data, which validated the theoretical model and demonstrated consistent results.

Future directions for this research might consider the study of incentive mechanism to reduce the PoA when only a marginal part of the network needs to participate in the task. Another interesting direction might consider the interaction of sensors with different capabilities and characteristics, thus including the role of partial information in the analysis.

APPENDIX I

Alternative formulation for partial derivative of P_{succ}
Let X be a random variable with Poisson binomial distribution and let $G_X(z) = \prod_{i=1}^N (1-p_i + p_i z)$ be its probability generating function (PGF). It is a well known result that the relationship between the probability mass function (PMF) and PGF of any discrete distribution is

$$P[X = t] = \frac{1}{t!} \left. \frac{\partial^t G_X(z)}{\partial z^t} \right|_{z=0}. \quad (11)$$

Derivating (4) w.r.t. a generic p_i we get

$$\frac{\partial P_{\text{succ}}}{\partial p_i} = \sum_{t=m}^N w(t) \frac{\partial P[X = t]}{\partial p_i}, \quad (12)$$

where we substituted $w(t) = P[\text{succ} \mid X = t]$ for the sake of compactness. The partial derivative of the PMF is given by

$$\begin{aligned} \frac{\partial P[X=t]}{\partial p_i} &= \frac{1}{t!} \left. \frac{\partial^t}{\partial z^t} \frac{\partial G_X(z)}{\partial p_i} \right|_{z=0} \\ &= \frac{1}{t!} \left[(z-1) \frac{\partial^t G_{X-i}(z)}{\partial z^t} + t \frac{\partial^{t-1} G_{X-i}(z)}{\partial z^{t-1}} \right] \Big|_{z=0} \end{aligned} \quad (13)$$

in which we applied the product rule to the derivative of the PGF w.r.t. p_i and we have defined $G_{X-i}(z)$ as the PGF of the Poisson binomial random variable X_{-i} , that is a modified version of X where we have removed the participation of node i . After some algebra and substituting the result into (12) we finally obtain

$$\frac{\partial P_{\text{succ}}}{\partial p_i} = \sum_{t=m}^N w(t) (P[X_{-i} = t-1] - P[X_{-i} = t]). \quad (14)$$

In particular, for HT in which $w(t) = 1$ for every $t \geq m$ the expression further simplifies to

$$\frac{\partial P_{\text{succ}}}{\partial p_i} = P[X_{-i} = m-1]. \quad (15)$$

APPENDIX II

Proof of Theorem 1 We will explicitly prove the theorem only for the HT case. The ST case follows exactly the same steps. Let μ denote the set of active nodes, representing those participating in the communication process. An NE is obtained when for every player $i \in \mathcal{S}$ it holds that the derivative of the penalty w.r.t. the node's activation probability is 0. Consider the system of (9) for nodes i and j . Solving it, we get

$$\frac{\partial P_{\text{succ}}}{\partial p_i} = \frac{\partial P_{\text{succ}}}{\partial p_j}, \quad (16)$$

as c and P_{succ} are the same for both. By applying the results of Appendix I for HT, i.e. (15), we obtain

$$P[X_{-i} = \mu - 1] = P[X_{-j} = \mu - 1]. \quad (17)$$

We can further subdivide X_{-i} and X_{-j} into random variables Z , defined as the successes of $\mu - 2$ nodes excluding i and j and the Bernoulli trials Y_i and Y_j for nodes i and j , respectively. By applying the law of total probability we get

$$\begin{aligned} P[X_{-i} = \mu - 1] &= \sum_{k \in \{0,1\}} P[X_{-i} = \mu - 1 \mid Y_j = k] P[Y_j = k] \\ &= P[Z = \mu - 1](1 - p_j) + P[Z = \mu - 2]p_j, \end{aligned} \quad (18)$$

and with exchanged pedices for j . By plugging (18) into (17) and isolating p_i and p_j , we promptly get

$$(p_i - p_j)(P[Z = \mu - 2] - P[Z = \mu - 1]) = 0, \quad (19)$$

for which the only solution is $p_i = p_j$, because Z is unimodal and is not uniformly distributed, thus proving the claim of symmetry for the NEs.

APPENDIX III

Proof of Theorem 2 Let μ denote the count of active nodes, representing those participating in the communication process, and recall that for HT the success rate is binary, 1 if $\mu \geq m$ and 0 otherwise. The proof follows an iterative reasoning. As a base case, consider $\mu = m$. In this situation, the number of transmitting nodes is guaranteed to be $X = m$, and they all must transmit with probability $p^{(m)} = 1$, therefore $P_{\text{succ}} = 1$. In this scenario, this is a stable solution because if node i decides to defect by not transmitting with probability 1, then the task will immediately fail and its penalty will be $\mathcal{P}_i \rightarrow +\infty > 1+c$. Let us now consider the case $m < \mu \leq N$. In this scenario an equilibrium in mixed strategies emerges. To analytically obtain it, we solve (9) by rearranging the terms and applying symmetry considerations from Theorem 1 obtaining

$$R(p, \mu) = c, \quad (20)$$

with $X^{(\mu)}$ becoming a binomial random variable of parameters μ and p , $X_{-i}^{(\mu)}$ the same random variable without the effect of node i , and

$$\begin{aligned} R(p, \mu) &= \frac{\partial P_{\text{succ}}}{\partial p_i} \cdot \frac{1}{(P_{\text{succ}})^2} \\ &= \frac{P[X_{-i}^{(\mu)} = m-1]}{(P[X^{(\mu)} \geq m])^2}. \end{aligned} \quad (21)$$

By analyzing the boundary behavior of $R(p, \mu)$ in the interval $(0, 1]$ we get that for $p \rightarrow 0$ we get $R(p, \mu) \sim \frac{1}{p^{m-1}} \rightarrow +\infty$, and for $p = 1$ we have $R(1, \mu) = 0$. Because $R(p, \mu)$ is continuous, by the intermediate value theorem and the fact that it is a strictly decreasing function in p , we find that there exists a single solution $p^{(\mu)} \in (0, 1]$. We also analyze the behavior of the function for fixed p . Let $\mu' = \mu + 1$, we have $R(p, \mu') < R(p, \mu)$, because the denominator of $R(p, \mu')$ is for sure bigger than the one of $R(p, \mu)$, as one extra active node makes it easier to reach the threshold target. If we substitute $p^{(\mu)}$ in the previous relations, we get that $R(p^{(\mu)}, \mu') < R(p^{(\mu)}, \mu) = c$, meaning that in order for the μ' solution to match the value of c , it must hold that $p^{(\mu')} < p^{(\mu)}$. In addition, it is immediate to prove iteratively that for each $m < \mu \leq N$ there is a combinatorial number of NEs. Let us focus on the case $\mu = m + 1$. In this case we have a symmetric mixed strategy NE as described before that uses all $m + 1$ active nodes with probability $p^{(m+1)} < p^{(m)}$, but there are also $\binom{m+1}{1}$ pure strategy NEs with m nodes active with probability 1 as in the case $\mu = m$. Similarly, for $\mu = m + 2$ we have a symmetric mixed strategy NE that uses all $m + 2$ nodes active with probability $p^{(m+2)} < p^{(m+1)} < p^{(m)}$, $\binom{m+2}{1}$ mixed strategies NEs with $m + 1$ active nodes with probability $p^{(m+1)}$ and $\binom{m+2}{2}$ pure strategy NEs with m active nodes. By iterating the same reasoning we get the statement of the theorem for which there is a symmetric NE in mixed strategies with $p^{(\mu)} < p^{(m)}$, and there are $\binom{\mu}{\zeta}$ pure and mixed strategies NEs, with $\zeta = 1, \dots, \kappa$ and $\kappa = \mu - m$, where $N - \mu + \zeta$ nodes are silent.

APPENDIX IV

Proof of Theorem 3 The proof of the existence of multiple equilibria closely follows the same reasoning of Theorem 2, with the major difference being the threshold $m^* \geq m$. First, it is trivial to check the existence of the catastrophic equilibrium in which no node transmits. This emerges from the fact that there is no incentive for any node to deviate from the *not transmit* strategy when all the others are staying silent, as the task will inevitably fail. To prove the existence of this modified threshold m^* let us consider again the ratio function $R(p, \mu)$ for $\mu \geq m$, which now has to use the less handy version of the derivative of the success probability (14) and the results of Theorem 1 for the symmetry of the solution. With this we get that for $p \rightarrow 0$, $R(p, \mu) \sim \frac{1}{p^{m+1}} \rightarrow +\infty$. For $p = 1$, instead $R(1, \mu)$ approaches a positive value which depends on the specific shape of (2). Given the continuity of $R(p, \mu)$ in p , there must exist a global minimum $R_{\min}(\mu) \in (0, 1]$. Because increasing μ guarantees that more nodes are active and the probability of task success increases, $R_{\min}(\mu)$ is a decreasing function in μ . Therefore, there must exist $m^* = \min_{\mu} [c \geq R_{\min}(\mu)]$. In summary, if $\mu < m^*$ we have $c < R_{\min}(\mu)$ and the only stable NE is the one in which all nodes remain silent. If $\mu \geq m^*$, then $c \geq R_{\min}(\mu)$. By the continuity and monotonicity of $R(p, \mu)$, by the intermediate value theorem there exists a solution $p^{(\mu)} \in (0, 1]$ for which (20) is satisfied.

APPENDIX V

Proof of Theorem 4 Consider the HT. A strategy is Pareto efficient if no other strategy can decrease a player's penalty without increasing another's. This is equivalent to finding the NE that minimizes the social penalty defined as $\sum_{i=1}^N \mathcal{P}_i$. This is minimized when $P_{\text{succ}} = 1$ and thus the only component to be minimized is $\sum_{i=1}^N \mathcal{C}_i$. By Theorem 2, for any given m , there exists a set of pure strategy NEs where exactly m nodes transmit with probability $p = 1$. In these circumstances, the cost component of the social penalty becomes $\sum_{i=1}^m c_i$, and attains a minimum at $m = 1$. This is therefore the most Pareto efficient solution to the problem, as any other configuration, also in mixed strategies, would require a less cost-effective activation pattern as $\mu > m$ nodes will be required to actively transmit.

REFERENCES

- [1] B. Zhou and W. Saad, "Performance analysis of age of information in ultra-dense Internet of things (IoT) systems with noisy channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3493–3507, May 2022.
- [2] A. Bujari, A. Calvio, L. Foschini, A. Sabbioni, and A. Corradi, "A digital twin decision support system for the urban facility management process," *Sensors*, vol. 21, no. 24, 2021.
- [3] J. A. Rahal, G. de Veciana, T. Shimizu, and H. Lu, "Optimizing timely coverage in communication-constrained collaborative sensing systems," *IEEE Trans. Control Netw. Syst.*, vol. 11, no. 3, pp. 1717–1729, 2024.
- [4] Z. Song, J. Cheng, A. Chauhan, and E. Tilevich, "Pushing participatory sensing further to the edge," in *Proc. IEEE EDGE*, 2019, pp. 24–26.
- [5] P. Bellavista, L. Foschini, and A. Mora, "Decentralised Learning in Federated Deployment Environments: A System-Level Survey," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–38, 2021.
- [6] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.
- [7] M. Ma, V. W. Wong, and R. Schober, "Channel-aware joint AoI and diversity optimization for client scheduling in federated learning with non-iid datasets," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 6, pp. 6295–6311, 2024.
- [8] K. Saurav and R. Vaze, "Game of ages," in *Proc. IEEE Infocom Wkshps*, 2020, pp. 942–947.
- [9] L. Badia, A. Zanella, and M. Zorzi, "A game of ages for slotted ALOHA with capture," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4878–4889, May 2024.
- [10] T. AlSkaif, M. G. Zapata, and B. Bellalta, "Game theory for energy efficiency in wireless sensor networks: Latest trends," *J. Netw. Comput. Appl.*, vol. 54, pp. 33–61, 2015.
- [11] Q. He, G. Dán, and V. Fodor, "Joint assignment and scheduling for minimizing age of correlated information," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 1887–1900, 2019.
- [12] A. Buratto, A. Mora, A. Bujari, and L. Badia, "Game Theoretic Analysis of AoI Efficiency for Participatory and Federated Data Ecosystems," in *IEEE ICC Wkshps*, 2023, pp. 1301–1306.
- [13] Y. Li, F. Li, S. Yang, Y. Wu, H. Chen, K. Sharif, and Y. Wang, "MP-cooperation: Competitive and cooperative mechanism for multiple platforms in mobile crowd sensing," *IEEE Trans. Serv. Comput.*, vol. 14, no. 6, pp. 1864–1876, 2021.
- [14] A. Garnaev, W. Zhang, J. Zhong, and R. D. Yates, "Maintaining information freshness under jamming," in *Proc. IEEE Infocom Wkshps*, 2019.
- [15] Z. Lian and C. Su, "Decentralized federated learning for Internet of things anomaly detection," in *Proc. ACM Asia Conf. Comp. Commun. Sec.*, 2022, pp. 1249–1251.
- [16] G. Yang, J. Sang, H. Li, X. He, F. Sun, J. Wang, and H. Pervaiz, "Efficient group collaboration for sensing time redundancy optimization in mobile crowdsensing," *IEEE Internet Things J.*, vol. 11, no. 15, pp. 26091–26103, 2024.
- [17] X. Zhang, Z. Cai, X. Yang, J. Luo, R. Zhang, and Y. Zhang, "Game theory-based demand-side management for efficient energy collaboration in smart networks: a neighborhood-scale optimization framework," *Electr. Eng.*, vol. 107, no. 7, pp. 9071–9088, 2025.
- [18] L. He, G. Sun, D. Niyato, H. Du, F. Mei, J. Kang, M. Debbah, and Z. Han, "Generative AI for game theory-based mobile networking," *IEEE Wirel. Commun.*, vol. 32, no. 1, pp. 122–130, 2025.
- [19] K. Akkarajitsakul, E. Hossain, D. Niyato, and D. I. Kim, "Game theoretic approaches for multiple access in wireless networks: A survey," *IEEE Commun. Surv. Tutor.*, vol. 13, no. 3, pp. 372–395, 2011.
- [20] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE ISIT*, 2012, pp. 2666–2670.
- [21] S. Banerjee and S. Ulukus, "Age of information in the presence of an adversary," in *Proc. IEEE Infocom Wkshps*, 2022.
- [22] G. D. Nguyen, S. Kompella, C. Kam, and J. E. Wieselthier, "Information freshness over a Markov channel: The effect of channel state information," *Ad Hoc Netw.*, vol. 86, pp. 63–71, Apr. 2019.
- [23] K. Saurav and R. Vaze, "Game of ages in a distributed network," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1240–1249, May 2021.
- [24] V. S. Dasari, B. Kantarci, M. Pouryazdan, L. Foschini, and M. Girolami, "Game theory in mobile crowdsensing: A comprehensive survey," *Sensors*, vol. 20, no. 7, p. 2055, 2020.
- [25] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," in *Proc. ACM SIGCOMM*, 2009, pp. 280–293.
- [26] L. Giaretta, I. Savvidis, T. Marchioro, Š. Girdzjauskas, G. Pallis, M. D. Dikaiakos, and E. Markatos, "PDS 2: A user-centered decentralized marketplace for privacy preserving data processing," in *Proc. IEEE ICDEW*, 2021, pp. 92–99.
- [27] K. Yang, K. Zhang, J. Ren, and X. Shen, "Security and privacy in mobile crowdsourcing networks: challenges and opportunities," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 75–81, 2015.
- [28] Y. Lai, H. Liu, T.-T. Chan, H. Pan, and C. Jiang, "Age of collection with network-coded multiple access: An experimental study," in *Proc. IEEE VTC2024-Spring*, 2024, pp. 1–6.
- [29] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, May 2020.
- [30] R. D. Yates, "The age of gossip in networks," in *Proc. IEEE ISIT*, 2021, pp. 2984–2989.
- [31] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *IEEE ISIT*, 2017, pp. 1122–1126.

- [32] D. Zorbas, C. Caillouet, K. Abdelfadeel Hassan, and D. Pesch, “Optimal data collection time in lora networks—a time-slotted approach,” *Sensors*, vol. 21, no. 4, p. 1193, 2021.
- [33] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, “Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning,” *IEEE Access*, vol. 10, pp. 40 281–40 306, 2022.
- [34] M. Fernandez and S. Williams, “Closed-form expression for the poisson-binomial probability density function,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 2, pp. 803–817, 2010.
- [35] D. Paccagnan, R. Chandan, and J. R. Marden, “Utility design for distributed resource allocation—part I: Characterizing and optimizing the exact price of anarchy,” *IEEE Trans. on Automat. Contr.*, vol. 65, no. 11, pp. 4616–4631, 2019.
- [36] G. Hardin, “The tragedy of the commons,” *J. Nat. Resource Pol. Res.*, vol. 1, no. 3, pp. 243–253, 2009.



Leonardo Badia (Senior Member, IEEE) received the Ph.D. degree in Information Engineering from the University of Ferrara, Italy, in 2004. From 2002 to 2003, he was a Visiting Scholar at the Radio System Technology Labs (currently Wireless@KTH), Royal Institute of Technology, Stockholm, Sweden. In 2006, he joined the IMT Institute for Advanced Studies, Lucca, Italy. In 2011, he moved to the University of Padova, Italy, where he is currently a Full Professor. He has published more than 300 research papers. His research interests include cross-layer optimization in wireless communications, transmission protocol modeling, and the application of game theory to radio resource management. He has also served on the editorial boards of several scientific journals and remains an active reviewer for many international periodicals, as well as a technical program committee member for conferences in the broad areas of communications and networking.



Alessandro Buratto (Graduate Student Member, IEEE) is a PhD at the University of Padova, Italy. He received the Bachelor's degree in Information Engineering and the Master's degree in Telecommunications Engineering from the same University in 2020 and 2022 respectively. In 2025 he was visiting research scholar at University of California Irvine, California, and visiting researcher at German Aerospace Center, Germany. His research interests are in distributed optimization for participatory sensing scenarios, age of information, and applications of game theory to communications scenarios.



Alessio Mora (Member, IEEE) received his Ph.D. in Computer Science and Engineering in 2023 from the University of Bologna, Bologna, Italy. He is currently an assistant professor at the University of Bologna. His research interests include Decentralized Learning, with a particular focus on Federated Learning, Deep Learning, and Edge Intelligence.



Armir Bujari (Member, IEEE) is an Associate Professor at the University of Bologna, Italy, in the Department of Computer Science and Engineering. His research interests include edge/fog computing, next-generation networks, with a focus on Industrial IoT, and mobile multimedia systems and services. On these topics, he has participated in numerous national and EU-funded research projects, often serving in key coordination and leadership roles.