IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 March 2025, accepted 11 April 2025, date of publication 17 April 2025, date of current version 28 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3561790

RESEARCH ARTICLE

Symbiotic Positioning, Navigation, and Timing via Game Theory and Reinforcement Learning

MD SADMAN SIRAJ[1], (Member, IEEE),
EIRINI ELENI TSIROPOULOU[1], (Senior Member, IEEE),
SYMEON PAPAVASSILIOU[2], (Senior Member, IEEE),
AND JIM PLUSQUELLIC[3], (Senior Member, IEEE)

[1]Performance and Resource Optimization in Networks–PROTON Laboratory, School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85282, USA
[2]School of Electrical and Computer Engineering, National Technical University of Athens, 157 80 Athens, Greece
[3]Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87106, USA

Corresponding author: Eirini Eleni Tsiropoulou (eirini@asu.edu)

The work of Eirini Eleni Tsiropoulou was supported in part by NSF under Award #2219617 and Award #2319994.

ABSTRACT Precise positioning, navigation, and timing (PNT) capabilities are essential for numerous critical infrastructure systems and advanced location-dependent applications. The challenges related to the reliability and accuracy of traditional Global Navigation Satellite Systems (GNSS) have driven the pursuit of innovative alternative PNT methodologies. This paper presents a novel approach inspired by the concept of biological symbiosis and leveraging the advanced capabilities of Reconfigurable Intelligent Surfaces (RISs). The proposed framework establishes a cooperative interaction between targets with unknown positions and collaborator nodes with approximate location estimates. These interactions are supported by the RISs and the anchor nodes with known positions. The objective is to minimize errors in positioning and timing for both the targets and the collaborators. This challenge is modeled as a non-cooperative game, and the existence of a Nash Equilibrium is demonstrated using potential game theory. To solve the game, Best Response Dynamics and a log-linear Reinforcement Learning (RL)-based approach are developed to identify the equilibrium state. The proposed system is thoroughly evaluated through simulations, in order to demonstrate its performance and the key trade-offs between game-theoretic strategies and the RL-based solutions.

INDEX TERMS Game theory, reconfigurable intelligent surfaces, reinforcement learning, symbiotic positioning, navigation, timing (SPNT).

I. INTRODUCTION

Positioning, Navigation, and Timing (PNT) services are becoming more and more pivotal in many modern applications, such as healthcare monitoring, disaster management, autonomous driving, augmented and virtual reality, just to name few. The legacy Global Navigation Satellite System (GNSS) is the dominant satellite-based PNT system providing mainly outdoor PNT services, with the Global Positioning System (GPS) being primarily used in the United States. However, the GNSS availability becomes rather limited in urban canyons, hostile territory, and indoor environments. Also, the GNSS system's quality of information may degrade due to man-made or physical interference in the satellite signals, which suffer from long propagation distances, or spoofing and jamming [1]. Thus, the design of alternative PNT solutions, which can complement or even substitute the GNSS, in cases of GNSS deteriorated services or denial is of paramount importance [2]. In this paper, aligned with the latter vision, a novel bio-inspired PNT solution is introduced. Specifically, motivated by the concept of symbiosis in biological systems and by exploiting the key enabling technology of Reconfigurable Intelligent Surfaces (RISs), we introduce a symbiotic PNT solution, where the targets

The associate editor coordinating the review of this manuscript and approving it for publication was Jonathan Rodriguez.

collaborate among each other in a symbiotic manner in order to accurately determine their position and timing. Following the symbiosis paradigm, the relationship of the targets with the rest of the nodes in the systems, i.e., RISs, collaborator nodes, and anchor nodes, is studied to design the novel symbiotic PNT solution, going beyond the simple exchange of pseudoranges measurements. The RISs technology is exploited by the targets to improve their PNT accuracy. The targets' distributed decision-making within the proposed symbiotic PNT system is performed based on Game Theory and Reinforcement Learning, while exploiting the benefits of each one of those approaches.

A. RELATED WORK

In the recent years, several alternative PNT solutions have been designed for indoor and outdoor environments, such as vision-based methods, fingerprinting, inter-vehicle collaboration, by mainly exploiting Bluetooth, ZigBee, LoRa (Long Range), WiFi, and RFID (Radio Frequency Identification Device) technologies [3]. An automatic algorithm for constructing environmental fingerprints in multi-storey buildings is introduced in [4] using unlabelled crowdsourced smartphone data, multimodal sensor fusion, and deep neural networks, in order to eliminate the need for manual fingerprint collection. An RFID-based localization solution is proposed in [5] for a vehicles use case scenario. The vehicles are equipped with an RFID reader, receiving signals from the RFID tags installed on the road sides. A multi-anchor nodes approach is studied in [6], by introducing an RSSI-based least-squares multilateration method that exploits the measured pseudoranges from an increasing number of anchor nodes. The main drawback of all the aforementioned PNT solutions is the increased infrastructure cost introduced in order to deploy dedicated equipment for supporting the PNT services.

The interest of the research community focused in particular on developing PNT solutions for vehicles-focused applications [7]. A reinforcement learning-based model is proposed in [8] by developing an asynchronous advantage actor-critic algorithm to enable the vehicles to learn their optimal strategies in order to make corrections on raw GNSS observations and ultimately improve the accuracy of their PNT services. In [9], a fusion framework based on the sparse Gaussian-Wigner prediction method is developed by utilizing the random matrix theory and the sparse property in order to improve the vehicles' PNT services. A clustering-based cooperative relative positioning scheme for UAV swarms in GPS-denied environments is developed in [10] by leveraging a coalition formation game model to balance intracluster cooperation and intercluster packet loss. Complementary, a sparse Bayesian learning-based model is proposed to improve the DoA estimation and achieve higher precision of the PNT services. The main drawback of the above discussed PNT solutions, focused on vehicles-oriented use case scenarios, is their high computational complexity resulting in long execution times of the proposed models to determine the targets' positions, and ultimately high energy consumption in order to be executed. Thus, those PNT solutions cannot be easily applied to Internet of Things (IoT) devices, which are characterized by limited computing resources and battery.

The study of cooperation among the targets, in order to improve their experienced PNT services, is still in its infancy. In [11], a multi-agent collaborative localization algorithm id designed using reinforcement learning compensation filtering to enhance the localization accuracy and robustness in complex environments. A cooperative localization framework among Uncrewed Aerial Vehicles (UAVs) is proposed in [12] that determines the minimum number of distance measurements that are required in a swarm of UAVs in order to accurately determine their relative positions. A similar framework is developed in [13] to introduce a cooperative vehicle localization and trajectory prediction framework, utilizing a belief propagation based location approximation algorithm for vehicle localization and a transformer-based model for trajectory prediction. A different approach is discussed in [14], by allowing the targets to exchange velocity measurements and information stemming from their inertial measurement units. The overall set of information is fed into a multi-hypothesis extended Kalman filter that determines the relative position and orientation of the targets. A different type of cooperation is discussed in [15] by introducing a dual-system localization approach that combines the GPS and the BeiDou navigation satellite system (BDS) in order to provide PNT services to the targets.

Recently, novel next generation wireless networks technologies, such as the Reconfigurable Intelligent Surfaces (RISs) and Device-to-Device (D2D) communications, have been exploited to design alternative PNT solutions. A comprehensive review and classification of deep learning-based visual localization approaches for UAV navigation in GPS-denied environments is presented in [16] analyzing their advantages, challenges, and future research directions. An UAV-based PNT framework is proposed in [17] that considers Integrated Sensing and Communication (ISAC) technologies, reinforcement learning, and game theory to enhance the victims' positioning and emergency response efficiency in post-disaster scenarios. A novel deep learning-based approach for vehicle indoor positioning using smartphone built-in sensors is introduced in [18], which outperforms existing methods and offers a cost-effective and accurate solution for smart car parking and driverless cars. A secure ground-based PNT solution for GPS-denied environments is proposed in [19] based on the matching theory and coalition games to optimize the anchor node selection in search and rescue and military operations. The RIS technology is exploited in [20] to decrease the number of necessary anchor nodes in order to determine the targets' positions, by exploiting the reflected signals on the RISs. A similar approach is followed in [21], where the authors also study the impact of the near-field and far-field propagation conditions on the

accuracy of the PNT solution. The MERCURY mechanism is introduced in [22], which leverages reinforcement learning, game theory, and RIS technology to optimize alternative PNT solutions, minimizing positioning errors and enhancing system reliability. Furthermore, the impact of the RISs' reflection coefficients on the accuracy of the PNT services is analyzed in [23], following a very similar PNT solution as in [20] and [21].

B. CONTRIBUTIONS AND OUTLINE

The existing research and literature has also identified the challenges and drawbacks of the implemented and existing alternative PNT solutions, mainly in terms of high infrastructure cost and computationally expensive solutions. Aiming at filling this exact gap, the main contribution and originality of this work lie in the introduction of a novel symbiotic PNT framework that leverages game theory and reinforcement learning (RL) to address the limitations of traditional PNT systems, particularly in GPS-denied or degraded environments. The framework establishes a mutualistic relationship among the targets, collaborator nodes, anchor nodes, and RISs, and enables the nodes with unknown or approximate positions to collaboratively improve their positioning and timing accuracy. This approach is bio-inspired, drawing from the concept of symbiosis in biological systems, and extends beyond traditional pseudorange-based methods by incorporating advanced technologies like RISs and distributed decision-making. The formulation of the problem as a non-cooperative potential game ensures the existence of a Nash Equilibrium, while the introduction of Best Response Dynamics (ABRD and SBRD) and RL-based algorithms (Binary-Logit and Max-Logit) provides flexible solutions for different scenarios, and balances the accuracy and computational efficiency. This framework is particularly important for applications in urban canyons, indoor environments, and hostile territories where GPS signals are unreliable or unavailable, and offers a scalable and robust alternative to traditional PNT systems.

Also, this current paper expands upon our previous works, introducing several key advancements to alternative PNT solutions. Specifically, compared to [27], it explores RL-based algorithms alongside game-theoretic approaches, demonstrating that RL outperforms the latter work. It also introduces comparisons with state-of-the-art PNT solutions using Iterative Least Squares (ILS) algorithms compared to [28]. In contrast to the [28], which focused on UAV-based localization for emergency scenarios, the current paper develops a generalized PNT framework applicable to various situations, particularly GPS-denial cases. Finally, while [29] concentrated on optimizing RIS elements allocation for localization, the current paper introduces a more robust and dynamic PNT framework that incorporates neighborhood determination and enables mutual localization through symbiotic relationships among participating nodes.

The novel key contributions of our research work are summarized below.

1) **Novel symbiotic PNT framework**: We introduce a unique symbiotic environment that integrates targets, RISs, collaborator nodes, and anchor nodes. This framework establishes a mutualistic relationship that enables the targets with unknown coordinates and the collaborator nodes with approximate positions to collaboratively improve their positioning and timing accuracy.
2) **Autonomous neighborhood identification and RIS optimization**: We propose a new method for the targets and the collaborator nodes to autonomously identify nearby nodes and optimize the RIS phase shifts. This approach maximizes the signal strength and improves the pseudorange measurements for the targets in order to ultimately enhance their positioning precision.
3) **Game-theoretic and reinforcement learning-based error minimization**: We formulate the positioning and timing problem as a potential game, proving the existence of a Nash Equilibrium. We introduce two game-theoretic (Asynchronous and Synchronous Best Response Dynamics) and two reinforcement learning (Binary-Logit and Max-Logit) algorithms to balance exploration and exploitation in achieving optimal positioning outcomes.
4) **Comprehensive simulation and comparative analysis**: We conduct extensive simulations to evaluate the performance of the game-theoretic and the RL-based approaches, focusing on the execution time, estimation error, and scalability. Additionally, we compare our symbiotic PNT approach with traditional ground-based solutions in order to demonstrate its advantages in large-scale environments.

Our discussion begins with the symbiotic environment and provides an overview of the proposed symbiotic PNT model (Section II). Section III presents the targets' and collaborator nodes' neighborhood identification process, along with the RISs' phase shift optimization. Section IV discusses the formulation of the symbiotic PNT problem as a potential game and shows the existence of a Nash Equlibrium, which is determined following a game-theoretic and a reinforcement learning-based approach. Section V presents a thorough analysis of our model's performance, supported by numerical data, while the key innovations of the proposed framework are summarized in Section VI. We conclude our paper in Section VII with a summary of key findings and potential future directions for research in this domain.

II. OVERVIEW OF THE PROPOSED FRAMEWORK

A. SYSTEM MODEL

A novel symbiotic environment is considered, consisting of the set of anchor nodes $A = \{1, \ldots, a, \ldots, |A|\}$, RISs $R = \{1, \ldots, r, \ldots, |R|\}$, collaborator nodes $C = \{1, \ldots, c, \ldots, |C|\}$, and targets $U = \{1, \ldots, u, \ldots, |U|\}$. The anchor nodes, e.g., gNBs, and the RISs are deployed

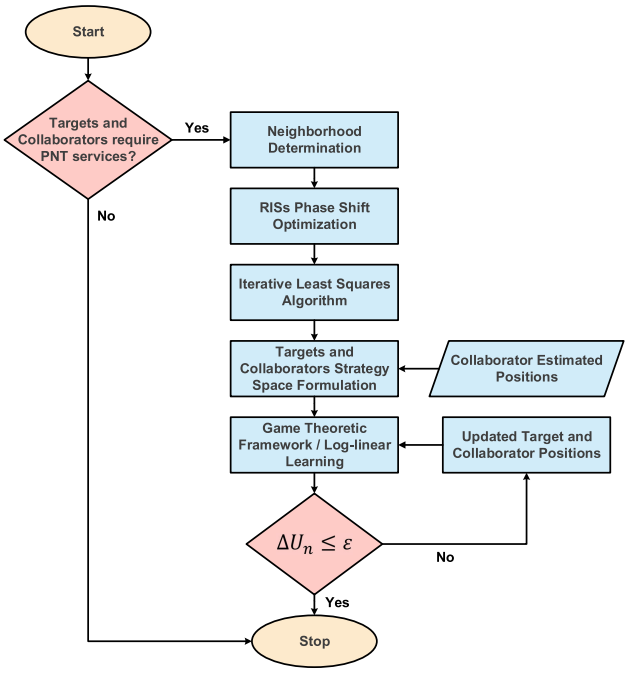

FIGURE 1. Overview of the symbiotic PNT solution's operation.

by the network service providers and their coordinates are known, i.e., $\mathbf{x}_a = (x_a, y_a, z_a), \forall a \in A$, and $\mathbf{x}_r = (x_r, y_r, z_r), \forall r \in R$, respectively. The collaborator nodes have an initial rough estimate of their position $\hat{\mathbf{x}}_c = (\hat{x}_c, \hat{y}_c, \hat{z}_c), \forall c \in C$, and the targets have unknown coordinates $\mathbf{x}_u = (x_u, y_u, z_u), \forall u \in U$. Also, in the general case, the clocks of the anchor nodes, and the collaborator nodes and targets may not be fully synchronized, thus, there is a corresponding clock offset Δt_c and Δt_u [sec], respectively.

The four types of entities engage in a mutualistic symbiotic relationship among each other, which is founded on the service-to-service mutualism basis. Specifically, all the four types of entities mutually collaborate among each other in order to minimize the overall positioning and timing error in the examined system, and no entity can achieve this goal by unilaterally making decisions in an isolated manner. The symbiotic relationship among all the involved entities benefits the collaborator nodes and the targets, in terms of accurately determining their positioning and timing, as well as the network service providers (who own the anchor nodes and RISs) to provide accurate services to their customers.

B. OPERATION OF THE PROPOSED FRAMEWORK

In this section, we provide an overview of the proposed symbiotic PNT solution, by highlighting its modules, and the flow of control and information among them, as presented in Fig. 1. Initially, each target and collaborator node, identifies its neighborhood, i.e., within communication range, via periodically transmitting beacon signals. It is noted that in the rest of the analysis, we refer to the targets and collaborators as ''nodes'', given that they are the ones that aim at accurately determining their positioning and timing. Then, each node will create a mutualistic symbiotic relationship with the anchor nodes, RISs, and other collaborator nodes belonging to its neighborhood in order to determine its positioning and timing by measuring the corresponding pesudoranges from them through exploiting their broadcasted beacon signals. Additionally, each node gets an equal share of each RIS's elements belonging to its neighborhood, and optimizes their phase shifts in order for the received signal strength to be maximized and the node to more accurately measure the corresponding pseudoranges. Then, a non-cooperative game is formulated among the nodes as a potential game in order to minimize their positioning and timing error. An Asynchronous and a Synchronous Best Response Dynamics algorithm are introduced in order to determine the Nash Equilibrium following the local search process, where all the nodes have determined their optimal positioning and timing estimation that minimizes their personal experienced error, as well as the system's overall estimation error. Alternatively, two log-linear RL algorithms are introduced, i.e., Binary Logit and Max Logit, to determine the nodes' optimal positioning and timing estimation, following the exploration ad learning processes. The overall process is repeated iteratively in the case of mobile nodes, where the topology dynamically changes. In the following sections, we provide a detailed analysis of the proposed symbiotic PNT solution's modules.

III. NEIGHBORHOOD IDENTIFICATION AND RIS PHASE SHIFT OPTIMIZATION

Initially, each node uses the beacon signals stemming from the anchor nodes, RISs, and collaborator nodes residing in its neighborhood to determine an initial estimate of its positioning and timing $\hat{\mathbf{P}}_j = (\hat{x}_j, \hat{y}_j, \hat{z}_j, \Delta\hat{t}_j), \forall j \in U \cup C$ based on the multilateration technique [24]. Each node emits a localization signal at a constant transmission power P [W], which is detected by its associated anchor nodes $A_j \subset A$ and collaborators $C_j \subset C$. Then, the later two types of nodes send out a response beacon signal with a constant transmission power P, which carries data related to their position and time. This data is represented as $\mathbf{P}_a = (x_a, y_a, z_a, \Delta t_a)$, $\forall a \in A_j$, and $\hat{\mathbf{P}}_c = (\hat{x}_c, \hat{y}_c, \hat{z}_c, \Delta\hat{t}_c), \forall c \in C_j$. Additionally, the anchor nodes broadcast digital information related to the locations of the RISs $\mathbf{x}_r$, for all $r \in R_j \subset R$ that lie within the coverage area of the target. This data is included in their transmitted signal. As a result, every node becomes aware of the coordinates $\mathbf{x}_a$, $\hat{\mathbf{x}}_c$, and $\mathbf{x}_r$, and can estimate the pseudoranges to adjacent nodes by analyzing the signal power strength received, as given by: $P^{rec}_{j,i} = P\frac{G^{trans}_i G^{rec}_j}{L_{j,i}}$, where $i = \{a, c, r\} \in A_j \cup C_j \cup R_j$. G^{trans}_i [dB] is the gain of the transmitting antennas, G^{rec}_j [dB] is the node's receiving antenna's gain, and P [W] is the fixed transmission power of the broadcasted beacon signal. The assumption of a fixed transmission power can ensure accuracy and consistency of the multilateration-based positioning process.

Specifically, each node derives its pseudorange estimates by analyzing the received signal strength from the anchor nodes, RISs, and collaborating nodes. It should be noted that practical implementations of localization systems, such as those based on beacon signals in urban environments, often use fixed transmission power to simplify the received signal strength-based distance estimation process. Many existing localization frameworks deploy beacons with predefined and stable power levels to mitigate power control-induced uncertainties. Also, it is highlighted that adaptive power control could have been introduced. However, the adaptive power control would require additional complexity in decoding the transmitted power levels at the receiver, which could introduce errors in the distance estimation and degrade localization accuracy [31]. In this paper, the main focus is the design of an alternative accurate PNT solution, where a fixed transmission power can support the measurement of the pseudorange by the target in a low complexity manner.

The path loss for the transmission is calculated using the Okumura/Hata model tailored for dense urban environments: $L_{j,i}(d_{j,i}) = 69.55 + 26.16\log f_c + (44.9 - 6.55\log h_i^{trans})\log d_{j,i} - 13.82\log h_i^{trans} - 3.2[log(11.75h_j)]^2 - 4.97$ [dB], where $f_c \geq 400$ [Hz] represents the carrier frequency, h_i^{trans} [m] denotes the height of the transmitting nodes $i \in A_j \cup C_j$, h_j [m] is the node's antenna's height, and $d_{j,i}$ [m] is the pseudorange measured among the node $j \in U \cup C$ and the node $i \in A_j \cup C_j \cup R_j$. The Okumura/Hata model has been extensively validated for urban environments and provides a well-established empirical formulation that captures key propagation characteristics, such as path loss variation with frequency, transmitter height, and distance. Moreover, in our proposed framework, each node autonomously determines its transmission power and association, and a computationally lightweight path loss model is needed. Advanced path loss models, like 3GPP TR 38.901, could have also been adopted as they are highly detailed and suitable for advanced cellular deployments, however, they introduce additional complexity due to their multi-environment applicability and reliance on extensive parameterization. It should be noted that the provided analysis in the paper would not be affected by the selection of the path loss model. By following this neighborhood identification process, each node j is informed about the coordinates $\mathbf{x}_a, \hat{\mathbf{x}}_c, \mathbf{x}_r$ and the pseudoranges $d_{j,a}, d_{j,c}, d_{j,r}, \forall a \in A_j, \forall c \in C_j, \forall r \in R_j$, and can implement the multilateration technique to determine an initial rough estimate of $\hat{\mathbf{P}}_j, \forall j \in U \cup C$.

Each node depends on the signal strength of the received beacon signals in order to measure the pseudoranges from the reference points $i \in A_j \cup C_j \cup R_j$ and ultimately estimate its positioning and timing $\hat{\mathbf{P}}_j$. The optimal control of the RISs elements' phase shifts can contribute in the software-defined design of a constructive beam with desirable propagation characteristics and maximized received signal strength at the node's side. Given that the anchor nodes are the only transmitting nodes with perfect knowledge of their position $\mathbf{x}_a$, each node selects the anchor node a^* from its neighborhood with the strongest incoming signal in order to optimize the RISs elements' phase shifts and further maximize the received signal's strength. Thus, each node can more accurately measure the pseudoranges and ultimately estimate more accurately its positioning and timing $\hat{\mathbf{P}}_j$.

Each RIS r consists of a set of elements $\mathcal{M}_r = \{1, \ldots, m, \ldots, M_r\}$. The RIS elements correspond to unit cells of a programmable metasurface, which enable the phase control and facilitate the signal reflections to improve the performance of the communication system. A RIS r may reside in the same neighborhood of multiple nodes, thus, following the principles of proportional fairness, its elements are equally shared among the nodes, i.e., $\frac{M_r}{J_r} = M_r^j$, where J_r denotes the corresponding number of nodes. Also, $\mathcal{M}_r^j = \{1, \ldots, m, \ldots, M_r^j\}$ denotes the set of elements of a RIS r that are allocated to node j in order to contribute to its PNT service. Given the multipath propagation of the anchor node's a^* transmitted signal due to the reflection on the RIS, there are three channel gain coefficients that should be defined among: (i) the node j and anchor node a^*, i.e., h_{j,a^*}, (ii) the anchor node a^* and RIS r, i.e., $\mathbf{h}_{a^*,r}$, and (iii) the RIS r and node j, i.e., $\mathbf{h}_{r,j}$.

The node's j direct communication link with the anchor node a^* is characterized by the channel gain coefficient $h_{j,a^*} = L_{j,a^*}(d_{j,a^*}) \cdot \tilde{h}$, where d_{j,a^*} [m] is the distance among j and a^*. The random scattering component in the system $\tilde{h} \sim CN(0,1)$ is modeled by a complex Gaussian random variable with zero mean and unit variance. The communication link between the anchor node a^* and the RIS r is defined by the channel gain coefficient $\mathbf{h}_{a^*,r} = \sqrt{\frac{1}{PL_{a^*,r}}}\left[1, e^{-j\frac{2\pi}{\lambda}d_s\phi_{a^*,r}}, \ldots, e^{-j\frac{2\pi}{\lambda}(M-1)d_s\phi_{a^*,r}}\right]^T$, where λ represents the carrier wavelength in meters, d_s is the antenna spacing in meters, and $\phi_{a^*,r}$ is the cosine of the angle between the signal's direction from the anchor node to the RIS. The path loss is denoted as $PL_{a^*,r} = \rho d_{a^*,r}^{a_r}$, where ρ (in dB) is the reference path loss at 1 meter, $d_{a^*,r}$ is the distance from the anchor node to the reference element (i.e., $m = 1$) of the RIS, and a_r is the path loss exponent. Each RIS element has its phase shift $\omega_m \in [0, 2\pi)$, for all $m \in \mathcal{M}_r^j$. The reflection matrix for the RIS elements, controlled in a software-defined manner by node j, is $\Omega = \text{diag}(e^{j\omega_1}, \ldots, e^{j\omega_{M_r^j}}) \in \mathbb{C}^{M_r^j \times M_r^j}$. The communication link between RIS r and node j is governed by the channel gain coefficient $\mathbf{h}_{r,j}$, which is formulated as: $\mathbf{h}_{r,j} = L_{r,j}(d_{r,j})\left(\sqrt{\frac{k}{k+1}}\mathbf{h}_{r,j}^{LoS} + \sqrt{\frac{1}{k+1}}\mathbf{h}_{r,j}^{NLoS}\right)$, where $d_{r,j}$ [m] is the distance among the r and j, k is the Rician factor, $\mathbf{h}_{r,j}^{NLoS} \sim CN(0,1)$ captures the Non-Line-of-Sight (NLoS) component, $\mathbf{h}_{r,j}^{LoS} = [1, e^{-j\frac{2\pi}{\lambda}d_s\phi_{r,j}}, \ldots, e^{-j\frac{2\pi}{\lambda}d_s(M_r^j-1)\phi_{r,j}}]^T$ captures the LoS component, and $\phi_{r,j}$ is the cosine of the angle of departure of the signal from the RIS r to the node j.

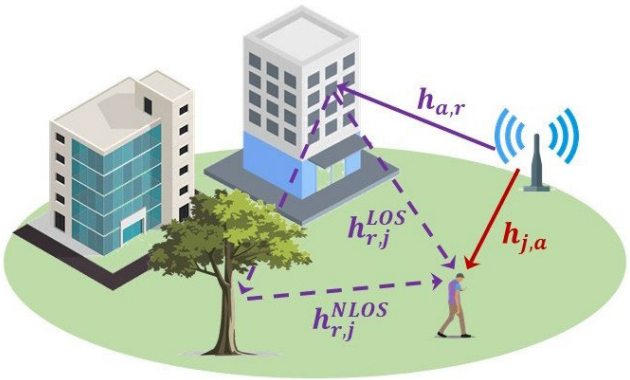

FIGURE 2. Communication links among the anchor nodes, RISs, and targets.

The communication links among the anchor nodes, RISs, and targets are presented in Fig. 2.

The overall channel power gain connecting the anchor node a^* with node j is $G_j^{a^*,r} = |h_{j,a^*} + \mathbf{h}_{a^*,r}\Omega\mathbf{h}_{r,j}|^2$. Node j focuses on adjusting the phase shifts of the RIS elements M_r^j assigned to it, with the goal of enhancing the intensity of the strongest signal arriving from the anchor node a^*.

$$\max_{\omega} |h_{j,a^*} + \mathbf{h}_{a^*,r}\Omega\mathbf{h}_{r,j}|^2 \tag{1a}$$

$$s.t.\ \ 0 \le \omega_m < 2\pi, \forall m \in \mathcal{M}_r^j \tag{1b}$$

where $\omega = [\omega_1, \ldots, \omega_m, \ldots, \omega_{\mathcal{M}_r^j}]$ is the RIS's phase shifts vector. The optimization problem is local to each node j, as it focuses on maximizing the signal strength for that specific node by adjusting the RIS phase shifts. However, the RIS elements are shared among multiple nodes, which introduces a coupling relationship. The phase shifts optimized for one node can affect the signal strength received by other nodes sharing the same RIS. This coupling is managed by the proportional fairness principle, where the RIS elements are equally shared among nodes in the same neighborhood. Towards solving the optimization problem (1a)-(1b), we set $v_m = e^{j\omega_m}, \forall m \in \mathcal{M}_r^j$, and we have $\mathbf{v} = [v_1, \ldots, v_m, \ldots, v_{M_r^j}] \in \mathbb{C}^{M_r^j \times 1}$. Then by substituting $\tilde{\mathbf{h}}_{a^*,r}^H = \mathbf{h}_{r,j}\,\mathrm{diag}(\mathbf{h}_{a^*,r}) \in \mathbb{C}^{1\times M_r^j}$ in (1a), the optimization problem (1a)-(1b) can be rewritten as follows:

$$\max_{\mathbf{v}} |h_{j,a^*} + \tilde{\mathbf{h}}_{a^*,r}\mathbf{v}|^2 \tag{2a}$$

$$s.t.\ \ |v_m| = 1, \forall m \in \mathcal{M}_r^j \tag{2b}$$

The formulation in Eq. 2a reaches its peak value when the direct path between the anchor node a^* and node j, along with the reflected path via the RIS, are perfectly phase-aligned and coherently superimposed. Thus, the following condition holds: $\angle h_{j,a^*} = -\angle\tilde{\mathbf{h}}_{a^*,r} + \angle\mathbf{v} \Leftrightarrow \mathbf{w}^* = \angle\mathbf{v} = \angle h_{j,a^*} + \angle\tilde{\mathbf{h}}_{a^*,r}$. Each node then calculates the optimal phase configuration ω^* for the RIS elements assigned to it within its vicinity. This information, including the corresponding RIS identifier r, is subsequently shared so the RIS controller can adjust the reflective phase settings accordingly. Following this step, the anchor nodes transmit a secondary ranging reply beacon signal at the same fixed power level P. Thus, node $j \in U \cup C$ can achieve more precise pseudorange measurements from the anchor a^* and the RISs $r, \forall r \in R_j$ by leveraging the increased signal strength. This process refines the initial estimates of node j's position and timing $\hat{\mathbf{P}}_j$.

IV. SYMBIOTIC POSITIONING, NAVIGATION, AND TIMING

In this section, we introduce the symbiotic positioning, navigation, and timing (SPNT) model that simultaneously achieves the following goals: (i) accurately determines each node's $j \in U \cup C$ positioning and timing $\hat{\mathbf{P}}_j^*$, and (ii) jointly minimizes the overall estimation error of the nodes, i.e., collaborators and targets, in a symbiotic manner. Initially, we define the Euclidean distance among the involved entities in the SPNT model, as follows:

$$\hat{d}(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i) = \begin{cases} ||\hat{\mathbf{P}}_j - \mathbf{P}_i||, \ \text{if} \ \ i = \{a, r\}, \forall a \in A_j, \forall r \in R_j \\ ||\hat{\mathbf{P}}_j - \hat{\mathbf{P}}_i||, \ \text{if} \ \ i = c, \forall c \in C_j \end{cases} \tag{3}$$

where $\hat{\mathbf{P}}_j$ is the initial estimate of the node's positioning and timing based on the multilateration technique, as discussed in Section III. The distances $\hat{d}$ [m] is an estimation that each node can derive based on the neighborhood identification process that has already taken place. Similarly, each node j has already measured the pseudoranges $d_{j,i}, \forall i \in A_j \cup C_j \cup R_j$, following the same process. Therefore, the position and estimation error of each node is defined as follows: $\epsilon(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i) = [d_{j,i} - d(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i)]^2$. Obviously, each node aims at accurately determining its positioning and timing, thus, minimizing the experienced estimation error $\epsilon(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i), \forall j \in U \cup C$. Therefore, the corresponding optimization problem for each node can be defined as a distributed minimization problem of its estimation error, as follows:

$$\min_{\hat{\mathbf{P}}_j} \sum_{i \in A_j \cup R_j \cup C_j} \epsilon(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i), \forall j \in U \cup C \tag{4}$$

where $\sum_{i \in A_j \cup R_j \cup C_j} \epsilon(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i)$ is the overall estimation error that node j experiences from its neighboring nodes.

Also, the goal of the overall examined symbiotic environment is to minimize the overall estimation error in the system, and the corresponding optimization problem can be defined as follows:

$$\min_{\hat{\mathbf{P}}_j} E(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i) = \sum_{\forall j \in U \cup C} \sum_{\forall i \in A_j \cup R_j \cup C_j} \epsilon(\hat{\mathbf{P}}_j, \hat{\mathbf{P}}_i) \tag{5}$$

This is a global optimization problem that considers the sum of the estimation errors for all nodes (targets and collaborator nodes) in the system. The coupling between the nodes arises from the fact that the positioning and timing errors of one node depend on the pseudorange measurements from neighboring nodes, i.e., anchor nodes, RISs, and collaborator nodes. Each node's estimation error is influenced by the positions and timing of other nodes,

and concludes to a distributed optimization problem where the decisions of one node affect the outcomes of others. Thus, we are adopting a game-theoretic approach to solve this problem compared to existing distributed optimization approaches [30], which require explicit coordination or communication among the nodes.

The optimization problems (4) and (5) can be solved by formulating a non-cooperative game among the targets and the collaborator nodes in order to determine their positioning and timing. The non-cooperative game is defined as $G = [N, \{S_n\}_{\forall n \in N}, \{U_n\}_{\forall n \in N}]$, where $N = U \cup C$ is the set of players, S_n is their strategy set with strategies $\mathbf{s}_n = (\hat{x}_n, \hat{y}_n, \hat{z}_n, \Delta\hat{t}_n)$, and $U_n(\mathbf{s}_n, \mathbf{s}_{-n}) = \sum_{i \in A_n \cup R_n \cup C_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i)$ is the payoff function. Our goal is to prove the existence of at least one Nash Equilibrium, where all the players, i.e., targets and collaborator nodes, have accurately determined their positioning and timing strategies $\mathbf{s}_n^*$ that minimize their estimation error, as defined in Eq. 4. It is noted that in real-world scenarios, the nodes in a ground-based positioning system may not always have complete trust or a mechanism to fully collaborate in optimizing a joint objective. They rely on their own measurements and optimizations to improve their own accuracy, which naturally fits the framework of a non-cooperative game. Modeling the problem as a cooperative game among the nodes would imply the existence of explicit coordination or shared control over resources, which is not the case in GPS-denied environments, which commonly suffer from a lack of infrastructure, e.g., gNBs. The primary objective of each node (whether a target or collaborator) is to minimize its own estimation error in determining its position and timing. As we define in the optimization problem of each node (Eq. 4), each node aims to minimize the difference between its measured pseudoranges and its estimation error in relation to its neighboring nodes. This optimization is inherently individual, meaning each node is acting in its own interest to improve its accuracy independently of the other nodes' optimization efforts. In the described SPNT model, the nodes do not share a common goal that would lead to a collective optimization (as would be the case in a cooperative game). Instead, they make their decisions independently based on their own strategies. The payoff function for each node (Eq. 4) reflects the local minimization of its error without explicit cooperation with other nodes. Even though the overall system goal (Eq. 5) is to minimize total error, this emerges as the collective outcome of each node's individual optimization efforts rather than through a shared objective or cooperative action plan. Although the environment is termed symbiotic, the nodes are essentially competing for optimal accuracy within their own localized network, which naturally lends itself to non-cooperative game theory. The symbiotic nature of the proposed SPNT solution stems from a deeper form of cooperation where the nodes mutually benefit by minimizing both their individual and the system-wide positioning and timing errors. This symbiosis is of paramount importance, as isolated targets (especially in scenarios of limited or denied GNSS services) would face inaccurate positioning and timing estimations.

Definition 1 (Nash Equilibrium – NE): A collection of strategies $\mathbf{s}^*$ represents a Nash Equilibrium in the game G if, for every participant n, the utility obtained by choosing their strategy $\mathbf{s}_n^*$ while all others adhere to their respective strategies $\mathbf{s}_{-n}^*$ is at least as great as the utility they could achieve by selecting any alternative strategy $\mathbf{s}_n'$.

In the following analysis, we show the existence of at least one Nash Equilibrium for the non-cooperative game G, by using the theory of Potential Games [25].

Definition 2 (Exact Potential Game): A non-cooperative game G is an exact potential game, if $\Phi(\mathbf{s}_n, \mathbf{s}_{-n}) - \Phi(\mathbf{s}_n', \mathbf{s}_{-n}) = U_n(\mathbf{s}_n, \mathbf{s}_{-n}) - U_n(\mathbf{s}_n', \mathbf{s}_{-n}), \forall \mathbf{s}_n' \in S_n, \forall n \in N$ where $\Phi(\mathbf{s}_n, \mathbf{s}_{-n})$ is the potential function.

Theorem 1: The non-cooperative game $G = [N, \{S_n\}_{\forall n \in N}, \{U_n\}_{\forall n \in N}]$ is an exact potential game with potential function $\Phi(\mathbf{s}_n, \mathbf{s}_{-n}) = \frac{E(\mathbf{s}_n, \mathbf{s}_{-n})}{2}$.

Proof: We calculate the difference of the payoff function $U_n(\mathbf{s}_n, \mathbf{s}_{-n})$ for two strategies of player n, $\mathbf{s}_n \neq \mathbf{s}_n'$, given the strategies of the rest of the players $\mathbf{s}_{-n}$, as follows:

$$U_n(\mathbf{s}_n, \mathbf{s}_{-n}) - U_n(\mathbf{s}_n', \mathbf{s}_{-n}) = \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) - \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n', \hat{\mathbf{P}}_i),$$

where $N_n = A_n \cup R_n \cup C_n$. We analyze the potential function:

$$\begin{aligned}
\Phi(\mathbf{s}_n, \mathbf{s}_{-n}) &= \frac{1}{2} \sum_{\forall n \in N} \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) = \frac{1}{2}[\sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) \\
&+ \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\forall i \in N_k} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i)] = \frac{1}{2}[\sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) \\
&+ \sum_{\substack{\forall k \in N \\ k \neq n}} [(\sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i)) + \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n)]] \\
&= \frac{1}{2}[\sum_{i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) + \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i) \\
&+ \sum_{\substack{\forall k \in N \\ k \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n)].
\end{aligned}$$

However, if two players, i.e., nodes k, n are not neighbors, then, they are not able to measure the pseudoranges among them, thus $\epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n) = 0$, for $k, n \notin N_n$. Therefore, the last term of the potential function is analyzed as follows:

$$\begin{aligned}
\sum_{\substack{\forall k \in N \\ k \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n) &= \sum_{\forall k \in N_n} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n) + \underbrace{\sum_{\substack{\forall k \notin N_n \\ k \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n)}_{=0} \\
&= \sum_{\forall k \in N_n} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n).
\end{aligned}$$

Based on this analysis, we can rewrite the potential function:

$$\Phi(\mathbf{s}_n, \mathbf{s}_{-n}) = \frac{1}{2}[\sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) + \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i) + \sum_{\forall k \in N_n} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_n)] = \frac{1}{2}[2 \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) + \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i)] = \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) + \frac{1}{2} \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i).$$

Then, we take the difference of the potential function $\Phi(\mathbf{s}_n, \mathbf{s}_{-n})$ for two strategies $\mathbf{s}_n \neq \mathbf{s}'_n$ of player n, as follows:

$$\Phi(\mathbf{s}_n, \mathbf{s}_{-n}) - \Phi(\mathbf{s}'_n, \mathbf{s}_{-n}) = \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) + \frac{1}{2} \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i) - [\sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}'_n, \hat{\mathbf{P}}_i) + \frac{1}{2} \sum_{\substack{\forall k \in N \\ k \neq n}} \sum_{\substack{\forall i \in N_k \\ i \neq n}} \epsilon(\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_i)] = \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}_n, \hat{\mathbf{P}}_i) - \sum_{\forall i \in N_n} \epsilon(\hat{\mathbf{P}}'_n, \hat{\mathbf{P}}_i) = U_n(\mathbf{s}_n, \mathbf{s}_{-n}) - U_n(\mathbf{s}'_n, \mathbf{s}_{-n}).$$

Thus, we can conclude that the non-cooperative game G is an exact potential game and admits at least one Nash Equilibrium. ■

Algorithm 1 Asynchronous BRD (ABRD) Algorithm

Input: $\mathbf{P}_a, \forall a \in A, \mathbf{P}_r, \forall r \in R, \hat{\mathbf{P}}_c, \forall c \in C$
Output: s^*
Initialization: *iteration* $= 0$, *stability* $=$ false, initialize $\mathbf{s}^{ite=0}$ randomly.
while not *stability* **do**
  iteration $\leftarrow$ *iteration* $+ 1$;
  Choose a random node $n \in N = U \cup C$
  Node n *updates its strategy* $\mathbf{s}_n^{*ite}$ *(Eq. 4) and* $U_n(\mathbf{s}_n^{*ite}, \mathbf{s}_{-n}^{*ite})$, *based on* $\mathbf{s}_{-\mathbf{n}}^{\mathbf{ite-1}}$
  if $|U_n(\mathbf{s}_n^{*ite}, \mathbf{s}_{-n}^{ite-1}) - U_n(\mathbf{s}_n^{*ite+1}, \mathbf{s}_{-n}^{ite})| \leq \delta$, δ small positive number, $\forall n \in N$ **then**
    stability $\leftarrow$ true
  end if
end while

Algorithm 2 Synchronous BRD (SBRD) Algorithm

Input: $\mathbf{P}_a, \forall a \in A, \mathbf{P}_r, \forall r \in R, \hat{\mathbf{P}}_c, \forall c \in C$
Output: s^*
Initialization: *iteration* $= 0$, *has_converged* $=$ **False**, $\mathbf{s}^{ite=0}$ randomly initialized.
while *has_converged* == **False do**
  Increment *iteration* by 1;
  for all $n \in N = U \cup C$ **do**
    Determine $\mathbf{s}_n^{*ite}$ *(Eq. 4) and* $U_n(\mathbf{s}_n^{*ite}, \mathbf{s}_{-n}^{ite-1})$, *given* $\mathbf{s}_{-\mathbf{n}}^{\mathbf{ite-1}}$
  end for
  if $|U_n(\mathbf{s}_n^{*ite}, \mathbf{s}_{-n}^{ite-1}) - U_n(\mathbf{s}_n^{*ite+1}, \mathbf{s}_{-n}^{ite})| \leq \delta$, δ small positive number, $\forall n \in N$ **then**
    has_converged $=$ **True**
  end if
end while

A. A GAME-THEORETIC APPROACH

The Best Response Dynamics (BRD) is a natural method that enables each node, i.e., target and collaborator node, to select its best response strategy to the strategies of the rest of the nodes in order for the non-cooperative game to converge to a NE via performing a local search. The BRD can be performed in an asynchronous or synchronous manner, where one node or all the nodes, play their best response strategies per iteration, respectively. The Asynchronous (ABRD) and the Synchronous BRD (SBRD) are described in Algorithms 1 and 2, respectively. The ABRD and SBRD converge to a Nash Equilibrium, given that the non-cooperative game G is an exact potential game [25]. The benefit of the SBRD over the ABRD is that no coordination is needed among the nodes regarding which node will play its best response strategy at each iteration. Another benefit is the expected lower execution time in order to converge to the NE as all the nodes play their best response strategies at the same iteration. On the other hand, the ABRD outperforms the SBRD, in terms of more accurately estimating the nodes' positioning and timing, as it avoids the herding effect that can be observed in the SBRD. The ABRD and SBRD algorithms are complementary in the sense that ABRD offers higher accuracy but slower convergence, while SBRD offers faster convergence but potentially lower accuracy.

B. A REINFORCEMENT LEARNING-BASED APPROACH

The main drawback of the ABRD and the SBRD algorithms is that they perform a local search of the strategy space and they can be trapped in local optimum solutions without converging to the best NE in terms of minimizing the positioning and timing estimation error. Also, the quality of the NE reached, heavily depends on the order that the nodes play in the ABRD case. In order to overcome those problems, the log-linear reinforcement learning (RL) approach is studied which performs the exploration and learning processes in order to converge to the best NE in terms of minimizing the estimation error. Log-linear algorithms are a set of algorithms that are designed to deal with decision-making problems in multi-agent systems. In this case, the RL agents iteratively adjust their strategies by considering their environment and the payoffs that they receive by selecting an action. Specifically, the log-linear algorithms are based on the probabilistic

approach that balances the exploration and exploitation and allows the RL agents to explore different strategies in order to converge toward an optimal decision. The log-linear nature of those algorithms stems from the probabilistic rule where the likelihood of selecting a particular strategy is proportional to the exponential of its payoff. Based on this approach the log-linear algorithms allow the RL agents to escape from local optima and converge to a Nash equilibrium and in some cases to achieve a Pareto optimal solution [26].

Two representative log-linear RL algorithms are studied in this research work, i.e., Binary-Logit (BL) and Max-Logit (MaxL), due to their inherent characteristics of fast learning of the NE, and converging to the Pareto Optimal Nash Equilibrium, if it exists, respectively. Both algorithms implement the exploration and learning processes. In the exploration phase, each node n randomly selects a strategy $\mathbf{s}_n$ with equal probability $P(\mathbf{s}_n) = \frac{1}{|S_n|}$ and explores the payoff $U_n(\mathbf{s}_n^{ite}, \mathbf{s}_{-n}^{ite})$ that it receives given the strategies of the rest of the nodes. Then, at the learning phase, each node updates its strategy based on the probabilistic rules Eq. 6a, 6b for the BL algorithm, and Eq. 7a, 7b for the MaxL algorithm. The BL and MaxL algorithms are described in Algorithm 3.

$$P(\mathbf{s}_n^{ite} = \mathbf{s}_n^{ite-1}) = \frac{e^{\beta U_n(\mathbf{s}_n^{ite-1})}}{e^{\beta U_n(\mathbf{s}_n^{ite-1})} + e^{\beta U_n(\mathbf{s}_n^{ite'})}} \tag{6a}$$

$$P(\mathbf{s}_n^{ite} = \mathbf{s}_n^{ite'}) = \frac{e^{\beta U_n(\mathbf{s}_n^{ite'})}}{e^{\beta U_n(\mathbf{s}_n^{ite-1})} + e^{\beta U_n(\mathbf{s}_n^{ite'})}} \tag{6b}$$

$$P(\mathbf{s}_n^{ite} = \mathbf{s}_n^{ite-1}) = \frac{e^{\beta U_n(\mathbf{s}_n^{ite-1})}}{\max\{e^{\beta U_n(\mathbf{s}_n^{ite-1})}, e^{\beta U_n(\mathbf{s}_n^{ite'})}\}} \tag{7a}$$

$$P(\mathbf{s}_n^{ite} = \mathbf{s}_n^{ite'}) = \frac{e^{\beta U_n(\mathbf{s}_n^{ite'})}}{\max\{e^{\beta U_n(\mathbf{s}_n^{ite-1})}, e^{\beta U_n(\mathbf{s}_n^{ite'})}\}} \tag{7b}$$

It is noted that $\beta \in \mathbb{R}^+$ is the learning parameter. For large values of β, the nodes thoroughly explore their strategy space, thus, converging to a better NE in terms of estimation error, by sacrificing longer convergence time. The proposed RL problem involves the optimization of the strategies of the nodes, i.e., targets and collaborator nodes, aiming at minimizing their positioning errors, especially in scenarios with limited GNSS services. Specifically, the environment consists of the nodes, which interact with each other, and each node's strategy impacts the overall system's performance. The examined RL problem is designed based on the Markov Decision Process (MDP) where the interactions among the nodes and the resulting states are captured by the transition mechanisms which are of probabilistic nature, i.e., Eq. 6a, 6b, 7a, 7b, and the selected strategies. The state at any given time is characterized by the current strategies of all the nodes and the corresponding payoffs that capture the system's configuration. The action is the selection of a strategy by each node from its available strategy space. The reward is defined as the payoff each node receives based on its strategy and the strategies of the other nodes. Modeling this problem as an RL problem is meaningful because it enables the nodes to adaptively learn their optimal strategies through exploration and exploitation, and address the limitations of the local search methods (ABRD and SBRD). By leveraging the log-linear RL algorithms, we facilitate the rapid learning of the nodes' strategies that converge toward a Pareto Optimal Nash Equilibrium. The complexity of the SBRD, BL and MaxL algorithms is $O(Ite)$, where Ite is the number of iterations that each algorithm needs in order to converge to the NE. The complexity of the ABRD algorithm is $O(Ite \cdot |N|)$. Please note that detailed numerical results and discussion about the overall algorithm complexity, in terms of actual execution time, are presented later in Section V.

Algorithm 3 BL [MaxL] Algorithm

Input: $\mathbf{P}_a, \forall a \in A, \mathbf{P}_r, \forall r \in R, \hat{\mathbf{P}}_c, \forall c \in C$
Output: s^*
Initialization: $ite = 0$, $Converged = 0$, $\mathbf{s}_n^{ite=0}, \forall n$.
while $Converged == 0$ **do**
  $ite = ite + 1$;
  Each node n selects $\mathbf{s}_n^{ite'}$ *with equal probability* $\frac{1}{|S_n|}$, *receives a payoff* $U_n(\mathbf{s}_n^{ite'})$ *and updates* $\mathbf{s}_n^{ite}$ *based on Eq. 6a, 6b* [Eq. 7a, 7b].
  The rest of the nodes keep their previous strategies, i.e., $\mathbf{s}_{-\mathbf{n}}^{\mathbf{ite}} = \mathbf{s}_{-\mathbf{n}}^{\mathbf{ite-1}}$.
  if $|\frac{\sum_{ite=0}^{T}\sum_{\forall n\in N} U_n^{ite}}{T} - \sum_{\forall n\in N} U_n^{ite}| \leq \delta$, δ small positive number **then**
    $Converged = 1$
  end if
end while

Based on the provided analysis in this Section, it is highlighted that the proposed game-theoretic approach for symbiotic PNT differs from gradient-free source-seeking algorithms [32], in several key aspects. The source-seeking algorithms focus on cooperative, gradient-free optimization to locate a source under disturbances and communication constraints, while on the other hand, the game-theoretic approach in this work models the problem as a non-cooperative potential game, where the nodes independently minimize their positioning and timing errors without explicit coordination. This decentralized framework leverages Best Response Dynamics (ABRD and SBRD) and Reinforcement Learning (Binary-Logit and Max-Logit) to converge to a Nash Equilibrium, and ensures its scalability and adaptability in dynamic environments. Unlike source-seeking, which relies on cooperative behavior and gradient-free methods, the game-theoretic approach provides theoretical guarantees for convergence and is specifically tailored for PNT applications, where the nodes do not necessarily share a common goal or

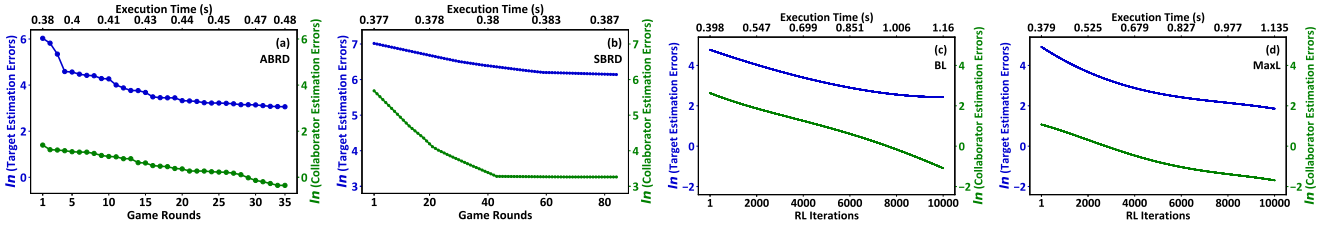

FIGURE 3. Estimation error and execution time under the: (a) ABRD, (b) SBRD, (c) BL, and (d) MaxL algorithms.

have the ability to coordinate explicitly. This makes the proposed method more suitable for large-scale, GPS-denied environments where the infrastructure and communication are limited.

V. NUMERICAL RESULTS

In this section, the performance of the proposed symbiotic PNT solution is evaluated via modeling and simulation, and it is also compared against other state-of-the-art approaches in the literature, in order to quantify its drawbacks and benefits. Specifically, the operational characteristics of our novel symbiotic PNT solution are initially presented in Section V-A, while a detailed scalability analysis for an increasing number of targets and collaborator nodes is demonstrated in Section V-B. The impact of targets' strategy space discretization is quantified in Section V-C, followed by a thorough comparative evaluation to existing alternative ground-based PNT solutions in the literature in Section V-D. Throughout our evaluation, unless otherwise explicitly stated, we adopted a symbiotic environment consisting of $|A| = 9$ anchor nodes, $|R| = 5$ RISs, $|C| = 4$ collaborator nodes, and $|U| = 5$ targets. The anchor nodes and RISs have been positioned following an intelligent deployment to reflect real-world use case scenarios. Specifically, the targets with higher IDs receive less support from the existing infrastructure (i.e., anchor nodes and RISs) and collaborator nodes, meaning fewer anchor nodes, RISs, and collaborator nodes are available to them. However, each target is guaranteed a minimum of three anchor nodes to receive signals. This intelligent deployment effectively captures how the estimation error of targets is influenced by the available support from anchor nodes, RISs, and collaborator nodes. Throughout our simulation-based results, the nodes' coordinates are consistently presented in meters, and the clock offset is measured in seconds. The variance in the clock offset remains small, and in realistic environments, it is negligible in comparison to the distance errors. Therefore, the results in our experiments primarily reflect the distance error, appropriately adjusted for the logarithmic scale and squared distances. Also, we consider $P = 2$ W, $G_j^{trans} = 0$ dB, $\forall j \in A_u \cup C_u$, $h_u = 1.5$ m, $\forall u \in U$, $M_r = 300$, $d_s = \lambda/2$ m, $k = 2.8$, $\alpha_r = 2.8$, $\beta = 0.1$, unless otherwise explicitly stated. The parameters are carefully selected to reflect realistic scenarios and ensure the robustness of the proposed symbiotic PNT framework. The chosen parameter values are based on typical settings in wireless communication and localization systems.

A. PURE PERFORMANCE AND OPERATION

Fig. 3a-3d illustrate the targets' and collaborator nodes' estimation error and the execution time of the ABRD, SBRD, BL, and MaxL algorithms, respectively. The results show that the RL-based algorithms outperform the game-theoretic ones in terms of the nodes' estimation error. This is due to the fact that the former ones follow the exploration and learning phases, thus, more thoroughly examining their available strategy space and converging to a better NE. However, the latter desirable outcome comes at the cost of longer execution times experienced by the RL-based approaches. The results also demonstrate that the MaxL algorithm converges to the best NE, achieving the lowest nodes' estimation error compared to all the other approaches. Focusing on the game-theoretic BRD algorithms, the results reveal that the SBRD converges faster to a NE than the ABRD algorithm, as it allows all the nodes to simultaneously update their strategies. However, the latter benefit comes at the cost of higher nodes' estimation error, as the herding phenomenon is observed, where all of them may update their strategies towards a less optimal direction.

The results presented in Fig. 3 stem from a single execution of the ABRD, SBRD, BL, and MaxL algorithms. However, all the four algorithms are characterized by stochastic components in their execution. In particular, the stochasticity in ABRD stems from the sequence of nodes playing their best response strategy, while in SBRD, the selection of the nodes' strategy is characterized by the herding phenomenon among the nodes. Also, the BL and MaxL RL-based algorithms are stochastic by their nature in terms of exploring random strategies during the exploration phase, and probabilistically selecting a strategy during the learning phase. Thus, we have performed a Monte Carlo analysis with 10,000 executions of the four algorithms. The box plots in Fig. 4a-4b present the means (yellow line) and the 25 (lower edge) and 75 (upper edge) percentile of the system's estimation error and the execution time, respectively, for all the four examined algorithms. Fig. 4c presents the system's estimation error, execution time, and their trade-off, i.e., product of the two previous parameters, for the four algorithms. The results confirm that the MaxL achieves the best results in terms of

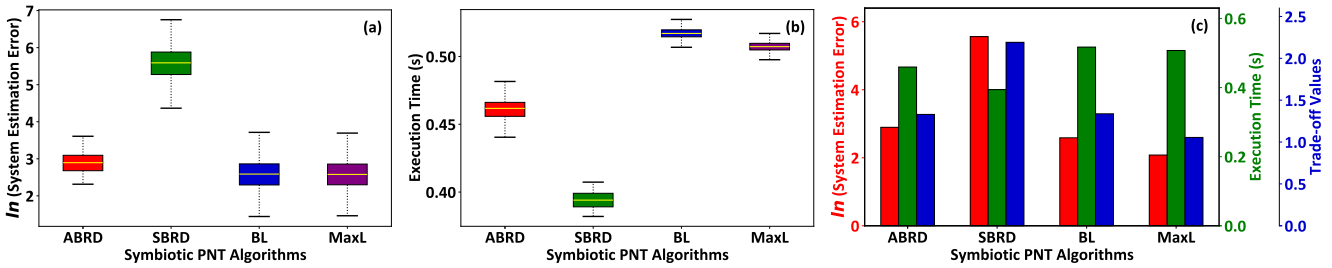

FIGURE 4. Mean, 25 and 75 percentile, of (a) system's estimation error, and (b) execution time for the ABRD, SBRD, BL, and MaxL algorithms, and (c) comparative evaluation of the four algorithms.

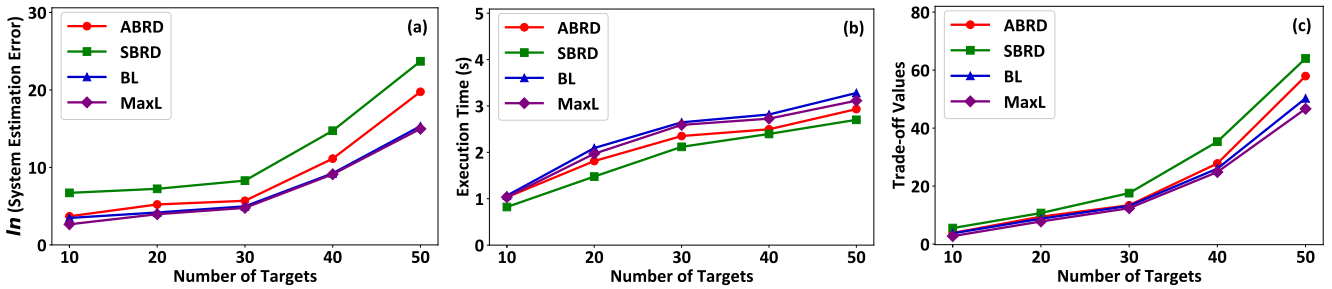

FIGURE 5. Scalability analysis for an increasing number of targets for the ABRD, SBRD, BL, and MaxL algorithms.

system's estimation error as it converges to the best NE. Also, the RL-based algorithms outperform the BRD algorithms with respect to the system's estimation error at the cost of a longer execution time. We further observe that higher stochasticity exists in the RL-based algorithms in terms of the system's estimation error (Fig. 4a) given the highly stochastic processes of exploration and learning. However, the stochasticity of the RL-based algorithms is lower than the one of the BRD algorithms in terms of execution time (Fig. 4b). This outcome is observed given that the RL algorithms explore the whole strategy space of the nodes, while the execution time of the BRD algorithms depends on the evolution of the nodes' strategies selection.

B. SCALABILITY ANALYSIS

Fig. 5a-5c demonstrates the system's estimation error, the execution time, and their trade-off, respectively, for the ABRD, SBRD, BL, and MaxL algorithms considering a large number of targets. The results show that as the number of targets increases the system's estimation error, the algorithms' execution time, and their trade-off increase for all the proposed symbiotic PNT algorithms. By taking a closer look into the results, we observe that the same percentage increase in the number of targets results in a higher percentage increase in the system's estimation error achieved by the RL-based algorithms, as the stochasticity in the system increases due to the impact of the nodes' strategies selection among each other (Fig. 5a). On the other hand, we observe that the BRD algorithms are more heavily impacted in terms of their execution time compared to the RL-based algorithms (Fig. 5b). This observation stems from the sequential decision-making process for the ABRD and from the herding effect for the SBRD, compared to the RL-based algorithms, where all the nodes perform in parallel the exploration and learning phases. The rapid deterioration of the system's estimation error in the case of RL-based algorithms drives the trade-off value (Fig. 5c) to increase rapidly for the MaxL and BL algorithms, with respect to the number of targets.

Fig. 6a-6c present the same set of results as above, as a function however of the number of collaborator nodes considering a large number of them. The results confirm the theoretical analysis, i.e., as the number of collaborator nodes' increases, the system's estimation error decreases, the execution time of the algorithms increases, and their trade-off decreases. This outcome is very well-expected, as the targets can measure more pseudoranges from more collaborator nodes and more accurately determine their positioning and timing. Also, we observe that the BRD algorithms benefit more compared to the RL-based algorithms in terms of lowering more rapidly the system's estimation error for an increasing number of collaborator nodes (Fig. 6a). On the other hand, the RL-based algorithms benefit in terms of a slower increase of their execution time compared to the BRD algorithms for an increasing number of collaborator nodes (Fig. 6b). This outcome is reasonable as the collaborator nodes added to the system have already an initial rough estimate of their coordinates, thus, the stochasticity introduced in the RL-based algorithms is limited. In contrast, the BRD algorithms have to deal with the higher cardinality of nodes

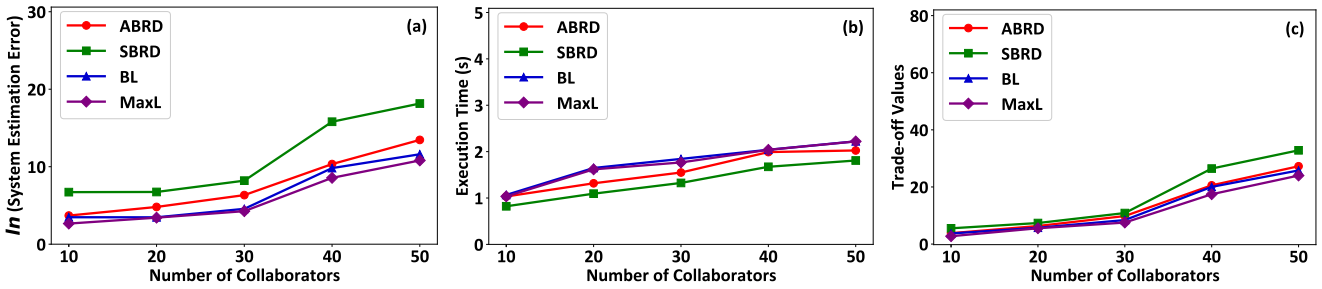

FIGURE 6. Scalability analysis for an increasing number of collaborator nodes for the ABRD, SBRD, BL, and MaxL algorithms.

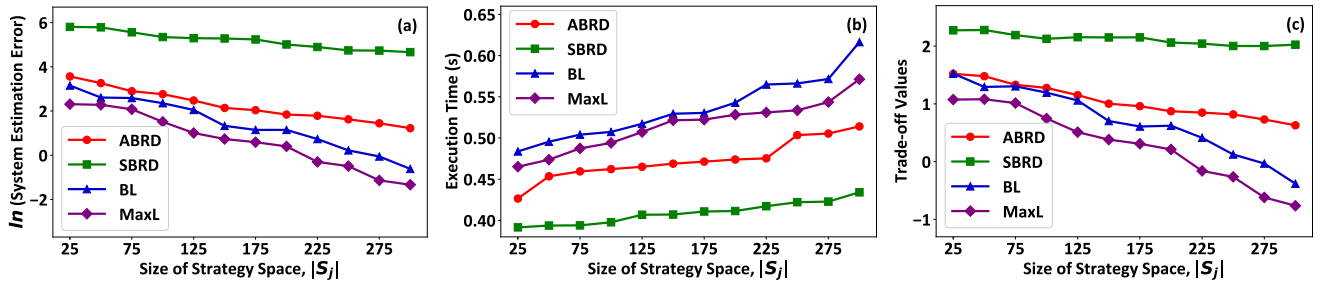

FIGURE 7. Trade-off between system's estimation error and execution time for increasing strategy space discretization for the ABRD, SBRD, BL, and MaxL algorithms.

added in the system, which contributes to the faster increase of their execution time due to the sequential decision-making (ABRD) or the herding effect (SBRD). The combined effect of the system's estimation error and the algorithm's execution time is presented via the trade-off value in Fig. 6c.

C. TRADING OFF ACCURACY AND COMPLEXITY VIA STRATEGY SPACE DISCRETIZATION

Fig. 7a-7c depicts the system's estimation error, the algorithm's execution time, and the trade-off value, respectively, as a function of the strategy space's cardinality $|S_j|, \forall j \in U \cup C$, for the four algorithms. It is noted that the nodes' polar coordinates $\hat{\mathbf{x}}_j, \forall j \in U \cup C$ are discretized considering an equal step for the radius and the angle, respectively, while the maximum radius, i.e., neighborhood, is derived by the node's j signal strength (Section III). The results reveal that as the cardinality of the strategy space increases, all the four symbiotic PNT solutions achieve a more accurate estimation of the nodes' positioning and timing (Fig. 7a), with the cost of higher execution time (Fig. 7b), resulting in decreasing trade-off values (Fig. 7c). Also, we conclude that the RL-based algorithms are more positively impacted by the increasing discretization of the nodes' strategy space compared to the BRD algorithms, as their system's estimation error decreases more rapidly (Fig. 7a). This benefit is achieved given that the RL-based algorithms explore thoroughly a larger strategy space and converge to a more accurate PNT solution with the cost of a rapidly increasing execution time (Fig. 7b). Thus, the RL-based algorithms experience a faster decrease in their trade-off values (Fig. 7c) due to their substantial improvement of the system's estimation error compared to the BRD algorithms.

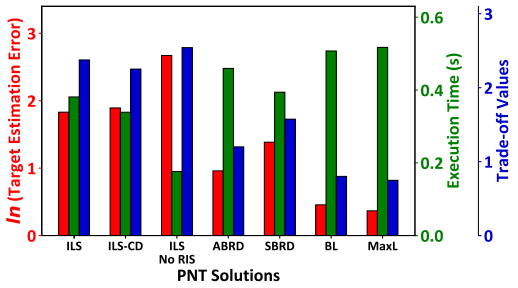

FIGURE 8. Comparative evaluation.

D. COMPARATIVE EVALUATION

In this section, a thorough comparative evaluation of the proposed symbiotic PNT solution against three dominant existing ground-based PNT solutions is performed. The comparative scenarios include the four proposed symbiotic PNT algorithms, and the traditional iterative least square (ILS) algorithm: (i) ILS: exploiting the multilateration technique based on the anchor nodes' signals, and the RISs' reflected signals [20]; (ii) ILS – No RIS: the traditional ILS algorithm exploiting only the anchor nodes' signals [6]; and (iii) ILS-CD: the ILS algorithm selecting four in total anchor nodes and RISs that reside in the closest distant (CD) to the targets. Fig. 8 presents the targets' estimation error, the algorithms' execution time, and their trade-off, under all

the comparative scenarios. It is noted that only the targets' estimation error is presented for fairness in the comparison, as the ILS, ILS-No RIS, and ILS-CD algorithms do not consider the collaborator nodes. The results reveal that the MaxL algorithm outperforms all the comparative scenarios in terms of PNT accuracy, bearing however higher execution time. On the other hand, the traditional ILS-No RIS algorithm presents the worst results in terms of accuracy, as this solution neither exploits the benefits of the RISs nor of the collaborator nodes. Focusing on the broader picture derived by the trade-off value, we observe that the existing ground-based PNT solutions present the worst results, either because they do not consider the symbiotic relationship among the nodes, i.e., ILS and ILS-CD algorithm, or because additionally, they do not exploit the most recent advances in next-generation wireless networks through the RIS technology, i.e., ILS-No RIS. Also, it is noted that the ILS-CD worse accuracy than the ILS as it limits the target to select the four closest anchor nodes and RISs compared to the ILS that exploits all the nearby infrastructure. However, ILD-CD converges faster than the ILS. Thus, there is a persistent tradeoff between the accuracy and execution time.

VI. KEY INNOVATIONS OF SPNT FRAMEWORK

Focusing on the mitigation of the high infrastructure costs and computational complexity of existing PNT solutions, we have introduced several innovative features in our proposed SPNT framework that directly tackle these challenges.

Firstly, we reduce the reliance on costly infrastructure by introducing the concept of collaboration between targets, collaborator nodes, and Reconfigurable Intelligent Surfaces (RISs). Instead of depending on dense deployments of high-cost anchor nodes (e.g., gNBs), the SPNT system utilizes bio-inspired collaborative behaviors among nodes with rough estimates of their positions. This approach leverages a mutualistic relationship, where targets benefit from each other's presence and from the relatively low-cost RISs, which are easier to deploy than additional anchor nodes. The RISs passively reflect signals to enhance positioning accuracy without the need for extensive signal transmission infrastructure. This reduces the overall infrastructure required for accurate PNT services and lowers the associated costs.

In terms of computational complexity, we recognize that existing PNT solutions rely on computationally expensive algorithms for signal processing and optimization. In contrast, our SPNT framework introduces two distinct classes of algorithms, i.e., Game Theory-based and Reinforcement Learning-based approaches, each with its own tradeoffs. We specifically address the computational overheads by offering a range of solutions:

1) The **Asynchronous Best Response Dynamics (ABRD)** and **Synchronous Best Response Dynamics (SBRD)** game-theoretic algorithms provide relatively lower computational complexity by employing local, decentralized decision-making. These algorithms, by focusing on local strategy updates and decentralized optimization, allow for scalable computation without centralized heavy processing.
2) For scenarios demanding higher accuracy, the **Binary-Logit (BL)** and **Max-Logit (MaxL) RL algorithms** support the more thorough exploration of the strategy space. These algorithms have higher computational costs compared to the previously proposed game-theoretic algorithms due to the exploration phase, and they ultimately converge to more accurate positioning solutions, particularly useful in scenarios where precise positioning is important, e.g., rescue missions. The MaxL algorithm, in particular, achieves the lowest estimation error while balancing this against the execution time through stochastic processes, as shown in our simulation results (please refer to Section V-D).

Moreover, our framework optimizes the RIS phase shifts to enhance the signal strength received by the targets and the collaborator nodes and improves the accuracy of pseudorange measurements. This additional layer of optimization provides substantial accuracy improvements without significantly increasing computational costs, as RISs operate with low energy consumption and minimal processing needs. Lastly, we performed an extensive simulation-based evaluation, comparing the performance of the game-theoretic and RL-based methods. This comparison clearly demonstrates that while the RL algorithms may involve longer execution times, they offer superior accuracy in complex or large-scale topologies and ensure that our SPNT solution scales effectively without overwhelming computational resources. Also, based on the design of the proposed SPNT framework, it is noted that an increasing number of anchor nodes and collaborator nodes generally reduces estimation error, as more reference points are available for pseudorange measurements. However, this also increases the execution time due to higher computational complexity. Moreover, the number of RIS elements significantly impacts the signal strength and, consequently, the accuracy of the pseudorange measurements. Specifically, more elements can lead to better accuracy but increase the complexity of the phase shift optimization. Additionally, the choice of the learning parameter β affects the trade-off between exploration and exploitation in the RL algorithms, where a higher β supports more exploration and potentially more accurate solutions but at the cost of longer execution times. Finally, the path loss exponent α_r affects the signal attenuation over distance, and a higher value can result in faster signal decay, and thus, make the accurate positioning more challenging, especially in environments with many obstacles.

VII. CONCLUSION

In this work, we introduced a novel symbiotic Positioning, Navigation, and Timing (PNT) solution by exploiting the mutualistic benefits among the anchor nodes, collaborator nodes, Reconfigurable Intelligent Surfaces (RISs), and targets. All the symbiotic entities collaborate among each other in order to minimize their positioning and timing error.

The problem of minimizing the nodes' positioning and timing error was formulated as a non-cooperative game among them and the principles of potential games were followed in order to show the existence of a Nash Equilibrium. Best Response Dynamics and log-linear RL-based algorithms were introduced to determine the Nash Equilibrium. A detailed numerical and comparative evaluation was performed to demonstrate the operational characteristics and benefits of the proposed symbiotic PNT solution.

Part of our future work is to further extend the introduced bio-inspired PNT concept into the parasitism PNT solution, in order to be able to identify, detect, and eject malicious anchor or collaborator nodes, i.e., parasites, from the symbiotic PNT system. The parasitism PNT solution can be realized by exploiting the received signal strength indicator along with positioning and timing information provided by the other collaborator nodes in order to detect the malicious actor(s). Then, a symbiotic trust score mechanism can be developed to eject the malicious nodes from the symbiotic environment.

REFERENCES

[1] L. Wielandner, E. Leitinger, and K. Witrisal, "RSS-based cooperative localization and orientation estimation exploiting antenna directivity," *IEEE Access*, vol. 9, pp. 53046–53060, 2021.

[2] *National Research and Development Plan for Positioning, Navigation, and Timing Resilience*, Nat. Sci. Technol. Council, Washington, DC, USA, 2011.

[3] C. L. Nguyen, O. Georgiou, G. Gradoni, and M. Di Renzo, "Wireless fingerprinting localization in smart environments using reconfigurable intelligent surfaces," *IEEE Access*, vol. 9, pp. 135526–135541, 2021.

[4] R. Santos, R. Leonardo, M. Barandas, D. Moreira, T. Rocha, P. Alves, J. P. Oliveira, and H. Gamboa, "Crowdsourcing-based fingerprinting for indoor location in multi-storey buildings," *IEEE Access*, vol. 9, pp. 31143–31160, 2021.

[5] R. Chen, X. Huang, Y. Zhou, Y. Hui, and N. Cheng, "UHF-RFID-Based real-time vehicle localization in GPS-less environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9286–9293, Jul. 2022.

[6] K. Cengiz, "Comprehensive analysis on least-squares lateration for indoor positioning systems," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2842–2856, Feb. 2021.

[7] J. Zhao, Y. Zhang, S. Ni, and Q. Li, "Bayesian cooperative localization with NLOS and malicious vehicle detection in GNSS-challenged environments," *IEEE Access*, vol. 8, pp. 85686–85697, 2020.

[8] E. Zhang and N. Masoud, "Increasing GPS localization accuracy with reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2615–2626, May 2021.

[9] V. Havyarimana, Z. Xiao, A. Sibomana, D. Wu, and J. Bai, "A fusion framework based on sparse Gaussian–Wigner prediction for vehicle localization using GDOP of GPS satellites," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 680–689, Feb. 2020.

[10] L. Ruan, G. Li, W. Dai, S. Tian, G. Fan, J. Wang, and X. Dai, "Cooperative relative localization for UAV swarm in GNSS-denied environment: A coalition formation game approach," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 11560–11577, Jul. 2022.

[11] R. Wang, C. Xu, J. Sun, S. Duan, and X. Zhang, "Cooperative localization for multi-agents based on reinforcement learning compensated filter," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 10, pp. 2820–2831, Oct. 2024.

[12] B. Jiang, B. D. O. Anderson, and H. Hmam, "3-D relative localization of mobile systems using distance-only measurements via semidefinite optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 1903–1916, Jun. 2020.

[13] F. Jin, K. Liu, C. Liu, T. Cheng, H. Zhang, and V. C. S. Lee, "A cooperative vehicle localization and trajectory prediction framework based on belief propagation and transformer model," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2746–2758, Feb. 2024.

[14] A. Chakraborty, K. M. Brink, and R. Sharma, "Cooperative relative localization using range measurements without a priori information," *IEEE Access*, vol. 8, pp. 205669–205684, 2020.

[15] S. Zhao, X.-P. Zhang, X. Cui, and M. Lu, "A closed-form localization method utilizing pseudorange measurements from two nonsynchronized positioning systems," *IEEE Internet Things J.*, vol. 8, no. 2, pp. 1082–1094, Jan. 2021.

[16] O. Y. Al-Jarrah, A. S. Shatnawi, M. M. Shurman, O. A. Ramadan, and S. Muhaidat, "Exploring deep learning-based visual localization techniques for UAVs in GPS-denied environments," *IEEE Access*, vol. 12, pp. 113049–113071, 2024.

[17] M. S. Siraj, P. Charatsaris, M. Diamanti, E. E. Tsiropoulou, and S. Papavassiliou, "HEROES: Humanitarian emergency response based on UAV-enabled integrated sensing and communication, positioning, and satisfaction games," *ACM J. Auto. Transp. Syst.*, vol. 2025, pp. 1–25, Feb. 2025, doi: 10.1145/3715334.

[18] B. Zhou, Z. Gu, F. Gu, P. Wu, C. Yang, X. Liu, L. Li, Y. Li, and Q. Li, "DeepVIP: Deep learning-based vehicle indoor positioning using smartphones," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13299–13309, Dec. 2022.

[19] M. S. Siraj, J. R. Atencio, and E. E. Tsiropoulou, "Dead-on-target: An accurate alternative positioning, navigation, and timing solution," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2024, pp. 3377–3382.

[20] M. S. Hossain, N. Irtija, E. E. Tsiropoulou, J. Plusquellic, and S. Papavassiliou, "Reconfigurable intelligent surfaces enabling positioning, navigation, and timing services," in *Proc. ICC - IEEE Int. Conf. Commun.*, May 2022, pp. 4625–4630.

[21] A. Elzanaty, A. Guerra, F. Guidi, and M.-S. Alouini, "Reconfigurable intelligent surfaces for localization: Position and orientation error bounds," *IEEE Trans. Signal Process.*, vol. 69, pp. 5386–5402, 2021.

[22] J. Bingham, M. S. Siraj, and E. E. Tsiropoulou, "MERCURY: Multilateral bargaining on reconfigurable intelligent surfaces for alternative positioning, navigation, and timing," in *Proc. IEEE 10th World Forum Internet Things (WF-IoT)*, Nov. 2024, pp. 924–929.

[23] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Towards ubiquitous positioning by leveraging reconfigurable intelligent surface," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 284–288, Jan. 2021.

[24] M. S. Siraj, M. S. Hossain, R. Brown, E. E. Tsiropoulou, and S. Papavassiliou, "Incentives to learn: A location-based federated learning model," in *Proc. Global Inf. Infrastruct. Netw. Symp. (GIIS)*, Sep. 2022, pp. 40–45.

[25] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, no. 1, pp. 124–143, May 1996.

[26] Y. Xu, J. Wang, and Q. Wu, "Distributed learning of equilibria with incomplete, dynamic, and uncertain information in wireless communication networks," in *Game Theory Framework Applied to Wireless Communication Networks*. Information Science Reference, 2016, pp. 63–86.

[27] M. S. Siraj, A. B. Rahman, E. E. Tsiropoulou, S. Papavassiliou, and J. Plusquellic, "Symbiotic positioning, navigation, and timing," in *Proc. 19th Int. Conf. Distrib. Comput. Smart Syst. Internet Things (DCOSS-IoT)*, Jun. 2023, pp. 261–268.

[28] M. S. Siraj, A. B. Rahman, P. Charatsaris, E. E. Tsiropoulou, and S. Papavassiliou, "Positioning, navigation, and timing on the air," in *Proc. 19th Int. Conf. Distrib. Comput. Smart Syst. Internet Things (DCOSS-IoT)*, Jun. 2023, pp. 661–668.

[29] M. S. Siraj, A. B. Rahman, M. Diamanti, E. E. Tsiropoulou, and S. Papavassiliou, "Alternative positioning, navigation, and timing enabled by games in satisfaction form and reconfigurable intelligent surfaces," *IEEE Syst. J.*, vol. 17, no. 3, pp. 5035–5046, Mar. 2023.

[30] T. Liu, Z. Qin, Y. Hong, and Z.-P. Jiang, "Distributed optimization of nonlinear multiagent systems: A small-gain approach," *IEEE Trans. Autom. Control*, vol. 67, no. 2, pp. 676–691, Feb. 2022.

[31] M. S. Siraj, J. R. Atencio, and E. E. Tsiropoulou, "PANTHER: A power-optimized and accurate positioning, navigation, and timing with high efficiency and reliability," *IEEE Open J. Commun. Soc.*, early access, Dec. 23, 2024, doi: 10.1109/OJCOMS.2024.3521293.

[32] Z. Jin, H. Li, Z. Qin, and Z. Wang, "Gradient-free cooperative source-seeking of quadrotor under disturbances and communication constraints," *IEEE Trans. Ind. Electron.*, vol. 72, no. 2, pp. 1969–1979, Feb. 2025.

MD SADMAN SIRAJ (Member, IEEE) received the bachelor's degree in electrical and electronic engineering from the University of Dhaka, in 2020, and the master's degree in computer engineering from The University of New Mexico, in 2023. He is currently a Ph.D. Student and a Research Assistant with the School of Electrical, Computer, and Energy Engineering, Arizona State University. His research interests include alternative positioning, navigation, and timing (APNT) services and novel localization techniques, resource management, game theory, and reinforcement learning.

EIRINI ELENI TSIROPOULOU (Senior Member, IEEE) is currently an Associate Professor with the School of Electrical, Computer, and Energy Engineering, Arizona State University, and the Director of the Performance and Resource Optimization in Networks–PROTON Laboratory. Her main research interests include cyber-physical social systems and wireless heterogeneous networks, with an emphasis on network modeling and optimization, resource orchestration in interdependent systems, reinforcement learning, game theory, network economics, and the Internet of Things. Four of her articles received the Best Paper Award at IEEE WCNC, in 2012, ADHOCNETS, in 2015, IEEE/IFIP WMNC 2019, and INFOCOM 2019 by the IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling. She received the NSF CRII Award, in 2019, and the Early Career Award from the IEEE Communications Society Internet Technical Committee, in 2019. In 2017, she was selected by the IEEE Communication Society–N2Women–as one of the Top Ten Rising Stars in the communications and networking field.

SYMEON PAPAVASSILIOU (Senior Member, IEEE) is currently a Professor with the School of Electrical and Computer Engineering (ECE), National Technical University of Athens. From 1995 to 1999, he was a Senior Technical Staff Member with AT&T Laboratories, New Jersey. In August 1999, he joined the ECE Department, New Jersey Institute of Technology, USA, where he was an Associate Professor, until 2004. He has an established record of publications in his field of expertise, with more than 450 technical journals and conference-published papers. His main research interests include modeling, optimization, and performance evaluation of distributed complex systems, and social networks.

JIM PLUSQUELLIC (Senior Member, IEEE) received the M.S. and Ph.D. degrees in computer science from the University of Pittsburgh. He is currently a Professor in electrical and computer engineering with The University of New Mexico. He received the "Outstanding Contribution Award" from IEEE Computer Society, in 2012 and 2017, for co-founding and for his contributions to the Symposium on Hardware-Oriented Security and Trust (HOST).