Alternative Positioning, Navigation, and Timing enabled by Games in Satisfaction Form and Reconfigurable Intelligent Surfaces

Md Sadman Siraj, Aisha B Rahman, Maria Diamanti, *Graduate Student Member, IEEE*, Eirini Eleni Tsiropoulou, *Senior Member, IEEE*, and Symeon Papavassiliou, *Senior Member, IEEE*

***Abstract*—The potential degradation of the Global Positioning System and other Global Navigation Satellite Systems under several circumstances gives rise to the development of alternative position navigation and timing (PNT) technologies aiming at maintaining efficient and safe operations. In this paper, we exploit the use of Reconfigurable Intelligent Surfaces (RIS) as enablers for the design of an alternative PNT solution, with improved accuracy and efficiency. The specific problem of RISs' orchestration and configuration is treated via the adoption of Game Theory and Reinforcement Learning (RL). Initially, a Satisfaction Game is formulated and solved among the targets, enabling them to autonomously determine the optimal number of RISs that will contribute to their PNT service, while the specific set of RISs to be used is determined by a novel RL algorithm. In order to further maximize the received signal strength at each target of the reflected signals from the specific set of RISs, the phase shift optimization of the latter is performed. Based on the above, an Iterative Least Squares (ILS) algorithm is adopted, following the multilateration technique, in order for each target to estimate its position and timing. The performance evaluation of the proposed approach is achieved via modeling and simulation.**

***Index Terms*—Positioning, Navigation, and Timing (PNT), Satisfaction Games, Reconfigurable Intelligent Surfaces (RISs), Reinforcement Learning (RL).**

I. Introduction

Positioning, Navigation, and Timing (PNT) systems are becoming more and more prevalent and ubiquitous in several technology and infrastructure domains [1], such as wireless communications, public safety and emergency management, energy distribution, transportation, banking and finance, weather forecasting, agriculture, and military missions to name a few [1]. Besides, PNT systems constitute the vital structural component of the most popular and recognizable service of outdoor localization worldwide, i.e., the Global Navigation Satellite System (GNSS), with the Global Positioning System (GPS) being its most representative PNT provider. However, the GPS service is not invulnerable and can be denied for various reasons. The GPS signal is particularly susceptible to geographical and environmental changes (e.g., indoor environments, urban canyons), as well as unintentional or man-made interference to the satellite signals at the receiver (e.g., attenuation, jamming) [2]. As a consequence, the development of alternative PNT systems with increased resilience to the aforementioned vulnerabilities has been already identified as one of the national planning objectives in the USA [3].

In this paper, aligned with the latter vision, we scrutinize the synergy between conventional PNT systems and the key 6G technology of Reconfigurable Intelligent Surfaces (RISs) [4], as a means of ameliorating the availability and accuracy of the PNT services of mobile targets. The proper orchestration and configuration of the RISs is achieved by exploiting the theory of Satisfaction Games [5] and the Reinforcement Learning (RL). Specifically, we introduce a Satisfaction Game-based approach, according to which each target autonomously determines the optimal number of RISs required to contribute to its PNT service. Considering that each target's energy consumption increases as the number of signals reflected from the RISs increase due to decoding, each target seeks to strike a balance between its own energy availability and the received signal's shadowing probability. Then, a low-complexity RL-based selection of the specific RISs that will be actually used for each target's PNT service provisioning follows, along with the RISs' phase shift optimization. The ultimate goal of this process is to minimize the position and timing error for the mobile targets.

A. Related Work

Extensive research efforts have been recently devoted on the problem of designing alternative PNT systems to complement or even substitute GPS operation in cases of GPS-denial [6]. In [7], the authors adopt the strapdown inertial navigation system to provide PNT services, while its operation is improved by exploiting the information of the velocity attitude, as well as combining the benefits from the position from the doppler velocity log, the celestial navigation, and the GNSS systems. In [8], a Kalman filter-based approach is introduced, combining the outputs of multiple GPS systems with an Inertial Measurement Unit (IMU) to improve the PNT services for the ground vehicles. Several multisensor localization and navigation algorithms are presented in [9], by exploiting the exchange of information among neighboring targets, in order to provide PNT services in large-scale Internet of Things (IoT) environments.

Manuscript received: September 19, 2022. Corresponding author: E.E. Tsiropoulou (email: eirini@unm.edu).

Md S. Siraj, A. B Rahman, and E.E. Tsiropoulou are with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA, 87131 (email: {mdsadmansiraj96, arahman3, eirini}@unm.edu). The research of Dr. Tsiropoulou was partially supported by the NSF CNS Award # 2219617

M. Diamanti and S. Papavassiliou are with the School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece, 15780 (email: mdiamanti@netmode.ntua.gr, papavass@mail.ntua.gr).

Another area of research proposes the utilization of artificial intelligence for the design of alternative PNT systems, by especially placing the focus on the introduction of novel RL techniques [10]. A deep Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) is designed in [11] to estimate the target's position, by combining the target's position measurements derived from the holographic radar and the Radio Positioning System (RPS). A set of gradient ascent and log-linear learning models is used in [12] to enable the target to select the most beneficial signals and autonomously perform its position and timing, while minimizing the corresponding estimation error. A Support Vector Regression (SVR) model is presented in [13] and trained offline, by using the information of the targets' density, the number of anchor nodes, their transmission range, and the number of training iterations. The goal of this machine learning-based approach is to minimize the average localization error of the targets, at the cost of increased training time of the model.

Nevertheless, the majority of the aforementioned alternative PNT solutions are characterized by high infrastructure cost and increased computing complexity. Towards addressing these challenges, the key 6G technology of RISs [4] has been recently proposed to be incorporated into existing low-computational complexity PNT solutions [14]. RISs have attracted increasing research and commercial popularity due to their salient attributes, such as their low-cost and flexible deployment on any static or mobile surface compared to the traditional anchor nodes, the reconfigurability of their reflected signals' phase shifts in a software-defined manner, and their passive operation [15]. Concerning their utilization for enhanced-performance PNT services, the problem of Line of Sight (LoS) communication links absence is addressed in [16], where a single RIS is used to construct a virtual LoS path among the anchor nodes and the target. Subsequently, the latter one is able to determine its position through the multilateration technique. Similarly, in [17], a target determines its position by exploiting the reflected signals from the RISs, and its position and orientation error bounds are studied and analyzed. In [18], the authors optimize the RISs' phase shifts, as a means of improving the received signal strength at the target, and the latter one further determines its position. The same problem is addressed in [19], where the RISs' phase shift optimization and target's positioning are addressed in a fully distributed manner.

B. Contributions and Outline

Apparently, the design of alternative PNT systems that leverage the technology of RISs is still in its infancy, while the major problem of the existing proposed solutions' high computational complexity remains notably unsolved. Also, the critical technical issues of selecting an optimal number and set of RISs to contribute to the targets PNT services, while optimizing the RISs' elements phase shifts to further improve the PNT solution's accuracy, have not been addressed in the existing literature. Aiming to make a first step towards filling this gap, we tackle the problem of RISs' orchestration and configuration within a PNT system via the adoption of Game Theory and Reinforcement Learning (RL). The main contributions of this work are summarized as follows.

1) A Satisfaction Game is formulated and solved among the targets, enabling them to autonomously identify and determine the optimal number of RISs that will contribute to their PNT service. The Satisfaction Game's solution is an Efficient Satisfaction Equilibrium (ESE) point, at which the targets satisfy their personal energy and shadowing probability constraints, while being charged with the lowest possible energy cost.
2) A novel RL algorithm is proposed to enable the targets to autonomously select the specific set of RISs that will contribute to their PNT service, given the optimal number of RISs as already determined by the Satisfaction Game and the corresponding ESE point. The goal of the RL-based RIS selection is to minimize the estimation error of each target's PNT service, increasing in this way the service's accuracy.
3) The RISs' phase shift optimization is performed with the aim to maximize the received signal strength at each target of the reflected signals from the specific set of RISs that has been determined via the RL-based RIS selection algorithm. Then, the targets are allowed to accurately measure the pseudoranges, and thus, estimate their position, timing, and navigation (in case of moving targets). For the latter purpose, the Iterative Least Squares (ILS) algorithm is adopted, following the multilateration technique.

The remainder of this paper is organized as follows. Section II introduces the system model and provides an overview of the proposed framework. In Section III, the selection of the optimal number of RISs by the targets is presented following a Satisfaction Game-based approach, while in Section IV, the Reinforcement Learning-based selection of the specific set of RISs by each target is described. Section V performs the RISs' phase shift optimization, and Section VI implements the Iterative Least Squares algorithm to determine the targets' position and timing. A detailed set of numerical and comparative evaluation results is presented in Section VII. Finally, Section VIII concludes the paper.

II. Overview of the Proposed Framework

A. System Model

We consider an area $\mathcal{A}$, where a set of anchor nodes $A = \{1, \ldots, a, \ldots, |A|\}$, a set of RISs $R = \{1, \ldots, r, \ldots, |R|\}$, and a set of targets $N = \{1, \ldots, n, \ldots, |N|\}$ reside. The coordinates of the anchor nodes, the RISs, and the targets are denoted as $(x_a, y_a, z_a), \forall a \in A$, $(x_r, y_r, z_r), \forall r \in R$, and $(x_n, y_n, z_n), \forall n \in N$, respectively. The anchor nodes and RISs are static and their coordinates are known, while the targets can be either mobile or static and their coordinates are unknown. Each target measures the pseudorange $d_{a,n}$ [m] from each anchor node $a \in A$ based on the received signal strength of the direct signal transmitted by the anchor node a, in order to calculate its position and timing (details in Section VI). As a means of increasing the accuracy of its PNT

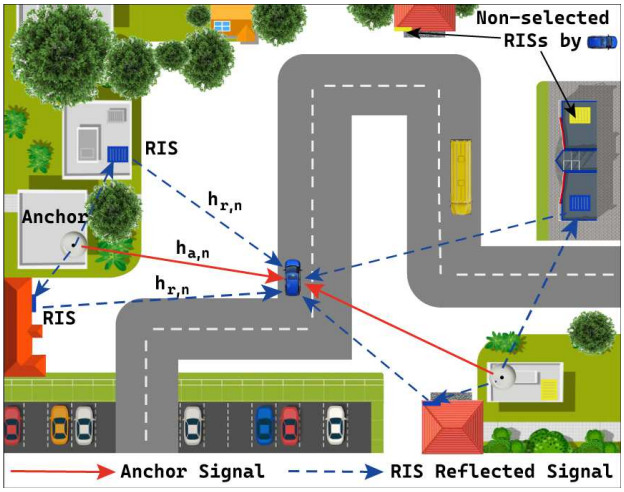

Fig. 1: Overview of the RIS-assisted system consisting of targets and anchor nodes.

service, each target leverages at the same time the reflected signals from a set of properly selected RISs [20], [21]. For this purpose, each target first selects the optimal number of RISs $|R_n|$ (see Section III) and identifies the specific set of RISs $R_n = \{1, \ldots, r, \ldots, |R_n|\}$ that will contribute to its PNT service (see Section IV). Specifically, each anchor node transmits a beacon signal with constant transmission power in order to support the targets' PNT services. The target receives the direct signals from the anchor nodes, as well as the reflected signals from the selected RISs, while treating the rest of the received signals as noise. Then, the measurement of the pseudoranges $d_{r,n}$ [m] from each selected RIS $r \in R_n$ follows, as described in Section VI [22]. An overview of the considered RIS-assisted network consisting of the targets and the anchor nodes is presented in Fig. 1.

As a result, each target receives beacon signals from $|A|$ anchor nodes and $|R_n|$ selected RISs, each of them consisting of B [bits] of information. The overall consumed energy $E_n(|R_n|)$ [J] by each target to decode the signals and measure the corresponding pseudoranges $d_{a,n}$ (from each anchor node $a \in A$) and $d_{r,n}$ [m] (from each selected RIS $r \in R_n$), is given as follows [23],

$$E_n(|R_n|) = (|A| \cdot B + |R_n| \cdot B) \cdot E_b, \tag{1}$$

where E_b [J/bit] is the target's unit energy consumed for receiving one bit of a beacon signal and calculating the respective pseudorange.

The targets experience a shadowing effect due to the physical obstacles in the surrounding environment, which can contribute to non-Line of Sight (NLoS) communication paths between the anchor nodes and the selected RISs with respect to the target. Apparently, as the number of selected RISs by each target increases, the probability of creating a constructive beam of Line of Sight (LoS) communication between the anchor nodes and the selected RISs also increases, mitigating practically the experienced shadowing effect. Based on the above observation, we define the shadowing probability for a target $n, \forall n \in N$, as follows [24],

$$P_s^n(|R_n|) = 1 - \frac{|A| + |R_n|}{|A| + |R|}. \tag{2}$$

TABLE I: Summary of Key Notations

Notation	Description
A, R, N	Set of anchor nodes, RISs, and targets, respectively
$(x_a, y_a, z_a), \forall a \in A$	Coordinates of anchor nodes
$(x_r, y_r, z_r), \forall r \in R$	Coordinates of RISs
$(x_n, y_n, z_n), \forall n \in N$	Coordinates of targets
$d_{a,n}$ [m]	Pseudorange among anchor node a and target n
R_n	Set of RISs selected by target n
E_n	Target's n consumed energy
B[bits]	Beacon signal's amount of information
E_b[J/bit]	Unit energy for receiving one bit of beacon signal
$d_{r,n}$ [m]	Pseudorange among RIS r and target n
P_s^n	Target's n shadowing probability
e_n[J]	Target's n energy constraint
p_s^n	Target's n shadowing probability constraint
$\mathbf{P}_n$	Target's n position and timing
M	Set of RIS elements
$\|N_r\|$	Number of targets selected RIS r
U_n	Target's n utility
f_n	Target's n satisfaction correspondence
$\mathbf{s}^+$	Satisfaction Equilibrium
ite	Iteration of the DRL algorithm
P_n^{ite}	Target's n selection probability of a strategy
λ	Leraning parameter of DRL algorithm
$\mathbf{s}^-$	Generalized Satisfaction Equilibrium
$\mathbf{s}^\dagger$	Efficient Satisfaction Equilibrium
s_n^c	Target's n clipping strategy
$g(R_n)$	Reward function
i	Iteration of RL-based RIS selection algorithm
$\mathcal{P}_n$	Target's n selection probability of a set of RISs
M_n^r	Set of RIS's r elements allocated to target n
ω_m	Phase shift of RIS's element m
$h_{a,n}$	Channel gain in the direct communication link
$PL_{a,n}$	Path loss in the communication link
$\tilde{h}$	Random variable capturing scattering effects
P^{LoS}, P^{NLoS}	Probabilities of LoS and NLoS communication
PL^{LoS},PL^{NLoS}	Path losses for the LoS and NLoS communication links
θ	Elevation angle
f_c [Hz]	Carrier frequency
δ	Path loss exponent
c [m/s]	Speed of light
$\mathbf{h}_{a,r}$	Channel gain in the communication link of a, r
ζ	Path loss at the reference distance 1 m
$\phi_{a,r}$	Signal's angle of arrival
d_s [m]	Antenna separation
λ [m]	Wavelength of the carrier signal
$\mathbf{h}_{r,n}$	Channel gain in the communication link of r, n
$\phi_{r,n}$	Signal's angle of departure

Last, each target is characterized by a personal energy constraint e_n [J] based on its device's physical properties, and a personal shadowing probability constraint p_s^n in order to be able to measure the corresponding pseudoranges. The key notations used in this paper are summarized in Table I.

B. Operation of the Proposed Framework

In this section, an overview of the proposed alternative PNT solution is provided, while describing the information and control flow among the individual building components of the overall framework. The operation of the proposed framework is presented in Fig. 2. Initially, each target estimates its position and timing $\mathbf{P}_n^{\tau=0} = (x_n^{\tau=0}, y_n^{\tau=0}, z_n^{\tau=0}, \Delta t_n^{\tau=0})$ by measuring the pseudoranges from the anchor nodes only, given that the target has still not selected the optimal number and set of RISs that will contribute to its PNT service subsequently. The anchor nodes perform an one-time broadcasting of their beacon signals in order for the targets to determine an initial estimation of their position and timing $\mathbf{P}_n^{\tau=0}$. The iterations of the overall PNT solution are denoted as τ. Then, the

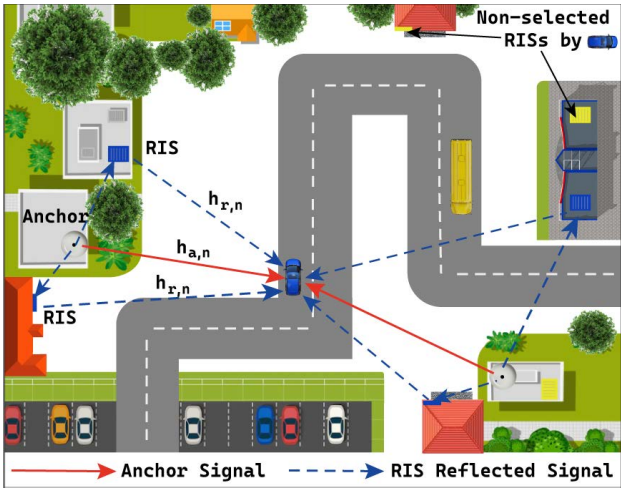

Fig. 1: Overview of the RIS-assisted system consisting of targets and anchor nodes.

service, each target leverages at the same time the reflected signals from a set of properly selected RISs [20], [21]. For this purpose, each target first selects the optimal number of RISs $|R_n|$ (see Section III) and identifies the specific set of RISs $R_n = \{1, \ldots, r, \ldots, |R_n|\}$ that will contribute to its PNT service (see Section IV). Specifically, each anchor node transmits a beacon signal with constant transmission power in order to support the targets' PNT services. The target receives the direct signals from the anchor nodes, as well as the reflected signals from the selected RISs, while treating the rest of the received signals as noise. Then, the measurement of the pseudoranges $d_{r,n}$ [m] from each selected RIS $r \in R_n$ follows, as described in Section VI [22]. An overview of the considered RIS-assisted network consisting of the targets and the anchor nodes is presented in Fig. 1.

As a result, each target receives beacon signals from $|A|$ anchor nodes and $|R_n|$ selected RISs, each of them consisting of B [bits] of information. The overall consumed energy $E_n(|R_n|)$ [J] by each target to decode the signals and measure the corresponding pseudoranges $d_{a,n}$ (from each anchor node $a \in A$) and $d_{r,n}$ [m] (from each selected RIS $r \in R_n$), is given as follows [23],

$$E_n(|R_n|) = (|A| \cdot B + |R_n| \cdot B) \cdot E_b, \tag{1}$$

where E_b [J/bit] is the target's unit energy consumed for receiving one bit of a beacon signal and calculating the respective pseudorange.

The targets experience a shadowing effect due to the physical obstacles in the surrounding environment, which can contribute to non-Line of Sight (NLoS) communication paths between the anchor nodes and the selected RISs with respect to the target. Apparently, as the number of selected RISs by each target increases, the probability of creating a constructive beam of Line of Sight (LoS) communication between the anchor nodes and the selected RISs also increases, mitigating practically the experienced shadowing effect. Based on the above observation, we define the shadowing probability for a target $n, \forall n \in N$, as follows [24],

$$P_s^n(|R_n|) = 1 - \frac{|A| + |R_n|}{|A| + |R|}. \tag{2}$$

TABLE I: Summary of Key Notations

Notation	Description
A, R, N	Set of anchor nodes, RISs, and targets, respectively
$(x_a, y_a, z_a), \forall a \in A$	Coordinates of anchor nodes
$(x_r, y_r, z_r), \forall r \in R$	Coordinates of RISs
$(x_n, y_n, z_n), \forall n \in N$	Coordinates of targets
$d_{a,n}$ [m]	Pseudorange among anchor node a and target n
R_n	Set of RISs selected by target n
E_n	Target's n consumed energy
B[bits]	Beacon signal's amount of information
E_b[J/bit]	Unit energy for receiving one bit of beacon signal
$d_{r,n}$ [m]	Pseudorange among RIS r and target n
P_s^n	Target's n shadowing probability
e_n[J]	Target's n energy constraint
p_s^n	Target's n shadowing probability constraint
$\mathbf{P}_n$	Target's n position and timing
M	Set of RIS elements
$\|N_r\|$	Number of targets selected RIS r
U_n	Target's n utility
f_n	Target's n satisfaction correspondence
$\mathbf{s}^+$	Satisfaction Equilibrium
ite	Iteration of the DRL algorithm
P_n^{ite}	Target's n selection probability of a strategy
λ	Leraning parameter of DRL algorithm
$\mathbf{s}^-$	Generalized Satisfaction Equilibrium
$\mathbf{s}^\dagger$	Efficient Satisfaction Equilibrium
s_n^c	Target's n clipping strategy
$g(R_n)$	Reward function
i	Iteration of RL-based RIS selection algorithm
$\mathcal{P}_n$	Target's n selection probability of a set of RISs
M_n^r	Set of RIS's r elements allocated to target n
ω_m	Phase shift of RIS's element m
$h_{a,n}$	Channel gain in the direct communication link
$PL_{a,n}$	Path loss in the communication link
$\tilde{h}$	Random variable capturing scattering effects
P^{LoS}, P^{NLoS}	Probabilities of LoS and NLoS communication
PL^{LoS},PL^{NLoS}	Path losses for the LoS and NLoS communication links
θ	Elevation angle
f_c [Hz]	Carrier frequency
δ	Path loss exponent
c [m/s]	Speed of light
$\mathbf{h}_{a,r}$	Channel gain in the communication link of a, r
ζ	Path loss at the reference distance 1 m
$\phi_{a,r}$	Signal's angle of arrival
d_s [m]	Antenna separation
λ [m]	Wavelength of the carrier signal
$\mathbf{h}_{r,n}$	Channel gain in the communication link of r, n
$\phi_{r,n}$	Signal's angle of departure

Last, each target is characterized by a personal energy constraint e_n [J] based on its device's physical properties, and a personal shadowing probability constraint p_s^n in order to be able to measure the corresponding pseudoranges. The key notations used in this paper are summarized in Table I.

B. Operation of the Proposed Framework

In this section, an overview of the proposed alternative PNT solution is provided, while describing the information and control flow among the individual building components of the overall framework. The operation of the proposed framework is presented in Fig. 2. Initially, each target estimates its position and timing $\mathbf{P}_n^{\tau=0} = (x_n^{\tau=0}, y_n^{\tau=0}, z_n^{\tau=0}, \Delta t_n^{\tau=0})$ by measuring the pseudoranges from the anchor nodes only, given that the target has still not selected the optimal number and set of RISs that will contribute to its PNT service subsequently. The anchor nodes perform an one-time broadcasting of their beacon signals in order for the targets to determine an initial estimation of their position and timing $\mathbf{P}_n^{\tau=0}$. The iterations of the overall PNT solution are denoted as τ. Then, the

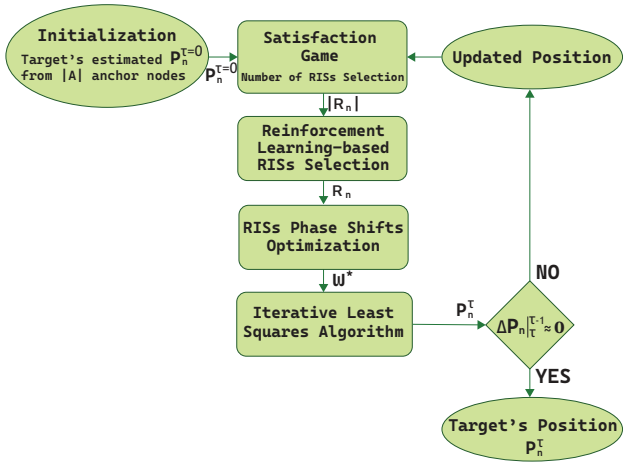

Fig. 2: Overview of the proposed alternative PNT solution's operation.

targets participate in a satisfaction game among them to determine the optimal number of RISs that will be used to improve the accuracy of their PNT services, while considering their personal energy and shadowing probability constraints (Section III). Following this step, a novel RL-based approach is proposed following the principles of the gradient ascent algorithms to enable each target to select the specific set of RISs that will contribute to its PNT service (Section IV), given the number of RISs as determined from the previous step (Section III). Afterward, the RISs phase shift optimization is performed to improve the signal strength of the signals received by the targets (Section V). Then, each target sends a control signal to its selected RISs' controllers to optimally control its selected RISs elements phase shifts, and the anchor nodes perform another broadcasting in order for the targets to receive the updated stronger received signals. Finally, the Iterative Least Squares (ILS) algorithm is implemented to determine the targets' position and timing (Section VI). The overall process is repeated iteratively, as presented in Fig. 2, until the position and timing estimation error of the targets is practically eliminated. It is noted that in the practical implementation of the proposed model, only few iterations of the overall process are needed given the intelligent exploitation of the beacon signals stemming both from the anchor nodes and the RISs.

III. Optimal Number of RISs via Games in Satisfaction Form

In this section, the theory of Games in Satisfaction Form is adopted to enable each target to autonomously identify and select the optimal number of RISs that will contribute to its PNT service. Each RIS consists of a set of RIS elements $M = \{1, \ldots, m, \ldots, |M|\}$. The targets that select the same RIS, equally share among each other the control over its elements, by being allocated with $\lfloor \frac{|M|}{|N_r|} \rfloor$ number of RIS elements, where $|N_r|$ denotes the number of targets that selected the RIS r. The goal of each target is initially to select the optimal number of RISs $|R_n|$ that satisfies its energy e_n and shadowing probability p_s^n constraints, while at the same time not over-exploiting the RISs elements. The latter condition is desired, since the more targets share the same RIS, the less number of elements each target is allocated, decreasing in this way their capability to create a constructive beam via controlling the phase shift of their allocated elements. Also, it is noted that the creation of few constructive beams based on the anchor nodes' reflected signals on the RISs can substantially mitigate the interference experienced in the overall system, as compared to the case of multi-path signals due to scattering, diffraction, etc., in a non-RIS supported wireless communication environment. Based on the above, we define the target's n satisfaction, i.e., utility, from selecting a number of RISs $|R_n|$, as follows,

$$U_n(s_n, \mathbf{s}_{-n}) = \begin{cases} (\frac{e_n - E_n}{e_n})(\frac{p_s^n - P_s^n}{p_s^n})(\frac{|R| \cdot |N|}{\sum_{\forall n \in N} |R_n|}), \\ \qquad \text{if } E_n < e_n \text{ and } P_s^n < p_s^n, \\ 0, \qquad \text{otherwise}, \end{cases} \tag{3}$$

where s_n is the strategy, i.e., the number of RISs $|R_n|$ selected by the target n, and $\mathbf{s}_{-n} = (s_1, \ldots, s_{n-1}, s_{n+1}, \ldots, s_{|N|})$ is the strategy vector of the rest of the targets. The physical meaning of Eq. 3 is that a target n receives a positive utility if its energy consumption and shadowing probability constraints are simultaneously satisfied, while being at the same time incentivized to not over-exploit the RISs elements.

Towards determining the optimal number of RISs selected by each target, we introduce a non-cooperative satisfaction game among the targets that is formally described as $G = [N, \{s_n\}_{\forall n \in N}, \{U_n\}_{\forall n \in N}, \{f_n\}_{\forall n \in N}]$, where N is the set of targets, s_n is each target's strategy, i.e., number of RISs, U_n is the target's utility (Eq. 3), and $f_n(\mathbf{s_{-n}}) = \{s_n | U_n(s_n, \mathbf{s}_{-n}) \geqslant 0\}$ is the target's satisfaction correspondence.

Definition 1. ***(Satisfaction Equilibrium (SE)):*** *A strategy vector* $\mathbf{s}^+ = (s_1^+, \ldots, s_n^+, \ldots, s_{|N|}^+)$ *is a SE, if* $\forall n \in N$, $s_n^+ \in f_n(\mathbf{s}_{-n}^+)$.

Practically, at the SE point $\mathbf{s}^+$, all targets achieve a non-negative utility, implying that they satisfy their personal constraints, as these have incorporated in the definition of their utility above.

Towards determining an SE point, we introduce the Distributed Reinforcement Learning algorithm (DRL), as presented in Algorithm 1. The DRL algorithm returns the SE strategy vector $\mathbf{s}^+ = (s_1^+, \ldots, s_n^+, \ldots, s_{|N|}^+)$, if it exists. Also, $\lambda \in [0, 1]$ is the learning parameter. The complexity of the DRL algorithm is $O(Ite)$, where Ite is the number of iterations required for the algorithm to converge to the SE. Detailed numerical results are presented in Fig. 4 and Fig. 7 in Section VII, showing that the number of iterations Ite are approximately 10 in order for the algorithm to converge.

Definition 2. ***(Clipping Strategy):*** *A target has a clipping strategy* s_n^c, *iff* $\forall \mathbf{s}_{-n}$, $s_n^c \in f_n(\mathbf{s}_{-n})$.

The physical meaning and interpretation of Definition 2 is that if a target n achieves a clipping strategy at an iteration ite of the DRL algorithm, then, for any consequent iteration, the target keeps the same strategy, i.e., selects the same number of RISs. Thus, if there is another target $n' \neq n$, such that $f_{n'}(\mathbf{s}_{-n'}) = \varnothing$ (i.e., the target is not satisfied), then, the target

Algorithm 1 Distributed Reinforcement Learning (DRL) algorithm

Input: $A, R, N, e_n, p_s^n, \forall n \in N$
Output: s^+
Initialization: $ite = 0$, $Convergence = 0$, $s_n^{ite=0}$, $\forall n \in N$, $P_n^{ite=0} = \frac{1}{|R|}$.
while *Convergence* $== 0$ **do**
  $ite = ite + 1$;
  if $s_n^{ite+1} = s_n^{ite}$ **then**
    Each node n selects $\mathbf{s}_n^{ite+1}$ with probability $P_n^{ite+1}(s_n^{ite+1}) = P_n^{ite}(s_n^{ite}) - \frac{1}{(ite+1)+1} \cdot \lambda \cdot P_n^{ite}(s_n^{ite})$.
  end if
  if $s_n^{ite+1} \neq s_n^{ite}$ **then**
    Each node n selects $\mathbf{s}_n^{ite+1}$ with probability $P_n^{ite+1}(s_n^{ite+1}) = P_n^{ite}(s_n^{ite}) + \frac{1}{(ite+1)+1} \cdot \lambda \cdot (\frac{1}{|R|-1} - P_n^{ite}(s_n^{ite}))$.
  end if
  if $U_n(s_n^{ite+1}, \mathbf{s}_{-n}^{ite+1}) \geqslant 0$ **then**
    $Convergence = 1$
  end if
end while

n will play its clipping strategy s_n^c and the target n' will not achieve the satisfaction of its personal constraints. Thus, the DRL algorithm will converge to a Generalized Satisfaction Equilibrium (GSE) point.

Definition 3. ***(Generalized Satisfaction Equilibrium):*** *A strategy vector $\mathbf{s}^-$ is a GSE if there exist two sets of targets N_{sat}, N_{unsat} with $N_{sat} \cup N_{unsat} = N$, such that $\forall n \in N_{sat}, s_n \in f_n(\mathbf{s}_{-n}^-)$ and $n' \in N_{unsat}, f_{n'}(\mathbf{s}_{-n'}^-) = \varnothing$.*

It should be highlighted that the non-cooperative satisfaction game G may have many GSE or many SE points that satisfy the targets' constraints. Thus, an GSE exists and can be determined by the DRL algorithm, but it is not necessarily unique, as several strategies $\mathbf{s}^-$ may satisfy the targets belonging in the set of satisfied targets N_{sat}. Out of all the potential SE points, the one SE that results in the lowest energy consumption $E_n(|R_n|)$ for the target, i.e., the Efficient Satisfaction Equilibrium (ESE) point, presents greater energy benefits for the target.

Definition 4. ***(Efficient Satisfaction Equilibrium (ESE)):*** *A strategy vector $\mathbf{s}^\dagger$ is an ESE, if $\forall n \in N, s^\dagger \in f_n(\mathbf{s}_{-n}^\dagger)$ and $E_n(s_n^\dagger) < E_n(s_n^{'}), \forall n \in N$, for any other possible strategy $s_n^{'}$.*

Practically, at the ESE point $\mathbf{s}^\dagger$, the targets achieve to not only satisfy their personal energy and shadowing probability constraints, but are also experiencing the lowest possible energy cost. In the general case, the ESE is not unique, as multiple strategies of number of RISs selection may provide the same lowest possible energy cost to the target. Towards determining the ESE of the formulated game, we utilize a similar methodology to the DLR algorithm presented earlier in this section, however, the strategy selection is performed under two joint criteria, i.e., non-negative utility and minimum energy consumption cost. The outcome of the algorithm determines the ESE, i.e., the optimal number of RISs $|R_n|$ that should be selected by each target in order to determine its position and timing, while considering its personal energy and shadowing probability constraints.

IV. Reinforcement Learning-based RIS Selection

In this section, a novel Reinforcement Learning (RL) algorithm is introduced to enable each target to autonomously select the specific set of RISs that will contribute to its PNT service, given the optimal number of RISs, as determined by the ESE (Section III). The metric that is utilized at this point of the proposed alternative PNT solution design is the Geometric Dilution of Precision (GDOP), which captures the accuracy of the PNT service. Particularly in this work, the GDOP value quantifies the success of the geometric constellation of the anchor nodes and the selected RISs in terms of accurately determining the target's position and timing. Low values of GDOP correspond to low estimation error of the target's position and timing, and thus, the ultimate goal of each target is to achieve a small GDOP value.

Based on the latter, we define each target's n reward by selecting a specific set of RISs R_n at the i^{th} iteration of the proposed RL algorithm, as follows,

$$g(R_n) = \frac{\sqrt{\frac{10}{|A|+|R_n|}}}{GDOP(R_n)}. \tag{4}$$

It is noted that the value $\sqrt{\frac{10}{|A|+|R_n|}}$ is the best GDOP value that is currently reported in the existing literature considering the received signals from $|A| + |R_n|$ transmitters, i.e., the anchor nodes and the selected RISs [25].

The proposed reward function is practically used to steer each target's n autonomous selection of the specific set of RISs R_n, via the implementation of a Linear Reward Inaction (LRI) algorithm-based RL scheme [26]. In particular, each target probabilistically selects a set of RISs R_n^i at the i^{th} iteration of the RL algorithm, and subsequently, updates the probabilities of selecting this specific set at future iterations, by evaluating the resulted value of reward $g(R_n^i)$. The probability of each target selecting the same (Eq. 5a) or a different (Eq. 5b) set of RISs R_n at each iteration i of the RL algorithm is given as follows,

$$\mathcal{P}_n(R_n^{i+1}) = \mathcal{P}_n(R_n^i) + \lambda_{RL} \cdot g(R_n^i)(1 - \mathcal{P}_n(R_n^i)), \text{if } R_n^i = R_n^{i+1}, \tag{5a}$$

$$\mathcal{P}_n(R_n^{i+1}) = \mathcal{P}_n(R_n^i) - \lambda_{RL} \cdot g(R_n^i)\mathcal{P}_n(R_n^i), \text{if } R_n^i \neq R_n^{i+1}, \tag{5b}$$

where $\lambda_{RL} \in [0, 1]$ is the target's learning rate. The proposed RL algorithm converges slower to a set of selected RISs for smaller values of the learning parameter λ_{RL}, with the benefit of more accurate decision-making, given that it thoroughly explores the available strategies. The devised RL algorithm is repeated iteratively by each target until the probability to select a specific set of RISs gets close to one. Also, it is highlighted that the strategic RISs selection by the targets accounting for the highest achieved GDOP, as captured in Eq. 4, has the potential of mitigating the interference in the overall system, as the targets exploit the existing strongest received signals

reflected on the RISs, without requesting higher transmission power levels.

V. RISs Phase Shift Optimization

In this section, the RISs' elements phase shift optimization is performed towards maximizing the received signal strength at each target of the reflected signals from the specific set of RISs that has been determined to accommodate its PNT service. As mentioned in Section III and IV, each target selects the optimal number of RISs $|R_n|$ and identifies the specific set of RISs R_n that will contribute to its PNT service. A single RIS r can be selected by multiple targets, the number of which is indicated by $|N_r|$. Each target n out of the $|N_r|$ has equal control over the RIS's r elements, by being allocated a number of $|M_n^r| = \lfloor \frac{|M|}{|N_r|} \rfloor$ elements, the set of which is indicated as M_n^r. Hence, the goal of each target is to optimally control the phase shift of the RIS elements that it has control over at each RIS. Before formulating the corresponding optimization problem, we analyze in detail the communication environment among the anchor nodes, the RISs, and the targets. The rest of the analysis is focused on one target and one of its selected RISs r, while targeting to optimize the phase shift of the corresponding $|M_n^r|$ allocated RIS elements. Similar analysis can be derived for each target, each selected RIS, and each group of $|M_n^r|$ allocated RIS elements.

Let $\omega_m \in [0, 2\pi], \forall m \in M_n^r$ denote the phase shift of RIS's r elements, and $\mathbf{\Omega} = \text{diag}(e^{j\omega_1}, \ldots, e^{j\omega_{|M_n^r|}})$, $\mathbf{\Omega} \in \mathbb{C}^{|M_n^r| \times |M_n^r|}$ be the corresponding diagonal reflection matrix [27]. We denote as $h_{a,n} = \sqrt{\frac{1}{PL_{a,n}}} \cdot \tilde{h}$, the channel gain experienced in the direct communication link between the anchor node a and the target n, with $\tilde{h}$ representing a random variable quantifying the random scattering effects and following a zero-mean unit-variance Complex Gaussian distribution. $PL_{a,n} = PL(d_{a,n})$ is the overall path loss of the corresponding communication link, with $d_{a,n}$ [m] denoting the Euclidean distance among the anchor node and the target. Especially, the path loss $PL_{a,n}$ is derived as $PL(d_{a,n}) = P^{LoS} \cdot PL^{LoS} + (1 - P^{LoS}) PL^{NLoS}$, where P^{LoS} and P^{NLoS} are the probabilities of LoS and NLoS communication, respectively, and PL^{LoS} and PL^{NLoS} are the path losses for the LoS and NLoS communication link, respectively. The formulas from which the latter are derived are as follows. We define $P^{LoS}(z_n, d_{a,n}) = \frac{1}{1+\alpha e^{-\gamma(\theta-\alpha)}}$, where $\theta = \frac{180}{\pi} \sin^{-1}(\frac{z_n}{d_{a,n}})$ [rad] denotes the elevation angle, z_n [m] is the target's altitude, and $\alpha, \gamma \in \mathbb{R}^+$ are positive constants depending on the carrier frequency f_c [Hz] and the communication environment, e.g., urban, rural. Also, we define $PL^{LoS}(d_{a,n}) = \eta^{LoS}(\frac{4\pi f_c d_{a,n}}{c})^{\delta}$ and $PL^{NLoS}(d_{a,n}) = \eta^{NLoS}(\frac{4\pi f_c d_{a,n}}{c})^{\delta}$, with η^{LoS}, η^{NLoS} denoting the excessive path loss coefficients, $\eta^{NLoS} > \eta^{LoS} > 1$, c [m/s] is the speed of light, and δ is the path loss exponent.

Focusing on the communication link between an anchor node a and a RIS r, the channel gain is derived as $\mathbf{h}_{a,r} = \sqrt{\frac{1}{PL_{a,r}}} \cdot [1, e^{-j\frac{2\pi}{\lambda} d_s \phi_{a,r}}, \ldots, e^{-j\frac{2\pi}{\lambda}(|M_n^r|-1) d_s \phi_{a,r}}]^T$, $\mathbf{h}_{a,r} \in \mathbb{C}^{|M_n^r| \times 1}$, with $PL_{a,r} = \zeta \cdot d_{a,r}^{\kappa}$ denoting the path loss of the communication link among the anchor node a and the RIS r, ζ [dB] is the path loss at the reference distance of 1 [m], $d_{a,r}$ [m] is the distance between the anchor node and the RIS, and κ is the path loss exponent. Also, $\phi_{a,r}$ denotes the signal's angle of arrival, d_s [m] is the antenna separation, and λ [m] is the wavelength of the carrier signal.

Concerning the communication link between the RIS r and the target n, the corresponding channel gain is derived as $\mathbf{h}_{r,n} = \sqrt{\frac{1}{PL_{r,n}}}(\sqrt{\frac{\mu}{1+\mu}} \mathbf{h}_{r,n}^{LoS} + \sqrt{\frac{1}{1+\mu}} \mathbf{h}_{r,n}^{NLoS}) \in \mathbb{C}^{|M_r^n| \times 1}$, where $\mathbf{h}_{r,n}^{LoS} = [1, e^{-j\frac{2\pi}{\lambda} d_s \phi_{r,n}}, \ldots, e^{-j\frac{2\pi}{\lambda}(|M_n^r|-1) d_s \phi_{r,n}}]^T$ is the LoS component, with $\phi_{r,n}$ denoting the signal's angle of departure, and $\mathbf{h}_{r,n}^{NLoS}$ is the NLoS component, represented by a random variable following a zero-mean unit-variance Complex Gaussian distribution. μ is the Rician factor, and $PL_{r,n} = PL(d_{r,n})$ is the overall path loss of the communication link, with $d_{r,n}$ [m] denoting the distance among the RIS r and the target n.

For each selected RIS, the goal of the target is to maximize the received signal strength, by optimizing the phase shifts of the RIS's elements that are allocated to the target. Based on the previous analysis and the modeling of the wireless channels' conditions within the considered communication environment, the overall channel gain between a target n and an anchor node a is $G_n = |h_{a,n} + \mathbf{h}_{r,n}^H \mathbf{\Omega} \mathbf{h}_{a,r}|^2$, while considering at the same time the reflection of the anchor node's transmitted signal from the RIS to the target. Thus, towards determining the optimal phase shift of the allocated RIS elements to a target, each target should solve the following optimization problem for each of its selected RISs:

$$\max_{\boldsymbol{\omega}} \sum_{a=1}^{|A|} |h_{a,n} + \mathbf{h}_{r,n}^H \mathbf{\Omega} \mathbf{h}_{a,r}|^2 \tag{6a}$$

$$\text{s.t.,} \ \ 0 \leqslant \omega_m < 2\pi, \forall m \in M_n^r, \tag{6b}$$

where $\boldsymbol{\omega}$ is the vector of the RIS's phase shifts $\omega_m, \forall m \in M_n^r$.

We rewrite the above optimization problem as follows:

$$\max_{\mathbf{v}} \sum_{a=1}^{|A|} |h_{a,n} + \hat{\mathbf{h}}_{a,r}^H \mathbf{v}|^2 \tag{7a}$$

$$\text{s.t.,} \ \ \sum_{m=1}^{M_n^r} |v_m| = 1, \forall m \in M_n^r, \tag{7b}$$

where $\mathbf{v} = [v_1, \ldots, v_m, \ldots, v_{|M_n^r|}] \in \mathbb{C}^{|M_n^r| \times 1}$, $v_m = e^{j\omega_m}$, and $\hat{\mathbf{h}}_{a,r}^H = \mathbf{h}_{r,n}^H \text{diag}(\mathbf{h}_{a,r}) \in \mathbb{C}^{1 \times |M_n^r|}$. The optimization problem (7a) – (7b) is non-convex, thus, in the following analysis, we introduce a heuristic approach in order to determine the optimal phase shift of the $|M_n^r|$ RIS's elements allocated to the target n.

We start our analysis with one anchor node and a target, and we generalize for multiple anchor nodes. In the case of one anchor node (e.g., $a = 1$), the direct and the reflected signal from the RIS should be aligned to maximize the received signal strength. Thus, it should hold true that $\angle h_{1,n} = -\angle \hat{\mathbf{h}}_{1,r} + \angle \mathbf{v} \iff \angle \mathbf{v} = \angle h_{1,n} + \angle \hat{\mathbf{h}}_{1,r}$, and the corresponding optimal phase shifts of the RIS's elements can be derived $\boldsymbol{\omega}^* = \angle \mathbf{v}$. Generalizing our analysis for multiple anchor nodes, we denote as $\mathbf{v}_a = [v_{a,1}, \ldots, v_{a,m}, \ldots, v_{a,|M_n^r|}] \in \mathbb{C}^{|M_n^r| \times 1}$ the reflection coefficient vector for each anchor node, where $\mathbf{v}_a = e^{j\angle h_{a,n}}$.

$$J^k = \begin{bmatrix} \frac{\partial f_1(\mathbf{X}_n^k)}{\partial x_n} & \frac{\partial f_1(\mathbf{X}_n^k)}{\partial y_n} & \frac{\partial f_1(\mathbf{X}_n^k)}{\partial z_n} & \frac{\partial f_1(\mathbf{X}_n^k)}{\partial \Delta t_n^k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{|A|+|R_n|}(\mathbf{X}_n^k)}{\partial x_n} & \frac{\partial f_{|A|+|R_n|}(\mathbf{X}_n^k)}{\partial y_n} & \frac{\partial f_{|A|+|R_n|}(\mathbf{X}_n^k)}{\partial z_n} & \frac{\partial f_{|A|+|R_n|}(\mathbf{X}_n^k)}{\partial \Delta t_n^k} \end{bmatrix} \quad (9) \qquad RES^k = \begin{bmatrix} f_1(\mathbf{X}_n^k) \\ \vdots \\ f_{|A|+|R_n|}(\mathbf{X}_n^k) \end{bmatrix} \quad (10)$$

$e^{j\angle \hat{\mathbf{h}}_{a,r}}$, $\forall a \in A$. However, it is evident that a single value $\mathbf{v}$ that maximizes the channel gain for all the anchor nodes simultaneously cannot be determined. Thus, we introduce a weight $w_a \in [0,1]$ to each reflection coefficient vector $\mathbf{v}_a$, with $\mathbf{v} = \sum_{a=1}^{|A|} w_a \cdot \mathbf{v}_a$, such that $\sum_{a=1}^{|A|} w_a = 1$, and we rewrite the optimization problem (7a) – (7b) as follows:

$$\max_{\mathbf{w}} \sum_{a=1}^{|A|} |h_{a,n} + \hat{\mathbf{h}}_{r,n}^H \mathbf{\Omega} \mathbf{h}_{a,r}|^2 \quad (8a)$$

$$\text{s.t.,} \;\; 0 \leqslant w_a < 1, \forall a \in A, \quad (8b)$$

$$\sum_{a=1}^{|A|} w_a = 1, \quad (8c)$$

with $\mathbf{w} = [w_1, \ldots, w_a, \ldots, w_{|A|}]$ and $\mathbf{\Omega} = \text{diag}(e^{j\angle \mathbf{v}})$. The latter optimization problem can be solved based on standard optimization tools, and derive the optimal reflection coefficient vector $\mathbf{v}^*$. By backward induction, the target's allocated RIS elements' optimal phase shift vector $\boldsymbol{\omega}^*$ can then be derived. Each target solves the optimization problem (8a)-(8c) for each selected RIS $r \in R_n$ and sends a control signal to the corresponding RIS controller in order to optimally control the phase shifts of the $|M_n^r|$ elements.

VI. Positioning, Navigation, and Timing Estimation

In this section, we introduce a low-complexity Iterative Least Squares (ILS) algorithm to estimate the targets' position and timing, given the derivation of the optimal number of RISs (Section III), the selection of the specific set of RISs by each target (Section IV), as well as the RISs' element phase shift optimization (Section V). Each target measures the pseudoranges from each anchor node $d_{a,n}^k = |\mathbf{X}_n^k - \mathbf{X}_a| - \Delta t_n^k \cdot c$, and each selected RIS $d_{a,r,n}^k = \mathbf{1}_{a,r}(\mathbf{X}_r - \mathbf{X}_a) + \mathbf{1}_r(\mathbf{X}_n^k - \mathbf{X}_r) - \Delta t_n^k \cdot c$, taking respective into account the direct and the reflected signals from the RISs. We denote as k the iteration of the ILS algorithm, $\mathbf{X}_n^k = (x_n^k, y_n^k, z_n^k)$, $\mathbf{X}_a = (x_a, y_a, z_a)$, $\mathbf{X}_r = (x_r, y_r, z_r)$ the coordinates of the target n, the anchor node a, and the RIS r, respectively. Also, Δt_n^k denotes the clock offset among the target n and the anchor nodes, assuming that the clocks of all the anchor nodes are synchronized. The main steps of the proposed ILS algorithm are summarized in Algorithm 2. The complexity of the ILS algorithm is $O(N \cdot K)$, where K is the number of iterations that the algorithm needs in order to converge. It is highlighted that in a practical implementation, the ILS algorithm is executed by each target, thus, the complexity is $O(K)$, experienced by each target.

Based on the above analysis, the ILS algorithm determines the targets' position and timing while exploiting the input of the Satisfaction Game, i.e., optimal number of selected RISs, the RL-based RIS selection approach, i.e., selection of a specific set of RISs, and the RISs phase shift optimization that maximizes the signal strength of the received signals. The goal of the ILS algorithm is to eliminate the estimation error.

Algorithm 2 Iterative Least Squares (ILS) Algorithm

1: **Input:** $d_{a,n}^k, d_{a,r,n}^k, \forall a \in A, r \in 1, \ldots |R_n|$
2: **Output:** $\mathbf{P}_n, \forall n \in N$
3: **Initialization:** $k = 0$, $Convergence = 0$,$\mathbf{P}_n^{k=0}, \forall n \in N$.
4: **while** *Convergence* $== 0$ **do**
5: $k = k + 1$;
6: **for** $n = 1$ **to** N **do**
7: The equations of the pseudoranges are set equal to zero, and the corresponding functions $f_j, \forall j \in A \cup \{1, \ldots, |R_n|\}$ are derived.
8: The Jacobian matrix J^k and the Residual matrix RES^k are calculated, as presented in Eq. 9 and Eq. 10, respectively.
9: The least squares problem is solved and the position and timing estimation error is calculated as $(\Delta x_n^k, \Delta y_n^k, \Delta z_n^k, \Delta(\Delta t_n^k)) = (J^{k^T} \cdot J^k)^{-1} \cdot J^{k^T} \cdot RES^k$.
10: The position and timing estimate is updated, as follows: $\mathbf{P}_n^{k+1} = (x_n^k + \Delta x_n^k, y_n^k + \Delta y_n^k, z_n^k + \Delta z_n^k, \Delta t_n^k + \Delta(\Delta t_n^k))$.
11: **end for**
12: **if** $|\mathbf{P}_n^k - \mathbf{P}_n^{k+1}| \approx 0, \forall n \in N$ **then**
13: $Convergence = 1$
14: **end if**
15: **end while**

VII. Numerical Results

In this section, a detailed simulation-based evaluation of the proposed framework is performed in order to demonstrate its pure operation and performance, as well as to highlight its benefits and drawbacks. Specifically, the pure operation and performance of the proposed alternative PNT solution is presented in Section VII-A, while its benefits are highlighted in Section VII-B through a comparative evaluation. The benefits of the proposed satisfaction game-theoretic approach in terms of selecting the number of RISs that contribute to the targets PNT service are quantified in Section VII-C. The performance of the proposed alternative PNT solution under the mobility scenario of the targets is illustrated in Section VII-D, while a detailed comparative evaluation to the state of the art is provided in Section VII-E.

Throughout our evaluation, we adopted a system topology consisting of 4 anchor nodes, 10 RISs and 5 targets. In particular the key parameters that have been used in the

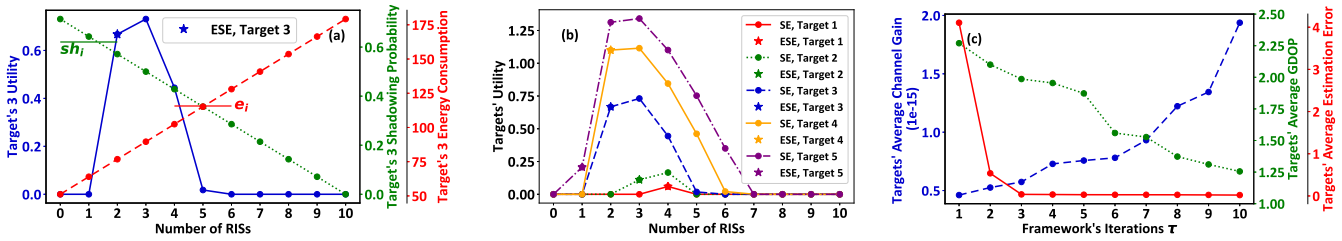

Fig. 3: Pure performance and operation of the proposed alternative PNT solution.

experiments are: $|A| = 4$, $|R| = 10$, $|N| = 5$, $B = 32$ bits, $E_b = 0.4$ J, $\mathbf{e} = [107.48, 110.93, 115.68, 128.58, 138.41]$ J, $\mathbf{p}_s = [0.48, 0.52, 0.62, 0.64, 0.65]$, $|M| = 300$, $\lambda = 0.01$, $\alpha = 11.95$, $\gamma = 0.14$, $f_c = 2 \times 10^9$ Hz, $\eta^{LoS} = 3$, $\eta^{NLoS} = 23$, $c = 3 \times 10^8$ m/s, $\delta = 2$, $d_s = \frac{\lambda}{2}$, $\mu = 2.8$, unless otherwise explicitly stated. For presentation purposes, it has been considered that the higher the target's ID is, the greater its energy consumption and shadowing probability constraints are. In this way, the effect of the stringency of the targets' constraints over the proposed framework's performance can be scrutinized and revealed. A detailed Monte Carlo analysis has been performed for all the presented experiments considering $10,000$ realizations of the RISs phase shift optimization problem. The evaluation of the proposed framework was performed in a Dell Tower Desktop with Intel i7 11700k 3.6GHz processor, 32GB available RAM.

A. Pure Operation & Performance

In this section, the pure performance and the operation of the proposed alternative PNT solution is demonstrated. Fig. 3a presents the utility (Eq. 3), the shadowing probability (Eq. 2), and the energy consumption (Eq. 1) of one indicative target (target's ID 3) as a function of the number of selected RISs. Fig. 3b shows all the targets' utility as a function of the number of selected RISs. Fig. 3c illustrates the targets' average channel gain, GDOP, and position and timing estimation error as a function of the iterations τ of the overall framework, as presented in Fig. 2.

The obtained results verify that the novel concept of satisfaction games enables the targets to achieve a feasible solution in terms of selecting the optimal number of RISs that satisfies their energy consumption and shadowing probability constraints (Fig. 3a), by securing a positive utility value (Fig. 3a,3b). Also, it is observed that the greater the targets' energy consumption and shadowing probability constraints are, the higher the achieved utility of the targets is, even when selecting a relatively small number of RISs to contribute to their PNT service. Furthermore, the proposed alternative PNT solution converges to low average estimation error and GDOP values for the targets, while simultaneously achieving high average channel gain (Fig. 3c), and converging practically in less than 10 iterations of the overall framework, which corresponds to less than 4 seconds.

B. Benefits of the Proposed Alternative PNT Solution

In this section, the benefits of the proposed alternative PNT solution are demonstrated through a detailed comparative evaluation. Specifically, four different scenarios are compared: (i) Proposed Approach, (ii) No PS: Our proposed solution without performing the RISs elements phase shift optimization, (iii) No PS & No RL: the RISs are randomly selected by the targets and no phase shift optimization is performed on their elements, and (iv) No RIS: the PNT service is exclusively supported by the anchor nodes' signals.

Fig. 4a–4c show the target's average channel gain, GDOP, and position and timing estimation error as a function of the frameworks' iterations τ, considering the four comparative alternative PNT solutions. The results reveal that as the alternative PNT solution becomes less intelligent i.e., no RISs' elements phase shift optimization (No PS scenario), random selection of RISs by the targets (No PS & No RL scenario), or not even using the RISs to support the PNT service (No PS & No RL scenario), the targets' average channel gain substantially decreases (Fig. 4a), while the targets' average GDOP and position and timing estimation error converge to significantly high values (Fig. 4b–4c). Thus, we conclude that the multifaceted intelligence that is introduced by our proposed alternative PNT solution, combining the benefits of the satisfaction games, the reinforcement learning, and the RIS technology, results in a highly accurate PNT service.

C. Intelligent Selection of the Number of RISs

In this section, we further corroborate on the benefits of the proposed alternative PNT solution, especially from the perspective of employing the theory of satisfaction games that enables the targets to select the optimal number of RISs that will contribute to their PNT service. Towards this direction, we provide a comparison between four alternative approaches for selecting the optimal number of RISs for each target, (i) Our Method; (ii) 50% of RISs: each target selects 50% of the total number of RISs available in the system, (iii) Average number of RISs: each target selects $\frac{|R|}{|N|}$ number of RISs, (iv) All RISs: each target uses all the available RISs in order to determine its position and timing. It is noted that the RISs' elements phase shift optimization presented in Section V is performed at all the comparative scenarios for fairness purposes, while in the second and third scenarios, the specific set of RISs that finally supports a target's PNT service is selected randomly.

Fig. 5a-5b present the targets' shadowing probability and energy consumption as a function of the target's ID for all four comparative scenarios, respectively. It is justified that only our proposed framework that follows the principles of satisfaction games can jointly respect shadowing probability (Fig. 5a) and the energy (Fig. 5b) consumption constraints. On the other hand, when the targets exploit all the RISs to determine their

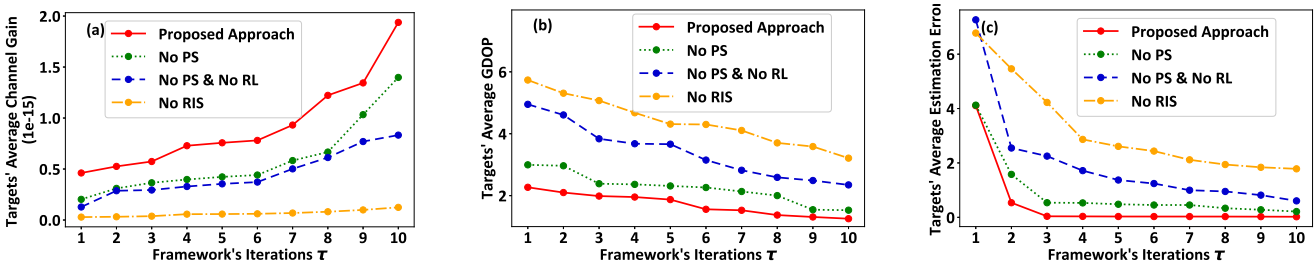

Fig. 4: Benefits of the proposed alternative PNT solution – A Comparative Evaluation.

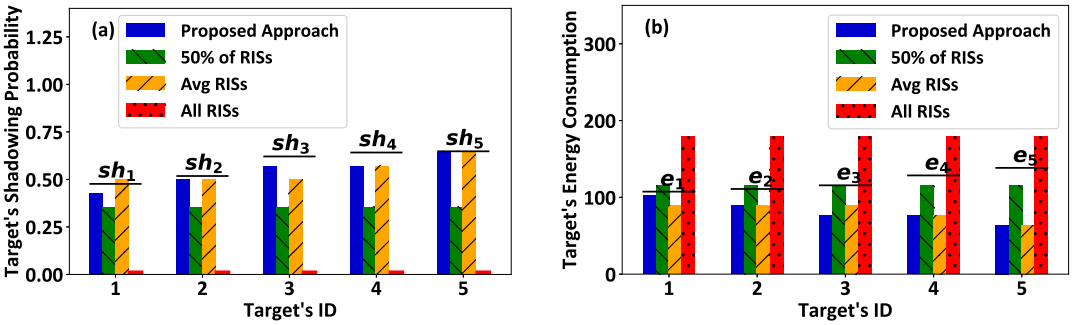

Fig. 5: Intelligent selection of the number of RISs – Benefits of satisfaction games.

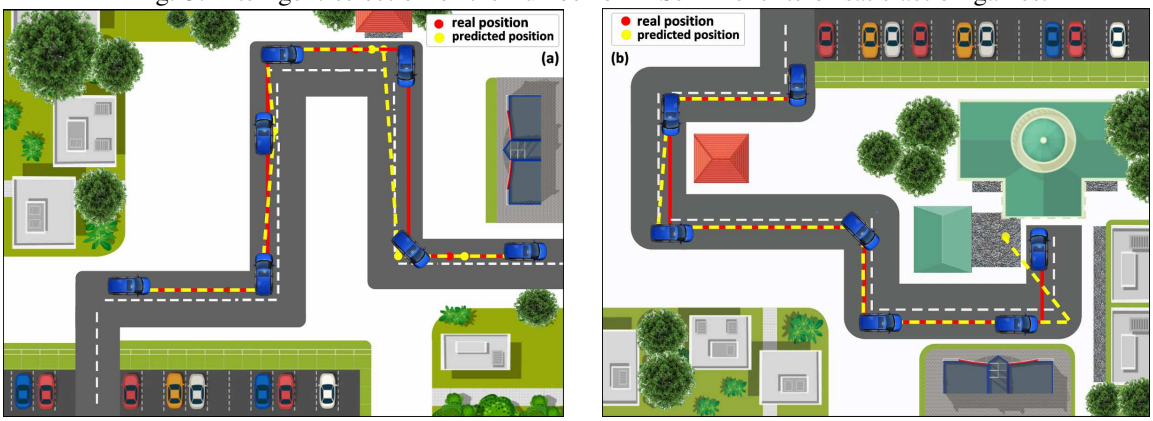

Fig. 6: Mobility scenario and accuracy of the proposed alternative PNT solution.

PNT, i.e., All RISs scenario, their shadowing probability is very low (Fig. 5a), given that there is higher probability of establishing LoS communication links. Nevertheless, the experienced energy consumption is extremely high (Fig. 5b), as the number of received signals at the target by both the anchor nodes and by reflection from the RISs increase, yielding at higher decoding overhead and thus, consumed energy. Also, the scenarios of selecting 50% or the average number of RISs existing in the system present similar trend, and there is no guarantee that the targets' energy consumption and shadowing probability constraints will be satisfied.

D. Targets Mobility and PNT

In this section, we are emulating a realistic mobility use case of the targets considering two different scenarios. In the first use case scenario, the target is in a favorable position (Fig. 6a) with respect to the obstacles in the surrounding environment, thus, experiencing a very limited shadowing effect. The exact opposite holds true in the second use case scenario (Fig. 6b) where the target is blocked by several obstacles during its navigation. The results demonstrate that the mobile target's positioning and navigation is successful and accurate for a practical and real-life implementation scenario. Also, it is observed that higher estimation error is experienced by the target, when it is blocked by surrounding obstacles (Fig. 6b). However, the intelligent exploitation of the RISs' reflected signals contribute to partially overcoming the topological defects and obstacles, and thus result in a highly accurate PNT service in both scenarios.

E. Comparative Evaluation

In this section, we provide a comparative evaluation of our proposed framework to the main existing approaches in the literature that provide alternative PNT solutions by exploiting the capabilities of the RIS technology. Six comparative scenarios have been considered: (i) Proposed Approach, (ii) Scenario I: the targets randomly select $\lceil \frac{\sum_{n=1}^{|N|} |R_n|}{|N|} \rceil$ RISs, (iii) Scenario II: the targets randomly select $\lceil \frac{|R|}{|N|} \rceil$ RISs, (iv) Scenario III and IV: the targets select $\lceil \frac{\sum_{n=1}^{|N|} |R_n|}{|N|} \rceil$ and $\lceil \frac{|R|}{N} \rceil$

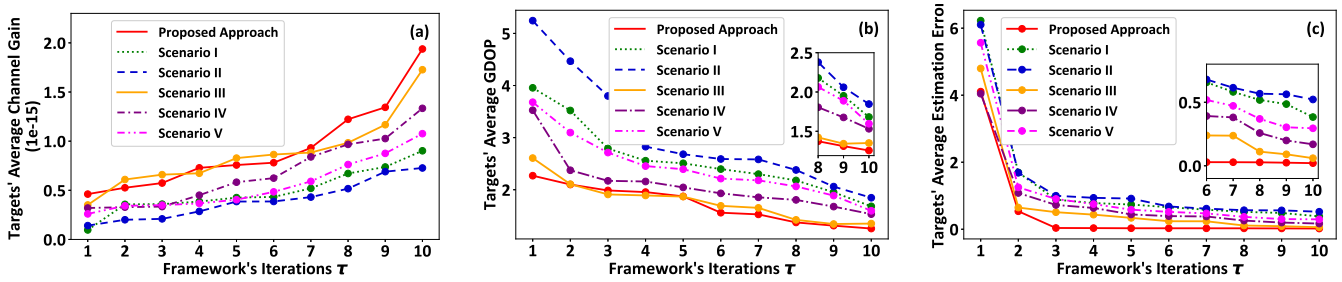

Fig. 7: Comparative evaluation.

RISs, respectively, based on the proposed reinforcement learning-based RIS selection mechanism (Section IV), and (v) Scenario V: the targets select $\lceil \frac{|R|}{|N|} \rceil$ RISs based on the best channel gain conditions contributed by the selected RISs, i.e., $\max_r \mathbf{h}_{r,n}^H \mathbf{\Omega} \mathbf{h}_{a,r}$ [21]. It is noted that in all of the aforementioned comparative scenarios, the RISs phase shift optimization is performed.

Fig. 7a-7c demonstrate the targets' average channel gain, the GDOP, and the position and timing estimation error as a function of the frameworks' iterations τ for all the six comparative scenarios. The results reveal that our proposed alternative PNT solution achieves the most accurate PNT service with the lowest estimation error (Fig. 7c), while simultaneously achieving the lowest GDOP (Fig. 7b) and the highest channel gain (Fig. 7a) for the targets. On the other hand, the importance of adopting the reinforcement learning approach in order to enable the targets to select the specific set of RISs that contribute to their PNT service is highlighted by comparing scenarios I and II, to scenarios III, IV, and V. Also, the selection of the RISs based on the best channel gain conditions contributed by the selected RISs outperforms the random-based RISs selection in terms of targets' average channel gain (Fig. 7a), GDOP (Fig. 7b), and estimation error (Fig. 7c). Scenarios I and II achieve overall better results in terms of providing an accurate PNT service compared to Scenarios III and IV.

VIII. Conclusions

In this paper, a novel alternative PNT solution is introduced that capitalizes on the key enabling technology of 6G networks, i.e., the RISs, while adopting and applying the theory satisfaction games and the reinforcement learning for their proper configuration and orchestration. Initially, a satisfaction game among the targets, which have unknown coordinates, is formulated to determine the number of selected RISs that will contribute to their PNT service, while guaranteeing their energy consumption and shadowing probability constraints. Then, a reinforcement learning-based approach enables the targets to select the specific set of RISs in order to minimize the position and timing estimation error, given the optimal number of RISs as determined at the previous stage. Afterwards, the RISs' elements phase shift optimization is performed to maximize the received signal strength of the signals transmitted by the anchor nodes, reflected from the selected RISs, and received by the targets. Finally, an Iterative Least Square (ILS) algorithm is used to accurately determine the targets' position and timing, capitalizing on the previous findings. A detailed set of numerical results are presented in order to show the pure operation and performance of the proposed framework, as well as its benefits compared to other prospective alternative PNT solutions. The main benefits of the proposed alternative PNT solution compared to the state-of-the-art can be summarized as follows: (i) improve the GDOP and the targets' estimation error by jointly exploiting the RIS technology and the intelligent selection of RISs, (ii) introduce a distributed PNT solution, where each target autonomously senses its environment, selects the RISs, and optimizes their phase shifts, and (iii) implement a low-computational complexity PNT solution that respects the targets' personal characteristics, in terms of energy and shadowing constraints.

Part of our current and future work is to extend the proposed alternative PNT solution in featureless terrains, aiming at minimizing the number of anchor nodes, the deployment of which is particularly costly and challenging under these circumstances.

References

[1] F. S. Prol, R. M. Ferre, Z. Saleem, P. Välisuo, C. Pinell, E. S. Lohan, M. Elsanhoury, M. Elmusrati, S. Islam, K. Çelikbilek, K. Selvan, J. Yliaho, K. Rutledge, A. Ojala, L. Ferranti, J. Praks, M. Z. H. Bhuiyan, S. Kaasalainen, and H. Kuusniemi, "Position, navigation, and timing (pnt) through low earth orbit (leo) satellites: A survey on current status, challenges, and opportunities," *IEEE Access*, vol. 10, pp. 83 971–84 002, 2022.

[2] E. Falletti, D. Margaria, G. Marucco, B. Motella, M. Nicola, and M. Pini, "Synchronization of critical infrastructures dependent upon gnss: Current vulnerabilities and protection provided by new signals," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2118–2129, 2019.

[3] "National research and development plan for positioning, navigation, and timing resilience," https://www.whitehouse.gov/wp-content/uploads/2021/08/Position_Navigation_Timing_RD_Plan-August-2021.pdf NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, 2021.

[4] X. Cao, B. Yang, C. Huang, C. Yuen, M. D. Renzo, D. Niyato, and Z. Han, "Reconfigurable intelligent surface-assisted aerial-terrestrial communications via multi-task learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3035–3050, 2021.

[5] S. Papavassiliou, E. E. Tsiropoulou, P. Promponas, and P. Vamvakas, "A paradigm shift toward satisfaction, realism and efficiency in wireless networks resource sharing," *IEEE Network*, vol. 35, no. 1, pp. 348–355, 2021.

[6] S. Han, Z. Gong, W. Meng, C. Li, and X. Gu, "Future alternative positioning, navigation, and timing techniques: A survey," *IEEE Wireless Communications*, vol. 23, no. 6, pp. 154–160, 2016.

[7] H. Xiong, R. Bian, Y. Li, Z. Du, and Z. Mai, "Fault-tolerant gnss/sins/dvl/cns integrated navigation and positioning mechanism based on adaptive information sharing factors," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3744–3754, 2020.

[8] D. DeVon, T. Holzer, and S. Sarkani, "Innovation-based fusion of multiple satellite positioning systems for minimizing uncertainty," *IEEE Systems Journal*, vol. 13, no. 1, pp. 928–939, 2019.

[9] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multisensor localization for the internet of things: Exploring a new class of scalable localization algorithms," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 153–167, 2018.

[10] W. Xue, R. Ying, X. Chu, R. Miao, J. Qian, and P. Liu, "Robust navigation under incomplete localization using reinforcement learning," in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, 2020, pp. 1618–1624.

[11] B.-J. Jeon, I. Petrunin, and A. Tsourdos, "Recurrent neural network based sensor fusion algorithm for alternative position, navigation and timing," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*. IEEE, 2021, pp. 1–7.

[12] M. S. Hossain, N. Irtija, E. E. Tsiropoulou, J. Plusquellic, and S. Papavassiliou, "Reconfigurable intelligent surfaces enabling positioning, navigation, and timing services," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 4625–4630.

[13] A. Singh, V. Kotiyal, S. Sharma, J. Nagar, and C.-C. Lee, "A machine learning approach to predict the average localization error with applications to wireless sensor networks," *IEEE Access*, vol. 8, pp. 208 253–208 263, 2020.

[14] Z. Lalama, S. Boulfekhar, and F. Semechedine, "Localization optimization in wsns using meta-heuristics optimization algorithms: A survey," *Wireless Personal Comm.*, vol. 122, no. 2, pp. 1197–1220, 2022.

[15] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser miso systems exploiting deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, 2020.

[16] S. Huang, B. Wang, Y. Zhao, and M. Luan, "Near-field rss-based localization algorithms using reconfigurable intelligent surface," *IEEE Sensors Journal*, vol. 22, no. 4, pp. 3493–3505, 2022.

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

[18] M. Luan, B. Wang, Y. Zhao, Z. Feng, and F. Hu, "Phase design and near-field target localization for ris-assisted regional localization system," *IEEE Trans. on Vehicular Tech.*, vol. 71, no. 2, pp. 1766–1777, 2022.

[19] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Metalocalization: Reconfigurable intelligent surface aided multi-user wireless indoor localization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 7743–7757, 2021.

[20] Y. Fang, S. Atapattu, H. Inaltekin, and J. Evans, "Optimum reconfigurable intelligent surface selection for wireless networks," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 6241–6258, 2022.

[21] N. Mensi and D. B. Rawat, "Reconfigurable intelligent surface selection for wireless vehicular communications," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1743–1747, 2022.

[22] M. S. Siraj, A. B. Rahman, M. Diamanti, E. E. Tsiropoulou, S. Papavassiliou, and J. Plusquellic, "Orchestration of reconfigurable intelligent surfaces for positioning, navigation, and timing," in *MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM)*, 2022, pp. 148–153.

[23] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[24] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, 2008.

[25] M. Zhang and J. Zhang, "A fast satellite selection algorithm: Beyond four satellites," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 740–747, 2009.

[26] G. Fragkos, E. E. Tsiropoulou, and S. Papavassiliou, "Artificial intelligence enabled distributed edge computing for internet of things applications," in *2020 16th international conference on distributed computing in sensor systems (DCOSS)*. IEEE, 2020, pp. 450–457.

[27] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, Z. Zhang, and M. Debbah, "Multi-hop ris-empowered terahertz communications: A drl-based hybrid beamforming design," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1663–1677, 2021.

Md Sadman Siraj is a Ph.D. student and a research assistant in the Department of Electrical and Computer Engineering at the University of New Mexico. He received his Bachelor's degree in Electrical and Electronic Engineering from the University of Dhaka in 2020. His research interests lie in the areas of Alternative Positioning, Navigation and Timing (APNT) services and novel localization techniques, resource management, orchestration and optimization, game theory, and reinforcement learning.

Aisha B Rahman is a Ph.D. student and a research assistant in the Department of Electrical and Computer Engineering at the University of New Mexico. She received her Bachelor's and Master's degree in Electrical and Electronic Engineering from the University of Chittagong in 2019 and 2021 respectively. Her research interests lie in the areas of wireless networking, resource management and optimization, game theory, Alternative Positioning, Navigation and Timing (APNT) services, and reinforcement learning.

Maria Diamanti is a Ph.D. student and a research assistant in the School of Electrical and Computer Engineering at the National Technical University of Athens. She received her Diploma in Electrical and Computer Engineering from the Aristotle University of Thessaloniki in 2018. Her research interests lie in the areas of 5G/6G wireless networks, resource management and optimization, game theory, contract theory, and reinforcement learning.

Eirini Eleni Tsiropoulou is currently an Assistant Professor at the Department of Electrical and Computer Engineering, University of New Mexico. Her main research interests lie in the area of wireless heterogeneous networks, with emphasis on network modeling and optimization. She was selected by the IEEE Communication Society - N2Women - as one of the top ten Rising Stars of 2017 in the communications and networking field. She received the NSF CRII Award in 2019 and the Early Career Award by the IEEE Communications Society Internet Technical Committee in 2019.

Symeon Papavassiliou is a Professor in the School of Electrical and Computer Engineering (ECE) at National Technical University of Athens. From 1995 to 1999, he was a senior technical staff member at AT&T Laboratories, New Jersey. In August 1999 he joined the ECE Department at the 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 400 technical journal and conference published papers. His main research interests lie in the areas of modeling, optimization and performance evaluation of distributed complex systems and social networks.

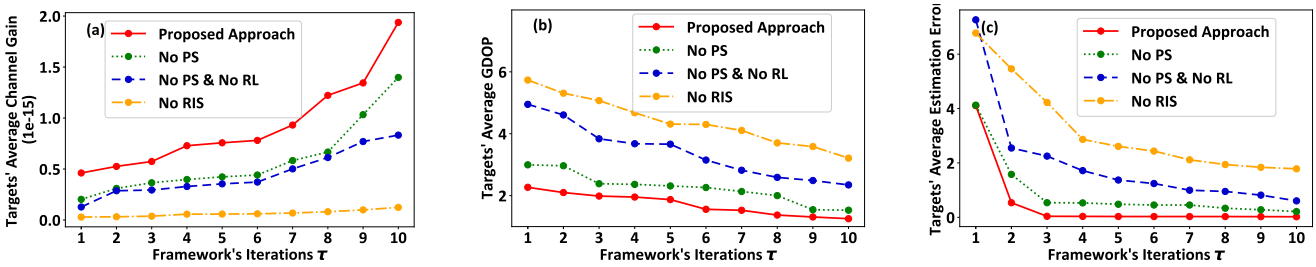

Fig. 4: Benefits of the proposed alternative PNT solution – A Comparative Evaluation.

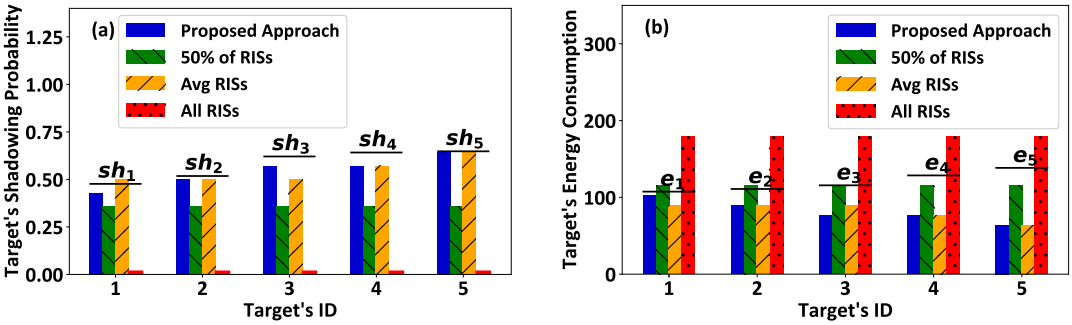

Fig. 5: Intelligent selection of the number of RISs – Benefits of satisfaction games.

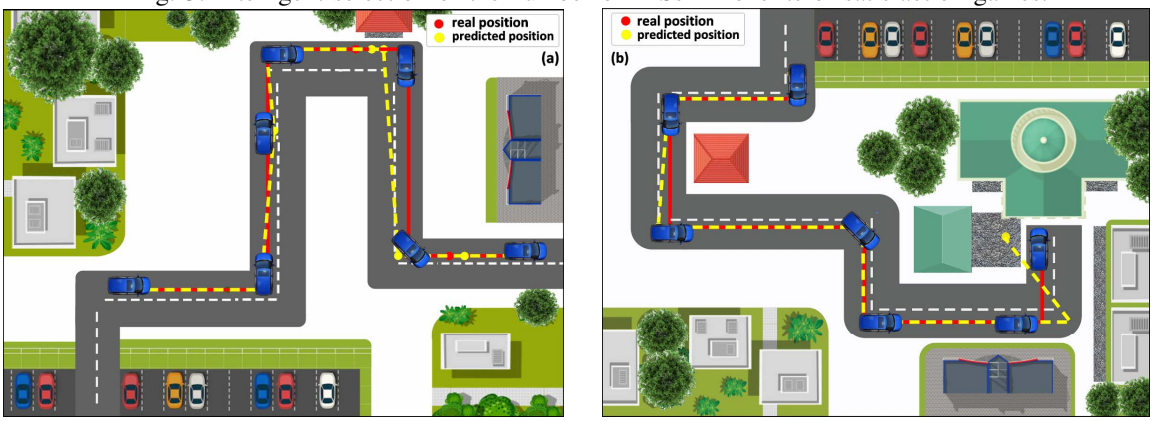

Fig. 6: Mobility scenario and accuracy of the proposed alternative PNT solution.

PNT, i.e., All RISs scenario, their shadowing probability is very low (Fig. 5a), given that there is higher probability of establishing LoS communication links. Nevertheless, the experienced energy consumption is extremely high (Fig. 5b), as the number of received signals at the target by both the anchor nodes and by reflection from the RISs increase, yielding at higher decoding overhead and thus, consumed energy. Also, the scenarios of selecting 50% or the average number of RISs existing in the system present similar trend, and there is no guarantee that the targets' energy consumption and shadowing probability constraints will be satisfied.

D. Targets Mobility and PNT

In this section, we are emulating a realistic mobility use case of the targets considering two different scenarios. In the first use case scenario, the target is in a favorable position (Fig. 6a) with respect to the obstacles in the surrounding environment, thus, experiencing a very limited shadowing effect. The exact opposite holds true in the second use case scenario (Fig. 6b) where the target is blocked by several obstacles during its navigation. The results demonstrate that the mobile target's positioning and navigation is successful and accurate for a practical and real-life implementation scenario. Also, it is observed that higher estimation error is experienced by the target, when it is blocked by surrounding obstacles (Fig. 6b). However, the intelligent exploitation of the RISs' reflected signals contribute to partially overcoming the topological defects and obstacles, and thus result in a highly accurate PNT service in both scenarios.

E. Comparative Evaluation

In this section, we provide a comparative evaluation of our proposed framework to the main existing approaches in the literature that provide alternative PNT solutions by exploiting the capabilities of the RIS technology. Six comparative scenarios have been considered: (i) Proposed Approach, (ii) Scenario I: the targets randomly select $\lceil \frac{\sum_{n=1}^{|N|} |R_n|}{|N|} \rceil$ RISs, (iii) Scenario II: the targets randomly select $\lceil \frac{|R|}{|N|} \rceil$ RISs, (iv) Scenario III and IV: the targets select $\lceil \frac{\sum_{n=1}^{|N|} |R_n|}{|N|} \rceil$ and $\lceil \frac{|R|}{N} \rceil$