sensors

MDPI

Article

Cross-Domain Knowledge Transfer for Sustainable Heterogeneous Industrial Internet-of-Things Networks

Zhenzhen Gong [1], Qimei Cui [1,*] and Wei Ni [2]

[1] National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China; gongzhenzhen0822@gmail.com
[2] Data61, Commonwealth Science and Industrial Research Organization (CSIRO), Marsfield, NSW 2122, Australia; wei.ni@data61.csiro.au
* Correspondence: cuiqimei@bupt.edu.cn

Abstract: In this article, a novel cross-domain knowledge transfer method is implemented to optimize the tradeoff between energy consumption and information freshness for all pieces of equipment powered by heterogeneous energy sources within smart factory. Three distinct groups of use cases are considered, each utilizing a different energy source: grid power, green energy source, and mixed energy sources. Differing from mainstream algorithms that require consistency among groups, the proposed method enables knowledge transfer even across varying state and/or action spaces. With the advantage of multiple layers of knowledge extraction, a lightweight knowledge transfer is achieved without the need for neural networks. This facilitates broader applications in self-sustainable wireless networks. Simulation results reveal a notable improvement in the 'warm start' policy for each equipment, manifesting as a 51.32% increase in initial reward compared to a random policy approach.

Keywords: industrial internet-of-things (IIoT); age of information (AoI); energy efficiency; cross-domain

Citation: Gong, Z.; Cui, Q.; Ni, W. Cross-Domain Knowledge Transfer for Sustainable Heterogeneous Industrial Internet-of-Things Networks. *Sensors* **2024**, *24*, 3265. https://doi.org/10.3390/s24113265

Academic Editor: Peter Han Joo Chong

Received: 7 April 2024
Revised: 29 April 2024
Accepted: 18 May 2024
Published: 21 May 2024

1. Introduction

The sustainability of communication networks is a critical goal for next-generation wireless systems (e.g., 6G and beyond [1]). Network sustainability is defined as an approach that successfully integrates and balances environmental responsibility, economic viability, and social equity. Despite the growing attention and hype surrounding the sustainability of 6G, there is a lack of a rigorous and practical definition to guide its implementation in networks. Sustainability has been mainly linked to green networking to achieve the United Nations' Sustainable Development Goals (SDGs) [2]. In practice, this is particularly related to energy efficiency of the versatile network elements. In particular, smart factories constitute a significant component in Industrial Internet-of-Things (IIoT) [3] and Industry 4.0 [4], playing a key role in enabling cyber-physical systems to function autonomously. IIoT applications typically requires the automation of a large number of devices in manufacturing with limited hardware capabilities and energy resources, usually with small batteries [5]. Industrial 4.0 [6] encompasses emerging technologies, such as artificial intelligence (AI), edge computing, and digital twin (DT) and so on. In particular, the work in [7] comprehensively investigated the intelligence maintenance in various aspects of maintenance. Specifically, it focused on the human-in-the-loop-based maintenance and its role in enhancing physical resilience in smart manufacturing. This paradigm requires increased flexibility, agility and resilience through the lifespan of the IIoT devices. Consequently, in the realm of IIoT, the smart factories are expected to integrate advanced autonomous capabilities along with enhanced energy-efficient functionality. Nevertheless, the robots, sensors and actuators in the factories are empowered with different sources of energy. Such sources include power grids [8], renewable technologies [9] (e.g., solar), and other energy harvesting techniques [10] (e.g., radio frequency (RF) energy). Subsequently, ensuring the

energy efficiency of each individual equipment necessitates adopting a unique mode of operation that is specifically tailored to the varying availability and abundance of their respective energy sources. This can have a direct implication on other critical performance metrics of operation in smart factory. Chief among these metrics is the *age of information (AoI)* [11] that represents the degree of freshness of the data acquired from the monitored autonomous physical systems [12]. With a focus on both energy efficiency and information freshness, the sustainability of each individual equipment can be significantly enhanced. However, assuring the sustainability of the IIoT as a whole requires looking beyond the individual equipment. In fact, the overall performance and environmental impact of the IIoT will crucially depend not only on the performance of single piece of equipment but also on long-term environmental friendliness of its solution. This encompasses considerations of the system's overall energy consumption and its ability to sustain prolonged operation without causing harmful impacts on the environment, by considering the associated complexity and energy efficiency of the solution.

The minimization of hybrid energy sources in smart factories has been extensively investigated in various scenarios [13,14]. For instance, the works in [13,14] study the minimization of grid energy consumption in a mixed energy supply scenario. Nonetheless, these works leverages reinforcement learning (RL) solutions [15] that assume a homogeneous model across equipment having heterogeneous energy utilities. In fact, these studies often assume uniformity of state and/or action spaces between heterogenous scenarios, which can barely hold true with the unique operation associated to each equipment [16]. Therefore, in practical real-world scenarios, a robust RL approach is needed to effectively address the heterogeneous nature of the cyber-physical system, while ensuring the sustainability of the solution. Notably, one should consider an RL solution that generalizes across multiple tasks. For instance, the works in [17,18] employ multiple experts to optimize the aggregated performance across different groups. However, the use of multiple agents hinders knowledge sharing among these groups and leads to increased costs as the number of groups grows. The work in [19] considers a federated imitation learning method for cross-domain knowledge sharing framewor. However, the utilization of neural networks slows down the learning process. Furthermore, the application of gradient descent (GD) [20] in such operations incurs additional energy costs as it requires a significant amount of resources to converge. Consequently, to ensure network sustainability, encompassing both the energy efficiency of individual equipment and the computational efficiency of the entire network, a more universally applicable and generalizable solution is essential for heterogeneous Internet of Things (IoT).

The main contributions of this paper is the development of a globally generalizable RL solution, designed to enhance the overall sustainability of cyber-physical systems comprising heterogeneous energy sources. In particular, we tackle the sustaibality issues at both the equipment and system levels by introducing a lightweight, cross-domain knowledge sharing solution. This innovative approach leverages a three-layered knowledge repository structure to facilitate efficient knowledge storage and transfer across the system. Numerical simulations demonstrate that the proposed method consistently outperforms other baseline methods in computational complexity while maintain a comparable performance for smart factories.

The rest of this paper is organized as follows. The system models and problem formulation are provided in Section 2. The proposed cross-domain knowledge sharing framework and the corresponding solutions are presented in Section 3. Simulation results are given in Section 4. Finally, conclusions and future works are drawn in Section 5.

2. System Models

Consider a set $\mathcal{N}$ of N smart factory equipment having heterogeneous energy resources in a smart factory. These pieces of equipment are distributed for various manufacturing purposes such as supply chain integration, pre-production setup, production, quality control and inspection, packaging and storage, delivery and so on. Each equipment collects

sensory data from its surrounding environment and subsequently executes actions that are tailored to the information gathered. As illustrated in Figure 1, these pieces of equipment are clustered into three distinct groups according to their energy sources. We use $x \in \mathcal{X}$ to index the three groups such that $x = 1, 2, \ldots, X$, whereby each group includes a set $\mathcal{N}_x$ of N_x equipment. In particular, three sources of energy supply are considered: (i) *grid power (GP), (ii) green sources (GS), and (iii) mixed sources (MS)*. Specifically, MSs encompasses both the grid and harvested energy resources. In addition, cyber-physical equipment within each group collects data packets from their respective surrounding environments and abstract useful information using their processing capabilities. As illustrated in Figure 2, the abstracted information is subsequently transmitted to a nearby base station (BS).

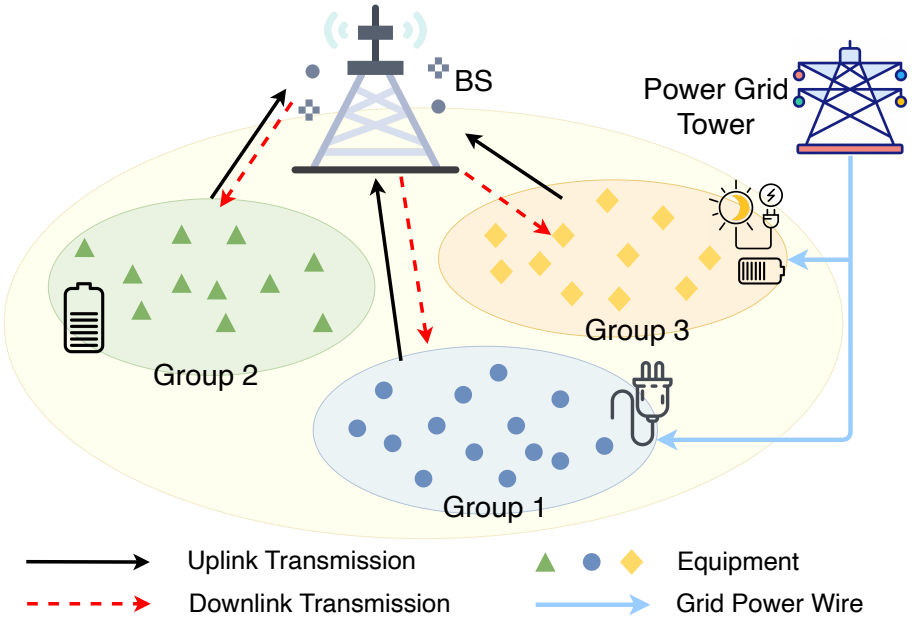

Figure 1. System model of three groups of smart factory equipment with diverse energy sources. Group 1 supplied by grid power wire, group 2 supplied by battery energy and group 3 supplied by grid power and green power source.

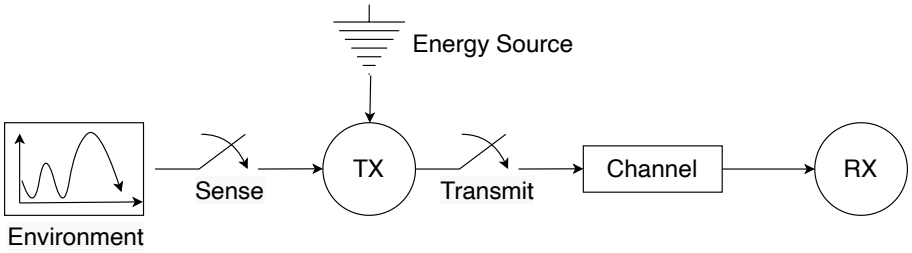

Figure 2. Illustrative figure of the system model representing an smart factory.

We consider a time-slotted system where each timeslot has a uniform length denoted as τ. These timeslots are indexed sequentially as $t = 1, 2, \ldots, T$. A Rayleigh fading channel is considered for the uplink communication between smart factory equipment and BS. The data transmission rate $\phi(c_{y(t)})$ for each equipment $y \in \mathcal{N}$ at time slot t can be obtained as below:

$$\phi(c_y(t)) = B \log_2\left(1 + \frac{g p_y(t)}{I + BN_0}\right), \ \forall y \in \mathcal{N}, \tag{1}$$

where $c_y(t)$ is the number of bits to be processed, $\phi(c_y(t))$ is the number of bits to be transmitted after processing, $p_y(t) \in [0, p_{y,\max}]$ (in dBm) is the transmitter power used to upload the abstracted information, B is the channel bandwidth, I is the interference from other pieces of equipment in corresponding group, N_0 is noise power spectral density, g is

the channel response, which is related to the distance between each equipment y and the BS, i.e., l_y. Next, we present the energy models of each group based on their energy sources:

1. *GP Source* : GP typically refers to power that is supplied through an electrical grid. Hence, GP-powered equipment does not have energy limitations. For instance, the robots and actuators in production line are connected to grid energy supply. The energy consumption $e_i(t)$ of each equipment $i \in \mathcal{N}_1$ can be divided into two categories: (a) transmission energy $e_i^T(t) = \tau p_i(t)$ consumed to transmit abstracted information to the BS and (b) computing energy $e_i^C(t) = \varsigma \kappa_i \vartheta^2 c_{i,t}$ used to process the collected data packets:
$$e_i(t) = e_i^T(t) + e_i^C(t) = \tau p_i(t) + \varsigma \kappa_i \vartheta^2 c_{i,t}, \tag{2}$$
where ς is the energy consumption coefficient depending on the chip of each IIoT equipment, κ_i is the number of central processing unit (CPU) cycles required for processing per bit data, assumed to be equal for all pieces of equipment and ϑ is the frequency of the CPU clock of each equipment [21].
2. *GS Source* : Renewable energy sources, such as wind power, solar power, thermal power and RF are used to enable the establishment of a self-sustainable green network. For example, drones and robots utilized for quality inspection and automated delivery systems are predominantly powered by battery technology. This reduces dependence on conventional grid energy and, consequently, enhances the mobilities while offering greater flexibility and efficiency in operational processes. These energy harvesting methods consistently capture energy from natural environments, converting it into electrical power and storing the collected energy in rechargeable batteries. We define $E_{\max}$ as the maximum amount of energy that can be stored in a battery. When the battery reaches its full capacity, any additional harvested energy will be discarded. Consider an ideal rechargeable battery with no energy loss during storage or retrieval processes. At each time slot, the harvested energy $e_j^h(t) \geq 0$ by equipment $j \in \mathcal{N}_2$ follows follows a Bernoulli distribution with probability $\sigma \in [0, 1]$, such that:
$$e_j^h(t) = p^{\text{solar}} \times \epsilon_0 \times \epsilon_1 \times \epsilon_2 \,, \tag{3}$$
where p^{solar} is the density of solar power to the equipment [22]. We consider a typical solar-powered equipment that is equipped with a photovoltaic panel with size ϵ_0 and the energy transfer efficiency ϵ_1. Considering the heterogeneity in solar power density, a uniformly distributed random variable, ϵ_2 is taken into account. Consequently, the energy level of the battery $e_j^b(t)$ will be given by:
$$e_j^b(t+1) = \min\{E_{\max},\, e_j^b(t) + e_j^h(t) - e_j^T(t) - e_j^C(t)\}, \tag{4}$$
where $e_j^T(t)$ and $e_j^C(t)$ are transmission energy and computing energy for equipment j. Moreover, the following constraint stands:
$$0 \leq e_j^b(t) \leq E_{\max} \tag{5a}$$
$$e_j^T(t) + e_j^C(t) = e_j(t) \leq e_j^b(t), \tag{5b}$$
where $e_j(t)$ is the energy consumption at time slot t for equipment in group 2. (5a) implies the battery limitation of equipment $j \in \mathcal{N}_2$. (5b) implies that the available energy, which can be used for processing and transmitting energy at the beginning of each time slot, must not exceed the energy level of the battery.
3. *MS Source* : The third group of cyber-physical equipment is powered by hybrid energy sources comprising both the grid and renewable energy sources. For example, the industrial sensors are strategically deployed to monitor a range of environmental parameters as well as the status of products. This design aims to reduce energy consumption from the grid power while mitigating the randomness and intermittency

associated with green energy. Accordingly, for an equipment $k \in \mathcal{N}_3$, the consumed energy at time slot t comprises two sources: grid energy $e_k^G(t)$ and battery energy $e_k^B(t)$. We assume the same energy harvesting model as previously defined, such that $e_k^h(t)$ is updated as in (3). Different from GS, the battery level is updated as:

$$e_k^b(t+1) = \min\{E_{\max},\ e_k^b(t) + e_k^h(t) - e_k^B(t)\} \tag{6}$$

where $e_k^b(t)$ is the battery level at each time slot. Furthermore, the following constraints are held:

$$e_k^T(t) + e_k^C(t) \leq e_k^B(t) + e_k^G(t)\,, \tag{7a}$$
$$e_k^B(t) \leq e_k^b(t)\,, \tag{7b}$$
$$e_k(t) = e_k^G(t) \tag{7c}$$
$$0 \leq e_k^b(t) \leq E_{\max} \tag{7d}$$

where $e_k^T(t)$ and $e_k^C(t)$ are the transmission and computing energy separately. Moreover, (7a) implies that the consumed energy must not exceed the total energy provided by both the battery and the grid. (7b) implies that the permissible battery energy must not exceed the available battery capacity. Since our optimization objective is to minimize grid energy consumption, we set the energy optimization variable $e_k(t)$ equal to $e_k^G(t)$ in (7c). (7d) indicates the battery limitation of each equipment $k \in \mathcal{N}_3$.

AoI Model for Heterogeneous Scenarios

At each time slot, sensing data packets arrive at the equipment with a probability λ_y. The size of the data packet, denoted as $a_y(t)$, follows a Gaussian distribution. Data packets, once collected, are placed in a waiting queue. The system processes and transmits these packets employing a first-come-first-served (FCFS) approach. Consequently, the queue size $b_y(t)$ can be updated as follows:

$$b_y(t+1) = b_y(t) + a_y(t) - c_y(t). \tag{8}$$

The AoI at time slot t is defined as the timestamp of the most recently processed and successfully received packet at the receiver. The entire process encompasses both data processing time and transmission time. Formally, the update of AoI $\Delta_y(t)$ is as follows:

$$\Delta_y(t+1) = \begin{cases} \Delta_y(t) + 1\,, & \text{if } \varpi_y(t) = 1 \\ \min\{(t+1) - U(t), \Delta_{\max}\}\,, & \text{otherwise.} \end{cases} \tag{9}$$

where, $U(t)$ represents the generation timestamp of the most recent packet and $\Delta_{\max}$ is the maximum AoI value and $\varpi_y(t) = 1$ indicates that the processing of a packet is finished. This limit is imposed to constrain the impact of AoI on performance after a certain level of staleness is reached.

With the aforementioned models, we can now proceed to our optimization objective. For each equipment, we define a cost function $w_y(t) = \eta_1 \Delta_y(t) + (1 - \eta_1) e_y(t)$, where η_1 is a tradeoff factor to balance AoI and energy cost and $e_y(t)$ represents the energy cost of any equipment of three groups. Our objective is to minimize the averaged cost $w_y(t)$ of all pieces of equipment throughout all the time, which can be written as:

$$\min_{\{p_y(t)\},\{e_y^G(t)\},\{e_y^B(t)\}} \sum_{x=1}^{3} \frac{1}{N_x} \sum_{y \in \mathcal{N}_x} w_y(t) \tag{10}$$

$$s.t. \quad 0 \le p_y(t) \le p_{y,\max} \tag{10a}$$

$$0 \le \Delta_y(t) \le \Delta_{\max} \tag{10b}$$

$$(2),\ \forall y \in \mathcal{N}_1 \tag{10c}$$

$$(3),\ (4),\ (5),\ \forall y \in \mathcal{N}_2 \tag{10d}$$

$$(6),\ (7),\ \forall y \in \mathcal{N}_3\,, \tag{10e}$$

where (10a) indicates that the transmission power must not surpass the maximum power of each equipment, (10b) indicates the limitation requirements of AoI. (10c) implies the conditions of equipment $i \in \mathcal{N}_1$ while (10d) and (10e) are constraints for all pieces of equipment in group GS and MS, separately.

3. Cross-Domain for Heterogeneous Scenarios

Problem (10) is NP complete, making it inherently computationally expensive. Furthermore, the unique constraints specified in (10c), (10d) and (10e) add to the complexity of this problem. Additionally, this problem is compounded by the absence of any presupposed knowledge regarding the distribution patterns of data and energy arrivals. To tackle these challenges, we adopt RL, a method that does not require prior knowledge of the underlying distribution patterns. First, three distinct RL models are presented for each group. Then, the overall minimization problem in (10) is considered as a cross-domain knowledge sharing problem.

3.1. Markov Decision Processes (MDPs) Models

Initially, the manufacturing related equipment is partitioned into three distinct groups. As such, the objective of each group is to minimize the averaged cost for all pieces of equipment. Without interfering the overall objective in (10), we model the objectives of each group using MDPs:

1. *GP Source* : The MDP tuple of the first group can be presented as $(\mathcal{S}_i, \mathcal{A}_i, \mathcal{R}_i)$, where $\mathcal{S}_i$ is the state space and $\mathcal{A}_i$ is the action space and $\mathcal{R}_i$ is the reward function separately. Particularly, $\mathcal{S}_i = \{\boldsymbol{s}_{i,t}\} = \{\Delta_{i,t}, b_{i,t} | \Delta_{i,t} \in [0, \Delta_{\max}],\ b_{i,t} \in \mathbb{N}\}$. The action space is the set of all possible transmitting powers such that $\mathcal{A}_i = \{a_{i,t}\} = \{p_{i,t} | p_{i,t} \in [0, p_{i,\max}]\}$. The reward function can be defined as $\boldsymbol{r}_i(\boldsymbol{s}_{i,t}, a_{i,t}) = w_i(t)$. The parameterized policies can be defined as $\pi_{\boldsymbol{\theta}_i}(a_{i,t}|\boldsymbol{s}_{i,t}) = \Pr\{a_{i,t}|\boldsymbol{s}_{i,t}, \boldsymbol{\theta}_i\}$, where $\boldsymbol{\theta}_i \in \mathbb{R}^{d_1}$, with $d_1 = 2$.
2. *GS Source* : The MDP tuple of the second group can be presented as $(\mathcal{S}_j, \mathcal{A}_j, \mathcal{R}_j)$. Particularly, $\mathcal{S}_j = \{\boldsymbol{s}_{j,t}\} = \{\Delta_{j,t}, b_{j,t}, e_{j,t}^b | \Delta_{j,t} \in [0, \Delta_{\max}],\ b_{j,t} \in \mathbb{N},\ e_{j,t}^b \in \{0, E_{\max}\}\}$. The action space is the the same as group 1, such that $\mathcal{A}_j = \{a_{j,t}\} = \{p_{j,t} | p_{j,t} \in [0, p_{j,\max}]\}$. Similarly, the reward function and the parameterized policies can be defined as $\boldsymbol{r}_j(\boldsymbol{s}_{j,t}, a_{j,t}) = w_j(t)$ and $\pi_{\boldsymbol{\theta}_j}(a_{j,t}|\boldsymbol{s}_{j,t}) = \Pr\{a_{j,t}|\boldsymbol{s}_{j,t}, \boldsymbol{\theta}_j\}$, where $\boldsymbol{\theta}_j \in \mathbb{R}^{d_2}$, with $d_2 = 3$.
3. *MS Source* : The MDP tuple of the third group can be presented as $(\mathcal{S}_k, \mathcal{A}_k, \mathcal{R}_k)$. Particularly, $\mathcal{S}_k = \{\boldsymbol{s}_{k,t}\} = \{\Delta_{k,t}, b_{k,t}, e_{k,t}^b | \Delta_{k,t} \in [0, \Delta_{\max}],\ b_{k,t} \in \mathbb{N},\ e_{k,t}^b \in \{0, E_{\max}\}\}$. The action space is $\mathcal{A}_k = \{a_{k,t}\} = \{p_{k,t}, e_{k,t}^B, e_{k,t}^G | p_{j,t} \in [0, p_{j,\max}], e_{k,t}^B \in [0, E_{\max}], e_{k,t}^G \in \mathbb{N}\}$. With the similar reward function $\boldsymbol{r}_k(\boldsymbol{s}_{k,t}, a_{k,t}) = w_k(t)$, the parameterized policies can be defined as $\pi_{\boldsymbol{\theta}_k}(a_{k,t}|\boldsymbol{s}_{k,t}) = \Pr\{a_{k,t}|\boldsymbol{s}_{k,t}, \boldsymbol{\theta}_j\}$, where $\boldsymbol{\theta}_k \in \mathbb{R}^{d_3}$, with $d_3 = 9$.

So far, we have successfully formulated the cost tradeoff between energy consumption and AoI as a series of MDPs, each corresponding to an individual equipment. For each equipment y, the optimization target can be writen as $\min_{\boldsymbol{\theta}_y} \mathcal{J}(\boldsymbol{\theta}_y) = \mathbb{E}[1/T \sum_{t=1}^{T} \boldsymbol{r}_{y,t}(\boldsymbol{s}_{y,t}, a_{y,t})]$. While RL demonstrates the capability to learn and optimize for each equipment individually, the scalability of this approach becomes a concern in large-scale factories due to the potentially large number of learning agents involved. More importantly, this approach can

be both time-consuming and energy-intensive. Consequently, there is a pressing need for a more efficient method that can collectively optimize across all pieces of equipment in the three distinct groups. Such a method is crucial not only for the energy efficiency of these equipment but also for the overall sustainability of the cyber-physical system.

3.2. Cross-Domain Knowledge Sharing

To facilitate learning and knowledge sharing across multiple groups, techniques like multi-task learning (MTL) and meta-learning are employed. These methods are adept at managing the simultaneous learning of multiple tasks. However, a significant limitation of these methods is their inherent assumption of model consistency across groups. This assumption poses a challenge when optimizing groups across heterogeneous groups, especially when there is a variation in the state and action spaces of these groups. Consequently, given the substantial resource costs associated with the requirement for numerous learning agents, the need for an efficient cross-domain knowledge transfer method becomes increasingly apparent.

To facilitate knowledge sharing among groups and within each group, a three-layered knowledge base is designed as in Figure 3. A global knowledge base, $\boldsymbol{L} \in \mathbb{R}^{d \times m}$, serves as the shared knowledge among groups. Three group-based knowledge matrices, denoted as $\boldsymbol{G}_x \in \mathbb{R}^{d_x \times d}$, where $x \in \{1, 2, 3\}$, are also utilized to store the knowledge specific to each group. These matrices serve as a bridge between the global knowledge base and equipment-specific mapping vectors, represented by $\boldsymbol{s}_y \in \mathbb{R}^m$, where $y \in \mathcal{N}$. As demonstrated by the MDP models of each group, the state and action spaces of each group can vary. In other words, the dimensions of the MDP policies, i.e., $\boldsymbol{\theta}_i$, $\boldsymbol{\theta}_j$, and $\boldsymbol{\theta}_k$, have different dimensions d_x, as illustrated in Figure 3. However, due to the existence of varied group knowledge bases $\boldsymbol{G}_x$, the variations of the policy vectors are mapped to the same space. This enables the achievement of a global knowledge base $\boldsymbol{L}$ that can be shared across different domains. As a result, the policy parameters of each equipment can be obtained as:

$$\boldsymbol{\theta}_y = \boldsymbol{G}_x * \boldsymbol{L} * \boldsymbol{s}_y \,. \tag{11}$$

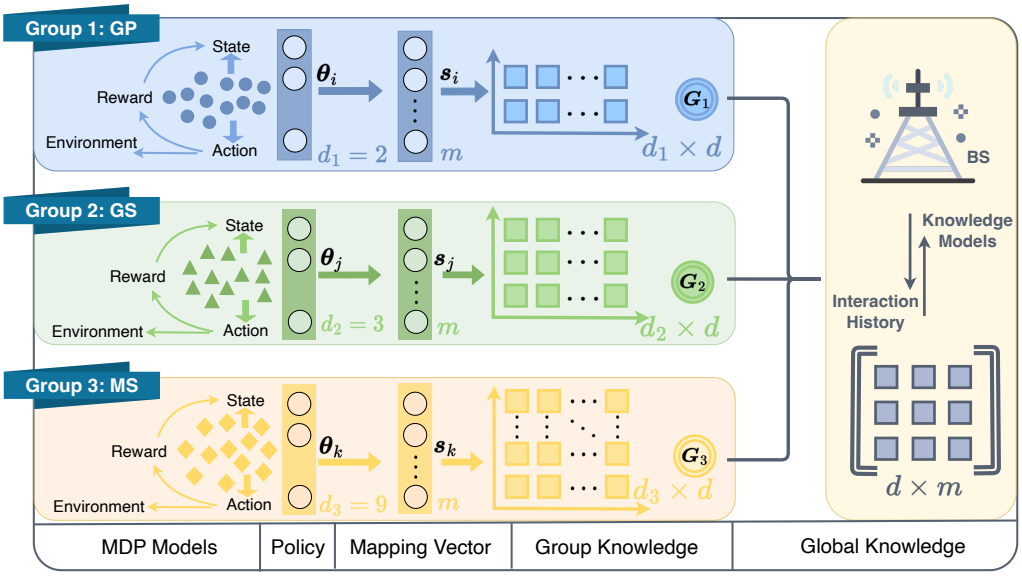

Figure 3. Illustration of three layers knowledge framework with varied state and/or action spaces.

Accordingly, our objective in (10) with the three-layered knowledge system can be represented as the minimization problem:

$$g(\boldsymbol{L}, \boldsymbol{G}_x) = \sum_{x=1}^{X} \left\{ \frac{1}{N_x} \sum_{y \in \mathcal{N}_x} \min_{\boldsymbol{s}_y} \left[\mathcal{J}(\boldsymbol{\theta}_y) + \mu_1 \|\boldsymbol{s}_y\|_1 \right] + \mu_2 \|\boldsymbol{G}_x\|_{\mathrm{F}}^2 \right\} + \mu_3 \|\boldsymbol{L}\|_{\mathrm{F}}^2$$

where L_1-norm approximates the vector sparsity and $\|\boldsymbol{L}\|_{\mathrm{F}} = (tr(\boldsymbol{L}\boldsymbol{L}'))^{1/2}$ is the Frobenius norm of matrix $\boldsymbol{L}$. The parameter μ_1 controls the balance between the policy's fit and the feature's fit. Also, μ_2 and μ_3 are two regularization parameters, where μ_2 controls the sparsity of $\boldsymbol{s}_y$. The penalty on the Frobenius norm of $\boldsymbol{G}$ and $\boldsymbol{L}$ regularizes the predictor weights to have low L_2-norm and avoids overfitting.

The above objective can be approximated by performing a second-order Taylor expansion towards $\mathcal{J}(\boldsymbol{\theta}_y)$ around the optimla policy $\boldsymbol{\alpha}_y$, which can be obtained using regular RL methods, such as policy gradient: $\boldsymbol{\alpha}_y = \arg\min_{\boldsymbol{\theta}_y} \mathcal{J}(\boldsymbol{\theta}_y)$. By operating first derivative and second derivative to $\mathcal{J}(\boldsymbol{\theta}_y)$, the above equation can be rewriten as:

$$\hat{g}(\boldsymbol{L}, \boldsymbol{G}_x) = \sum_{x=1}^{X} \left\{ \frac{1}{N_x} \sum_{y \in \mathcal{N}_x} \min_{\boldsymbol{s}_y} \left[\|\boldsymbol{\alpha}_y - \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_y\|_{\boldsymbol{\Gamma}_y}^2 + \mu_1 \|\boldsymbol{s}_y\|_1 \right] + \mu_2 \|\boldsymbol{G}_x\|_{\mathrm{F}}^2 \right\} + \mu_3 \|\boldsymbol{L}\|_{\mathrm{F}}^2$$

where $\boldsymbol{\Gamma}_y$ is the Hessian matrix and $\|\boldsymbol{\alpha}_y - \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_y\|_{\boldsymbol{\Gamma}_y}^2 = (\boldsymbol{\alpha}_y - \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_y)^\top \boldsymbol{\Gamma}_y (\boldsymbol{\alpha}_y - \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_y)$. The constant term was ignored because it has no effect on the minimization. The linear term was ignored because the $\boldsymbol{\alpha}_y$ is the estimated optimal policy.

In further, we can split the above equation by all the equipment. Such that, we only optimize the equipment specific $\boldsymbol{\theta}_y$ while fix the value of $\boldsymbol{\theta}_y$ for all other equipment. The improvement of $\boldsymbol{L}$ and $\boldsymbol{G}_x$ can be reflected to other equipment. As such, we can obtain the update function of $\boldsymbol{L}$ and $\boldsymbol{G}_x$:

$$\Delta \boldsymbol{L}(k) = \beta_1 \left[\sum_{x=1}^{3} \frac{1}{|\mathcal{Z}_x(k)|} \sum_{z \in \mathcal{Z}_x(k)} \left(-\boldsymbol{G}_x^\top \boldsymbol{\Gamma}_z \boldsymbol{\alpha}_z \boldsymbol{s}_z + \boldsymbol{G}_x^\top \boldsymbol{\Gamma}_z \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_z \boldsymbol{s}_z^\top \right) + \mu_3 \boldsymbol{L} \right], \tag{12}$$

$$\Delta \boldsymbol{G}_x(k) = \beta_x \left[\frac{1}{|\mathcal{Z}_x(k)|} \sum_{z \in \mathcal{Z}_x(k)} \left(-\boldsymbol{\Gamma}_z \boldsymbol{\alpha}_z (\boldsymbol{L} \boldsymbol{s}_z)^\top + \boldsymbol{\Gamma}_z \boldsymbol{G}_x (\boldsymbol{L} \boldsymbol{s}_z)(\boldsymbol{L} \boldsymbol{s}_z)^\top + \mu_2 \boldsymbol{G}_x \right) \right], \tag{13}$$

where, β_1 and β_x, $\forall x \in \{1, 2, 3\}$, are the learning rates for $\boldsymbol{L}$ and $\boldsymbol{G}_x$, separately. $z \in \mathcal{Z}_x$ is the set of observed equipment for each group. Such that, $\boldsymbol{L}(k+1) = \boldsymbol{L}(k) + \Delta \boldsymbol{L}(k)$ and $\boldsymbol{G}_x(k+1) = \boldsymbol{G}_x(k) + \Delta \boldsymbol{G}_x(k)$, where k means k-th update step. With the updated global knowledge base and group base, $\boldsymbol{s}_y$ can be obtained by solving a Lasso:

$$\boldsymbol{s}_y(k+1) \leftarrow \arg\min_{\boldsymbol{s}_y(k)} \ell\left(\boldsymbol{G}_x \boldsymbol{L}, \boldsymbol{s}_y(k), \boldsymbol{\alpha}_y, \boldsymbol{\Gamma}_y\right), \tag{14}$$

where $\ell\left(\boldsymbol{G}_x \boldsymbol{L}, \boldsymbol{s}_y, \boldsymbol{\alpha}_y, \boldsymbol{\Gamma}_y\right) = \|\boldsymbol{\alpha}_y - \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_y\|_{\boldsymbol{\Gamma}_y}^2 + \mu_1 \|\boldsymbol{s}_y\|_1$. Consequently, the full algorithm can be organized as in Algorithm 1: (1) Initialize the $\boldsymbol{L}$, $\boldsymbol{G}_x$ and $\boldsymbol{\alpha}_y$ for all equipment. (2) Estimate $\boldsymbol{\alpha}_y$ for all equipment. (3) Randomly choose a piece of equipment y and update $\boldsymbol{L}$ and corresponding $\boldsymbol{G}_x$ using (12) and (13). (4) Compute $\boldsymbol{s}_y$ according to (14). (5) Repeat steps (3) and (4) until the time period comes to an end. It is worth noting that, at each step, we update only the global knowledge base and the corresponding group knowledge base. The performance improvement of equipment in other groups can benefit from the updating of the global knowledge base. In this case, we have $\mathcal{N} = \mathcal{Z}_1 \cup \mathcal{Z}_2 \cup \mathcal{Z}_3$ and $\boldsymbol{\theta}_y = \boldsymbol{G}_x \boldsymbol{L} \boldsymbol{s}_y$.

Algorithm 1 Overview of the Proposed Algorithm

Require: $T \leftarrow 0$, $\boldsymbol{L} \leftarrow \text{zeros}_{d,m}$, $\boldsymbol{G}_x \leftarrow \text{zeros}_{d_x,d}$, $\forall x \in \{1,2,3\}$
Require: $\boldsymbol{\alpha}_y$, $\boldsymbol{s}_y$ for all pieces of equipment
  while $t \leq T$ **do**
    Randomly choose a piece of equipment
    Identify the group of the chosen equipment as $x \in \{1,2,3\}$
    Obtain interaction history and compute $\boldsymbol{\Gamma}_y$
    Update $\boldsymbol{L}$, $\boldsymbol{G}_x$ using (12) and (13)
    Update $\boldsymbol{s}_y$ for device i using (14)
    $t \leftarrow t+1$
  end while

3.3. Computing Complexity

Each update begins with the computing of $\boldsymbol{\theta}_y$ and $\boldsymbol{\Gamma}_y$ for each individual equipment. We adopt a base-learner, specifically the episodic Natural Actor Critic (eNAC), characterized by a computational complexity of $O(\xi(d_x, n_t))$ for each step. Here, n_t represents the number of trajectories obtained for a piece of equipment during the current iteration. The update of $\boldsymbol{L}$ includes multiplication of matrix and vectors, which yields $O(d_x^3 + d_x dm + m^2)$ for each step. Similarly, the update of $\boldsymbol{G}_x$ has a complexity $O(d_x dm + m^2 + d_x^2)$. The update of $\boldsymbol{s}_y$ requires solving an instance of Lasso, which typically would be $O(d^3h^2 + md_x^2 + d_x m^2)$. Therefore, the overall complexity of each update for an individual equipment is $O(d_x^3 + d_x dm + md_x^2 + d_x m^2 + \xi(d_x, n_t))$.

4. Simulation Results

4.1. Simulation Settings

For our simulations, we consider a circular network area with a radius of 500 m and one BS at its center serving three groups of equipment. Each group is distributed in a circle with a radius of 250 m. Within each group, we consider $N_x = 10$ pieces of equipment uniformly distributed. For each group, we have the following simulation parameters:

- GP: This group of equipment relies solely on the grid power as the energy source. Therefore, there is no limit on the amount of energy could be utilized, i.e., $e_i(t) \in [0, \infty]$. For this group, we consider the state vector dimension as $d_1 = 2$.
- GS: This group of equipment is equipped with grean energy harvesting capabilities. For solar energy collection, we consider solar panels with the following parameters: $p^{\text{solar}} = 300$ W/m^2, $\epsilon_0 = 3.8$ cm $\times$ 9 cm, $\epsilon_1 = 50\%$, and $\epsilon_2 \in [0.5, 1.5]$. The collected energy is stored in batteries with maximum capacity $E_{\max} = 10$ J. The dimension of the state vector $d_2 = 3$ is considered in this group.
- MS: The equipment in this group relies on both the grid energy and the green energy source. For the grid energy source, there's no limit on the amount of energy could be utilized, i.e., $e_k^B(t) \in [0, \infty]$. For the green energy harvesting, we consider solar energy collection as in group GS. Similarly, the same solar panel parameters are considered here. Moreover, the collected energy is stored in the batteries with the maximum capacity $E_{\max} = 10$ J. We consider a state vector dimension of the $d_3 = 9$ for this group.

Moreover, for our three layered knowledge model, the values of d and m are obtained through validation experiment. In addition to the above parameters, the parameters shared by all parties pieces are listed below. We consider a bandwidth $B = 180$ kHz and noise power spectral density $N_0 = -174$ dBm/Hz. In addition, the loss of the channel is $g = 128.1 + 37.6 \log 10(l_y)$, where l_y (in km) and the standard deviation of shadow fading is 8 dB. With regard to computing energy consumption, we utilize associated values such as: $\varsigma = 10^{-27}$, $\kappa_y = 40$ and $\vartheta = 10^9$. We consider the harvested energy arrival probability $\sigma = 0.7$. For each piece of equipment, we assume the average size of the arrived data packets for each group is randomly generated from the range $[20, 60]$. Within each group, the $a_{y,t}$ follows a Gaussian distribution specifically. For all the equipment, we assume

$p_{y,\max} = 0.01W$ and $\Delta_{\max} = 30$. The simulations in the article were conducted using a MacBook Pro with an M1 chip. The code was executed on Matlab R2024a, and the MacOS system version used was Ventura 13.2.1.

4.2. Results and Analysis

For comparison purposes, two benchmark algorithms are compared with our proposed algorithm. The first is a Random strategy, which employs a randomly initialized policy and regular Policy Gradient (PG) updating method. To be specific, any PG methods capable of estimating policy gradient can be utilized, such as REINFORCE [23] and Natural Actor Critic (NAC) [24]. In our simulation, we adopt the NAC method as the base learning method. Additionally, we compare our proposed algorithm with the policy gradient efficient lifelong learning (PGELLA) algorithm [25]. PGELLA facilitates learning and knowledge sharing within each individual group, which is different from our cross-domain approach.

In Figure 4, we examine the initial performance improvement of the two algorithms, which we refer to as the warm start policy, compared to the random initial policy. Figure 4 shows the improvement of warm start policies of our proposed method and PGELLA over random initial policies. Both methods surpass the random initial policies. To be specific, the overall averaged results show that our algorithm can achieves 51.32% warm start policy improvement over random policy while PGELLA achieves 28.91% in general. As shown in Figure 4, for the devices in group GP, the proposed algorithm can provide a slight better performance compared to PGELLA. This is because the group GP, with lower complexity, requires less knowledge from other domains. Therefore, simpler intra-cluster knowledge sharing and migration models can provide satisfactory performance compared to cross-domain methods. Furthermore, for both the group GS and group MS, the proposed algorithm obviously outperforms PGELLA. Particularly for the MS group, the algorithm proposed can achieve a performance improvement of 90.04% compared to the random initial policy, while a 23.32% performance improvement can be achieved by PGELLA. The differing performance of the algorithm proposed and PGELLA across different groups can be attributed to the varying model complexities of these groups. The proposed algorithm enables cross-domain knowledge sharing, leading to a higher degree of exploration for complex models and the ability to migrate learned knowledge to new tasks. The enhanced performance of our method is attributed to its capacity to retain a broader spectrum of knowledge, while PGELLA is limited to providing insights specific to individual groups.

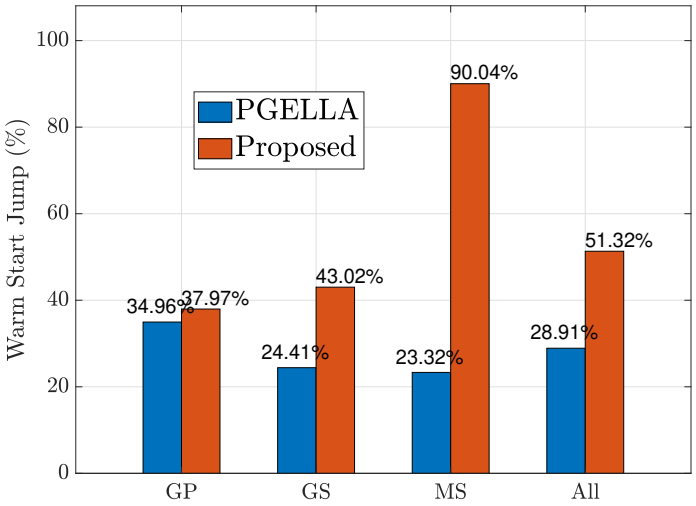

Figure 4. The warm start policy improvement for all the groups.

In Figure 5, a comprehensive overview of the learning trajectory is depicted across 1000 iterations for three groups. The implementation of more effective warm start policies can significantly reduce the convergence time for each group. This underscores the impact of initial policy selection on the efficiency of the learning process. In Figure 5a, it is evident that for group GP, both PGELLA and the proposed algorithm demonstrate improved warm start policies compared to the random initial policy. This enhancement effectively improves the performance of the initial policy. Additionally, both methods exhibit similar convergence speeds and final convergence performance. This similarity may be attributed to the simpler Markov models in group GP. These simpler models require less demanding algorithms to achieve satisfactory performance. In other words, PGELLA suffices for simpler models, despite lacking extensive knowledge collection and migration capabilities. Figure 5b illustrates that the enhanced capacity for broader exploration facilitates attaining globally optimal solutions. In Figure 5b, for group GS, our proposed algorithm significantly enhances the performance of the warm start policy and outperforms both PGELLA and random PG algorithm in terms of convergence speed. This improvement stems from the ability of the proposed algorithm to learn cross-domain knowledge and migrate accumulated knowledge from other domains to the current one, potentially achieving global optimality or sub-optimality and breaking out of local optima. It is worth noting that PGELLA also achieves superior results by enabling knowledge sharing within group devices, which helps overcome local optimality limitations. In Figure 5c, it can be observed that both PGELLA and the proposed algorithm achieve notable improvements in warm start policies compared to the random PG algorithm. Particularly, the proposed algorithm exhibits a significant enhancement, consistent with the performance of warm start depicted in Figure 4. Additionally, although PGELLA also contributes to the improvement in convergence rate, the algorithm proposed outperforms the other two algorithms significantly in terms of convergence speed. This is attributed to our proposed algorithm's capability in cross-domain knowledge transfer, enabling it to achieve greater performance improvements on complex models than on simpler ones. In Figure 5d, the average performance of different algorithms across three groups is presented. Based on the averaged results, it can be observed that overall, both in terms of warm start policies performance and convergence rate, our proposed algorithm outperforms PGELLA. In a nutshell, while the performance in each single group can vary, our algorithm shows better overall performance in respect to warm start policy and convergence speed.

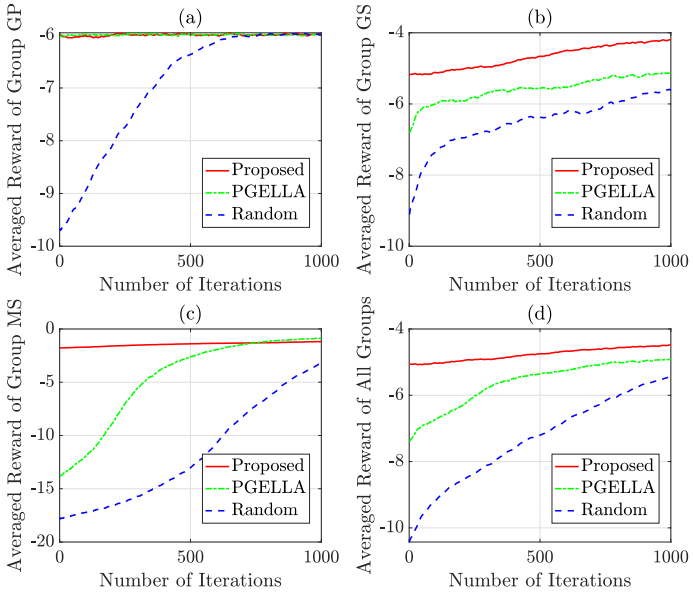

Figure 5. The learning methods comparison over 1000 iterations: (**a**) Group GP. (**b**) Group GS. (**c**) Group MS. (**d**) All groups.

Figure 6 considers the default groups and environmental settings as specified in Section 4.1. In this scenario, mixed energy sources, including grid power energy and green harvested energy, are considered to minimize grid energy consumption. In a nutshell, our proposed method can achieve a better balance between grid energy consumption, battery energy consumption and AoI. By contrast, the other methods, such as PGELLA and Random methods, fail to optimize the overall performance. This proves the sustainability improvement of our proposed method. To be specific, Figure 6 shows the performance attributes for the MS group, encompassing average AoI, queue length, grid energy consumption and battery energy consumption. This group is selected due to its model complexity, allowing for a more comprehensive comparison of the algorithms. Three algorithms are considered: random PG with a random initial policy, PGELLA with intra-group knowledge sharing capability, and the proposed algorithm, the cross-domain knowledge migration algorithm. Figure 6a demonstrates the significant impact of the proposed algorithm on reducing AoI as the number of learning steps increases. In contrast, algorithms with random initial policies or PGELLA exhibit inferior performance in this regard, albeit the latter showing initial superiority compared to random initial policy. Figure 6b reveals a similar trend in average queue length reduction across all the devices in the group MS with the algorithms proposed, indicating enhanced packet processing efficiency. Notably, queue length is closely correlated with AoI. A notable difference in grid energy consumption among the three algorithms is evident in Figure 6c. While both the random initial policy and PGELLA show a decrease in grid energy consumption with increasing learning steps, our proposed algorithms exhibits an increase. However, it's essential to highlight that this aligns with our goal of minimizing balanced cost. Our method achieves a better balance between AoI and grid energy consumption. Figure 6d indicates relatively consistent performance among the three algorithms in terms of battery energy consumption. The stability in battery energy consumption observed with the proposed algorithms is attributed to the significant performance enhancement with minimal energy consumption in the grid power network. Conversely, the battery energy consumption of the other two algorithms decreases with increasing learning steps, through with limited performance enhancement. Additionally, the PGELLA exhibits lower battery energy consumption than the random initial strategy.

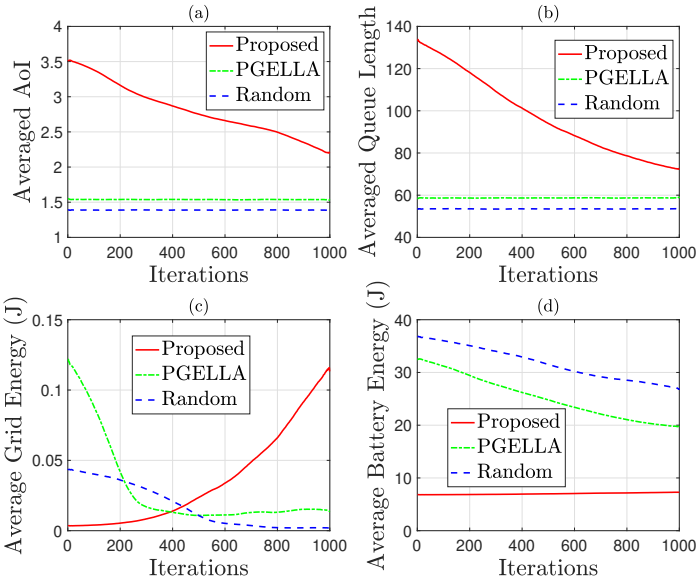

Figure 6. AoI and energy related performance comparison for group MS: (**a**) Average AoI. (**b**) Average queue length. (**c**) Average grid energy consumption. (**d**) Average battery energy consumption.

As depicted in Figure 7, the performance of group GS is evaluated in terms of average AoI and average battery energy consumption. From Figure 7a, it's evident that the performance of all three algorithms improve as the number of learning steps increases.

Particularly, the initial AoI of the random PG is lower than the other two algorithms, and this trend persists as the number of learning steps increases. In Figure 7b, the battery energy consumption of all three algorithms decreases with the increase in the number of learning steps. This reduction in energy consumption is accompanied by an increase in average AoI, indicating the optimization of multiple parameters rather than a single objective. Notably, the algorithms proposed demonstrate superior optimization for average energy consumption compared to the other two algorithms. While all algorithms exhibit a significant decrease in average energy consumption with increasing learning steps, the proposed algorithms perform the best in terms of overall performance improvement across multiple parameters, as demonstrated in Figure 5b.

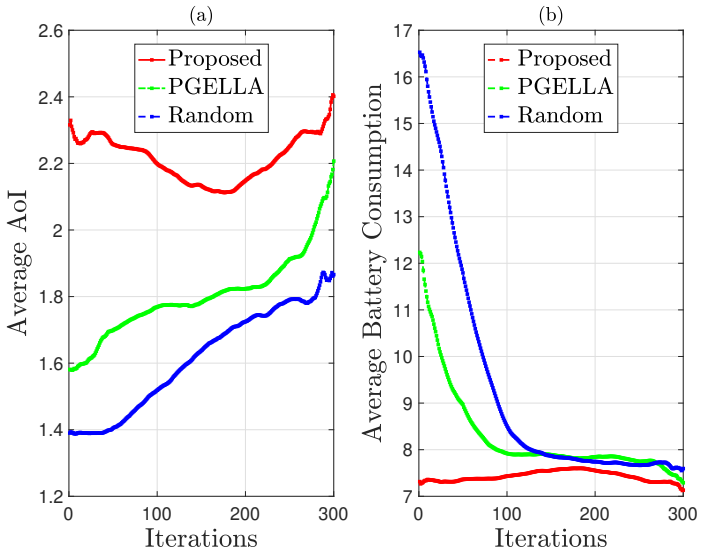

Figure 7. AoI and energy consumption comparison for group GS: (**a**) Average AoI. (**b**) Average total battery energy consumption.

4.3. Influence of the Number of Groups

As depicted in Figure 8, the impact of the number of groups on the proposed algorithm is illustrated. It can be observed that as the number of device groups increases, the proposed algorithm offers improved initial policies. This enhancement is attributed to the increased richness of knowledge in the knowledge base with each type of group, resulting in a more diverse global knowledge base. Furthermore, since the group knowledge base relies on the existence of the global knowledge base, optimizing the global knowledge base further enhances the performance of each group. Specifically, the warm start policy performance of the proposed algorithm improves from 29.73% for a single group to 51.32% for three groups compared to a random initial policy. Similarly, as the number of groups increases, the PGELLA algorithm maintains relatively stable performance, with the warm start policy performance ranging from 28.28% for one group to 29.91% for three groups. This is because, PGELLA, as an intra-cluster knowledge learning algorithm, does not perform knowledge sharing among groups, and its warm start policy performance variation is due to the differentiated performance of different groups. Nevertheless, our proposed algorithm still achieves a better warm start policy improvement than PGELLA, demonstrating its effectiveness in handling differentiated cluster data. Furthermore, when the number of groups is insufficient, the cross-domain knowledge sharing framework has less knowledge to abstract and share, leading to a degradation in performance compared to scenarios with a larger number of groups. As a result, as the number of groups increases, the advantages of the three-layer knowledge base framework proposed become apparent, significantly outperforming the performance of PGELLA. This highlights the advantages of the three-layer knowledge base framework in handling cross-domain knowledge trans-

fer. Therefore, the proposed algorithm exhibits more potential application scenarios and advantages when the number of groups is high.

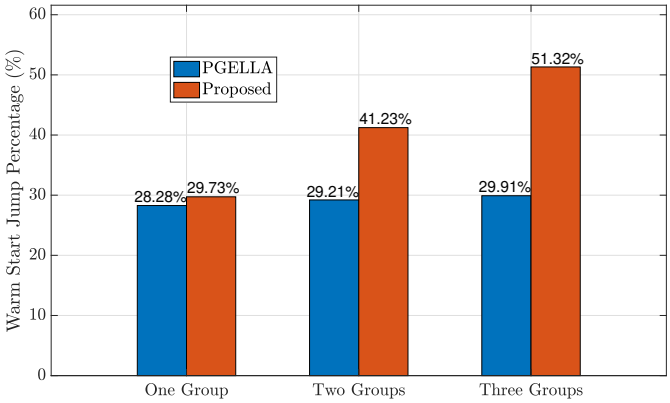

Figure 8. Warm start policy improvement when the number of groups increases.

As depicted in Table 1, the table provides a comparison of the running time and the running time difference between two algorithms: PGELLA and our proposed algorithm, for varying numbers of groups. From the table, it's evident that as the number of groups increases, the running time of both algorithms also increases approximately linearly. With PGELLA, each additional group necessitates a complete repetition of the algorithm's process. On the other hand, our proposed algorithm requires visiting a higher number of devices with each additional group, resulting in a longer time to visit all devices compared to PGELLA. Hence, with a single group, the runtime of our proposed algorithm (3.9225 s) is less than that of PGELLA. Yet, as the number of groups rises, PGELLA's runtime progressively diminishes compared to our proposed algorithm. In particular, when there are three groups, PGELLA's runtime surpasses that of the proposed algorithm by 0.5931 s. Combining the findings from Figure 8, we observe that our proposed algorithm achieves approximately a 15% performance enhancement, with a mere 4.94% increase in runtime. This suggests a perfect balance between runtime efficiency and performance improvement.

Table 1. The Running Time Comparison.

Running Time (Seconds)	One Group	Two Groups	Three Groups
CD	3.9225	9.8620	12.5984
PG	4.0014	9.4575	12.0053
Gap	−0.0793	0.4044	0.5931

5. Conclusions

The article introduces a lightweight cross-domain knowledge sharing model leveraging diverse energy supply methods. It employs a three-layered knowledge base, incorporating global, group-specific, and individual policy vectors. By integrating grid, harvested, and mixed energy sources, significant improvements in warm start policy performance are demonstrated compared to random initial policies. Moreover, the collaborative nature of the global knowledge base contributes to enhanced sustainability, surpassing that of two-layered models. Considering the significant energy savings and AoI optimization achieved, our approach can facilitate the sustainability of IIoT and Industrial 4.0 initiatives. To advance in the field of cross-domain knowledge sharing, several potential research directions can be considered. These include investigating the impact of mobility, addressing privacy and security concerns, and exploring the integration of edge computing.

Author Contributions: Conceptualization, Z.G. and Q.C.; methodology, Z.G. and W.N.; software, Q.C.; validation, Z.G., Q.C. and W.N.; investigation, W.N.; resources, Q.C.; data curation, Z.G.; writing—original draft preparation, Z.G.; writing—review and editing, W.N.; visualization, Z.G.; supervision, Q.C.; project administration, W.N.; funding acquisition, Q.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Regional Innovation and Development of the National Natural Science Foundation of China (No. U21A20449).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IIoT	Industrial Internet-of-Things
IoT	Internet of Things
AoI	Age of Information
SDGs	Sustainable Development Goals
RL	Reinforcement Learning
GD	Gradient Descent
GP	Graid Power
GS	Green Source
MS	Mixed Sources
BS	Base Station
CPU	Central Processing Unit
FCFS	Firsy-Come-Firsy-Served
MDP	Markov Decision Processes
MTL	Multi-Task Learning
eNAC	episodic Natural Actor Critic
PGELLA	Policy Gradient Efficient Lifelong Learning Algorithm
PG	Policy Gradient
CD	Cross Domain

References

1. Saad, W.; Bennis, M.; Chen, M. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Netw.* **2019**, *34*, 134–142. [CrossRef]
2. The Sustainable Development Goals Report. Available online: https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf (accessed on 6 April 2024).
3. Li, X.; Cui, Q.; Feng, D.; Gong, Z.; Tao, X. Deep Reinforcement Learning-Based Solution for Minimizing the Alterable Urgency of Information in UAV-Enabled IIoT System. In Proceedings of the GLOBECOM 2023—2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 4–8 December 2023; IEEE: New York, NY, USA, 2023; pp. 437–442.
4. Malik, P.K.; Sharma, R.; Singh, R.; Gehlot, A.; Satapathy, S.C.; Alnumay, W.S.; Pelusi, D.; Ghosh, U.; Nayak, J. Industrial Internet of Things and its applications in industry 4.0: State of the art. *Comput. Commun.* **2021**, *166*, 125–139. [CrossRef]
5. Moloudian, G.; Hosseinifard, M.; Kumar, S.; Simorangkir, R.B.; Buckley, J.L.; Song, C.; Fantoni, G.; O'Flynn, B. RF Energy Harvesting Techniques for Battery-less Wireless Sensing, Industry 4.0 and Internet of Things: A Review. *IEEE Sens. J.* **2024**, *24*, 5732–5745. [CrossRef]
6. Folgado, F.J.; Calderón, D.; González, I.; Calderón, A.J. Review of Industry 4.0 from the Perspective of Automation and Supervision Systems: Definitions, Architectures and Recent Trends. *Electronics* **2024**, *13*, 782. [CrossRef]
7. Cortés-Leal, A.; Cárdenas, C.; Del-Valle-Soto, C. Maintenance 5.0: Towards a Worker-in-the-Loop Framework for Resilient Smart Manufacturing. *Appl. Sci.* **2022**, *12*, 11330. [CrossRef]
8. Dileep, G. A survey on smart grid technologies and applications. *Renew. Energy* **2020**, *146*, 2589–2625. [CrossRef]
9. Zhang, H.; Lu, Y.; Han, W.; Zhu, J.; Zhang, Y.; Huang, W. Solar energy conversion and utilization: Towards the emerging photo-electrochemical devices based on perovskite photovoltaics. *Chem. Eng. J.* **2020**, *393*, 124766. [CrossRef]

10. Ibrahim, H.H.; Singh, M.J.; Al-Bawri, S.S.; Ibrahim, S.K.; Islam, M.T.; Alzamil, A.; Islam, M.S. Radio frequency energy harvesting technologies: A comprehensive review on designing, methodologies, and potential applications. *Sensors* **2022**, *22*, 4144. [CrossRef] [PubMed]
11. Kaul, S.; Yates, R.; Gruteser, M. Real-time status: How often should one update? In Proceedings of the 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 25–30 March 2012; IEEE: New York, NY, USA, 2012; pp. 2731–2735.
12. Dang, Q.; Cui, Q.; Gong, Z.; Zhang, X.; Huang, X.; Tao, X. AoI oriented UAV trajectory planning in wireless powered IoT networks. In Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022; IEEE: New York, NY, USA, 2022; pp. 884–889.
13. Hamdi, R.; Baccour, E.; Erbad, A.; Qaraqe, M.; Hamdi, M. LoRa-RL: Deep reinforcement learning for resource management in hybrid energy LoRa wireless networks. *IEEE Internet Things J.* **2021**, *9*, 6458–6476. [CrossRef]
14. Ye, J.; Gharavi, H. Deep reinforcement learning-assisted energy harvesting wireless networks. *IEEE Trans. Green Commun. Netw.* **2020**, *5*, 990–1002. [CrossRef] [PubMed]
15. Gong, Z.; Cui, Q.; Chaccour, C.; Zhou, B.; Chen, M.; Saad, W. Lifelong learning for minimizing age of information in Internet of Things networks. In Proceedings of the ICC 2021-IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
16. Gong, Z.; Hashash, O.; Wang, Y.; Cui, Q.; Ni, W.; Saad, W.; Sakaguchi, K. UAV-aided lifelong learning for AoI and energy optimization in non-stationary IoT networks. *arXiv* **2023**, arXiv:2312.00334.
17. Kolter, J.Z.; Maloof, M.A. Dynamic weighted majority: An ensemble method for drifting concepts. *J. Mach. Learn. Res.* **2007**, *8*, 2755–2790.
18. Wu, Q.; Iyer, N.; Wang, H. Learning contextual bandits in a non-stationary environment. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA , 8–12 July 2018; pp. 495–504.
19. Yu, A.; Yang, Q.; Dou, L.; Cheriet, M. Federated imitation learning: A cross-domain knowledge sharing framework for traffic scheduling in 6G ubiquitous IoT. *IEEE Netw.* **2021**, *35*, 136–142. [CrossRef]
20. Cui, Q.; Gong, Z.; Ni, W.; Hou, Y.; Chen, X.; Tao, X.; Zhang, P. Stochastic online learning for mobile edge computing: Learning from changes. *IEEE Commun. Mag.* **2019**, *57*, 63–69. [CrossRef]
21. Pan, Y.; Pan, C.; Yang, Z.; Chen, M. Resource allocation for D2D communications underlaying a NOMA-based cellular network. *IEEE Wirel. Commun. Lett.* **2017**, *7*, 130–133. [CrossRef]
22. Lu, X.; Wang, P.; Niyato, D.; Kim, D.I.; Han, Z. Wireless networks with RF energy harvesting: A contemporary survey. *IEEE Commun. Surv. Tutorials* **2014**, *17*, 757–789. [CrossRef]
23. Williams, R.J. Simple Statistical Gradient-following Algorithms for Connectionist Reinforcement Learning. *Mach. Learn.* **1992**, *8*, 229–256. [CrossRef]
24. Peters, J.; Schaal, S. Natural Actor-critic. *Neurocomputing* **2008**, *71*, 1180–1190. [CrossRef]
25. Ammar, H.B.; Eaton, E.; Ruvolo, P.; Taylor, M. Online multi-task learning for policy gradient methods. In Proceedings of the International Conference on Machine Learning. PMLR, Beijing, China, 21–26 June 2014; pp. 1206–1214.