Attack-Resistant, Energy-Adaptive Monitoring for Smart Farms: Uncertainty-Aware Deep Reinforcement Learning Approach

Qisheng Zhang, *Student Member, IEEE,* Dian Chen, Yash Mahajan, Ing-Ray Chen, *Member, IEEE,* Dong Sam Ha, *Fellow, IEEE,* and Jin-Hee Cho, *Senior Member, IEEE*

***Abstract*—This work proposes an energy-adaptive monitoring system for a smart farm using solar sensors attached to cows. The proposed system aims to achieve a high monitoring quality in the smart farm under fluctuating energy and cyber attacks disrupting the collection of sensed data from solar sensors, such as protocol non-compliance, false data injection, denial-of-service, and state manipulation. We adopt Subjective Logic, a belief model, to consider multidimensional uncertainty in sensed data. We employ Deep Reinforcement Learning (DRL) for agents on gateways to collect high-quality sensed data from the solar sensors. The DRL agents aim to collect high-quality sensed data with low uncertainty and high freshness under fluctuating energy levels in solar sensors. We analyze the performance of the proposed energy-adaptive smart farm system in accumulated reward, monitoring error rate, and system overload. We conduct a comparative performance analysis of the uncertainty-aware DRL algorithms against their counterparts in choosing the number of sensed data to be updated to collect high-quality sensed data to achieve high resilience against attacks. Our results prove that Multi-Agent Proximal Policy Optimization (MAPPO) using the uncertainty maximization technique outperforms other counterparts, showing about 4% lower monitoring error rate and the system overload.**

***Index Terms*—Smart farm, energy-adaptive, deep reinforcement learning, solar sensors, uncertainty, cyberattacks.**

I. Introduction

ACCORDING to the Food and Agriculture Organization (FAO) of the United Nations [17], the food production rate must increase by a factor of up to 70 percent to absorb the increase in population, estimated as more than 9 billion people by 2050 [44]. To support farms' productivity, flexibility, or availability, smart farm technologies have developed by leveraging sensors, Internet-of-Things (IoT), and edge and cloud computing technologies. Smart farm research is applied to develop agricultural business practices [10], improve monitoring of animal welfare [47], and provide data sensing and environmental controls [47]. Smart farm research also investigated efficient data transmission considering CPU usage, signal strength, and battery operation time [23] for wireless sensor networks [39]. However, existing research lacks secure solutions for wireless sensor networks with energy constraints, such as solar energy harvesting [1].

According to the World Health Organization (WHO), over half a million people died from food contamination caused by bacteria, viruses, toxins, or chemicals. Cyber attacks on farms, transportation systems, and food processing industrial control systems to distort and disrupt correct data handling can worsen the problem. Any distortion in the data received from the livestock [2] monitoring systems can lead to serious situations such as the spread of disease, possible pandemics, and the provision of wrong information to potential customers of the livestock [19].

In this work, we are interested in improving the accuracy of the livestock monitoring system in farms under the presence of cyber attacks that can forge, modify, or drop sensed data from sensors to gateways or edge devices or inject false data. Most wireless sensor networks (WSNs) cannot accurately record cattle biometrics because battery-powered sensors attached to the cattle collar can last only a few days or weeks. Frequent replacement or recharging of batteries for sensor nodes is laborious for a farm with many animals. To address this problem, we consider wireless solar sensor nodes attached to the livestock's ears powered by solar energy harvesting. Due to the small size of the solar panel attached to a sensor node, the amount of solar energy harvested is low. Moreover, the harvested energy level fluctuates as the livestock and/or its ear moves, making sensing and transmission of the data unstable.

To address the problem, we consider sensor nodes adopting two communication protocols, LoRa (Long Range) [49] and BLE (Bluetooth Low Energy) [53]. Note that we consider a solar sensor-based smart farm environment because of the following reasons. Jawad et al. [24] discusses that solar energy based on photovoltaic systems and methods has been the

 This research was partly sponsored by the National Science Foundation (NSF) under Grant Numbers CNS-2106987 and III-2107450. In addition, this research is also partly supported by Virginia Tech's Institute for Critical Technology and Applied Science (ICTAS) The Engineering Faculty Organization (EFO) Opportunity Seed Investment Grant, and The Commonwealth Cyber Initiative (CCI) Southwest Virginia Node Research, Innovation, & Workforce Development Funding. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein (*Corresponding author: Jin-Hee Cho*). Qisheng Zhang, Dian Chen, Yash Mahajan, Ing-Ray Chen, and Jin-Hee Cho are with the Department of Computer Science, Virginia Tech, Falls Church, VA, USA. Dong Sam Ha is with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. E-mail: {qishengz19, dianc, yashmahajan, irchen, dha, jicho}@vt.edu.

[1]Energy harvesting means the sensor nodes harvest different types of energy to prolong their lifetime.

[2]Livestock is the domesticated animals raised in an agricultural setting, e.g., cows, pigs, and sheep.

most popular agriculture-based energy-harvesting technique. In addition, solar cells provide a good solution to ensure the survivability of the agriculture monitoring system [58], which also has been used in Virginia Tech's smart farm testbed. The LoRa protocol aims for long-distance communication with a range of several kms, but the data rate is only 27 $kbps$. Contrarily, the BLE protocol aims for a short distance with a line-of-sight distance of 100 $meters$ and a higher data rate of 2 $Mbps$. Furthermore, the BLE protocol drains considerably less power than the LoRa one. Fig. 1 shows the WSN system considered in this paper. A sensor node monitors the energy level of its battery. A sensor node with a high energy level transmits its sensed data with LoRa to LoRa gateways, and the gateways upload the data to the cloud server accessible to the user. A sensor node with a low energy level may be unable to send its data directly to a LoRa gateway due to insufficient energy. Instead, the sensor node seeks a nearby sensor node with a high energy level. If it finds one, the sensor node sends its data by BLE to the nearby sensor node with a high energy level, and the nearby sensor node transmits the received data to LoRa gateways.

Under this scenario, since there may be multiple sensor nodes with low energy requesting to transmit their sensed data to the sensor node with high energy, a decision on which sensed data to transmit to the gateways can significantly impact the accuracy of monitored data. Instead of continuously transmitting sensed data that are already received sufficiently and hence high quality (i.e., low uncertainty), a sensor node with high energy can select a sensor node with low energy whose data have high uncertainty, and transmit the data, resulting increase in the data certainty. The process will significantly increase the certainty of the overall data monitored from all animals on the farm. To this end, we introduce an uncertainty-aware transmission policy based on the assessment by LoRa gateways. Specifically, a LoRa gateway can request sensor nodes to send sensed data of particular animals whose monitored data have trended high uncertainty (i.e., low certainty). In this work, we leverage deep reinforcement learning (DRL) to identify sensor nodes whose data need to be transmitted to improve the overall monitoring accuracy.

This paper substantially extends our prior work [56] by introducing the following **additional key contributions**:

- We propose an energy-adaptive monitoring system for WSN-based smart farms with solar-powered sensor nodes attached to cattle. This is the first work that considers how WSN-based smart farms can maintain high monitoring quality under limited and fluctuating energy availability due to solar energy harvesting in the smart farm.
- We develop uncertainty-aware DRL algorithms based on *Deep Reinforcement Learning* (DRL) [15] and *Subjective Logic* [25] (SL) to minimize the overall monitoring error and system overload for sensor nodes. The uncertainty in SL-based opinions is measured by *vacuity* due to a lack of evidence and *dissonance* due to conflicting evidence. This uncertainty measurement (i.e., vacuity and dissonance in SL-based opinions) represents uncertainty in data representing epistemic uncertainty. Note that other uncertainty measures, such as entropy or variance, represent aleatoric uncertainty, capturing uncertainty from statistical randomness.
- We provide mathematical proof to justify how SL's uncertainty maximization (UM) technique contributes to reducing monitoring errors. Our theoretical analysis proved that using the UM technique allows applying newly arrived evidence effectively, reflecting recent network dynamics properly.
- We consider a comprehensive attack model consisting of diverse cyber attacks (e.g., non-compliance to the protocol, false data injection, and Denial-of-Service attacks) as well as adversarial examples (e.g., four different state-of-the-art data poisoning attacks) disrupting deep learning models.
- We devise a novel *monitoring error rate metric* to evaluate the monitoring quality independent of monitoring data distributions. The developed monitoring error rate metric enables our proposed monitoring system to handle multi-dimensional heterogeneous data simultaneously.
- We identify an optimal deployment setting of LoRa gateways on which DRL agents run to maximize the chances for solar sensors to deliver their sensed data within the gateway wireless radio range.
- We provide the asymptotic complexity analysis of our proposed uncertainty-aware DRL-based algorithms. This analysis reveals a critical tradeoff between robustness/effectiveness vs. efficiency.
- We validate the performance of the proposed uncertainty-aware DRL-based monitoring system using real datasets obtained from Virginia Tech's Smart Farm Innovation Network. Furthermore, we design a framework where healthy sensor nodes generate synthetic datasets similar to real datasets, and compromised sensor nodes are modeled as attackers following the attack model for testing the robustness of our uncertainty-aware DRL-based algorithms against adversarial attacks.
- We conduct extensive sensitivity analyses to investigate the effect of key designs and environmental factors on the performance of two proposed uncertainty-aware DRL-based algorithms (i.e., deep Q-learning, DQN, and multi-agent proximal policy optimization, MAPPO) against three baseline models (i.e., greedy, random, data mitigation), including the attack severity, the initial sensor node energy level, the number of solar sensors, and the chance for sensor nodes to be exposed to the sun.

The rest of this paper is structured as follows. Section II provides a brief overview of the related work. Section IV describes the network model, node model, and attack model considered in this work. Section V describes our proposed uncertainty-aware DRL-based algorithms for animal monitoring. Section VI explains an experiment setup including datasets, parameterization of key design parameters, performance metrics, and baseline schemes considered for comparative performance analysis. Section VII demonstrates the key experimental results and their overall trends, along with the physical interpretations of the observed results. Section VIII concludes the paper and suggests future work directions.

II. Related Work

In this section, we provide a brief overview of related work in DRL-based optimization of WSNs, energy-adaptive smart

environments, and uncertainty-aware smart environments.

A. DRL-based Optimization of WSNs & Smart Farms

Yoon et al. [55] improved the scalability and usability of the smart farm system by leveraging the low-power Bluetooth and Low Power Wide Area Networks (LPWAN) communication modules. With the implementation of monitoring and control functions, the authors showed promising directions for developing the agricultural IoT in various applications. Moreover, Lau et al. [29] proposed a data processing model for smart city environments with renewable wireless sensor network (RWSN) and pyroelectric infrared (PIR) sound sensors. The proposed system prevents PIR sensors from false alarms under extreme weather and environments. Boursianis et al. [6] employed the UAV technology in smart farm environments for remote monitoring and management of agriculture. However, the existing approaches cannot handle complex issues in agriculture. Adelantado et al. [2] demonstrated the challenges of using LoRaWAN to manage many end devices for maximizing the operation period in real-time environments. Thus, the scalability of LoRaWAN technologies still remains unsolved.

DRL-based approaches have been shown in using WSNs and IoT technologies for smart farms. Energy-aware WSNs have been proposed in various WSN applications, including routing [27, 31], resource management [42], power control [7, 8, 37], and system/hybrid design [59, 54, 50]. A cluster-based routing protocol was proposed based on a Q-learning approach called *QL-Cluster* [27]. The QL-Cluster was designed to identify the best routes between individual nodes and remote healthcare stations to efficiently monitor a patient's health. Qi et al. [42] proposed an adaptive energy management strategy for a solar-powered WSN with hybrid storage, consisting of super-capacitors and batteries, to avoid high current charging/discharge of the batteries and fully utilize the super-capacitors.

Chen et al. [7, 8] leveraged DRL with Q-learning to control power for communications between the in-body sensor and Wireless Body Area Networks (WBANs) coordinator to build jamming attack-resistant healthcare applications. Their research aimed to develop a WBAN coordinator that chooses the sensor to transmit the data in the next time slot and decides the transmitting power of these sensors, which is then sent to the sensor. The WBAN coordinator uses Q-learning to achieve an optimal power control strategy. Naderializadeh et al. [37] used a multi-agent deep RL approach to tackle the problem of distributed user scheduling and downlink power control in multi-cell wireless networks. Compared against several decentralized and centralized baseline counterparts, the authors showed that the proposed algorithm outperforms two decentralized approaches while performing comparably to the centralized scheduling algorithms. Moreover, the agents are trained for a specific environment. However, the approach is only scalable to different network configurations.

Alonso et al. [5] proposed a Double Deep-Q Learning approach to optimize the dataflows in the networks under smart farming and smart energy environments. They deployed the Double Deep-Q Networks on software-defined networking architectures to decrease the latency and data integrity with bandwidth. Nguyen et al. [38] developed a Deep Q-Learning-based algorithm for a smart farm deployed with IoT cameras, UVAs, and a mobile edge computing (MEC) server. This work aimed to determine an optimal policy to address the maximum number of tasks while utilizing the minimum energy.

As discussed above, DRL has been applied to develop energy-efficient WSNs where privacy concerns are considered in applying deep neural networks (DNNs). However, no prior work has been done for energy-adaptive smart farms with solar sensors, which is tackled in this work. Specifically, we develop uncertainty-aware DRL algorithms to maximize uncertainty-aware monitoring quality with sensors having limited and fluctuating energy harvesting.

B. Energy-Adaptive Smart Environments

Igder et al. [22] introduced an energy-adaptive Fog Server that could handle all requests at the same time under low-energy conditions with a limited number of requests. Popa et al. [40] proposed an intelligent platform and DNN-based models to achieve energy-efficient smart environments. The authors applied two techniques, namely, energy load forecasting and non-intrusive load monitoring, for learning while reducing energy consumption. In a smart home environment, the former was used to predict patterns of energy consumption and identify unusual energy usage, while the latter was used to further find which appliance caused the anomaly to provide energy-saving tips. Modarresi and Symons [36] proposed a multidimensional framework with an instance of a high-level smart home network. They argued that the diversity of services affects routing levels while providing routing services for both high and low energy consumption. However, their evaluation did not include the evaluation of their approach in terms of resilience to various cyber attacks.

The works above focused on learning and predicting energy usage to minimize energy consumption. On the other hand, our work focuses on learning and deriving energy adaptation strategies for LoRa gateways to obtain sensed data from solar sensors having limited and fluctuating energy, to maximize uncertainty-aware monitoring quality.

C. Uncertainty-Aware Smart Environments

Due to the dynamic nature of the multi-sensor smart environments, uncertainty, and ambiguity can significantly impact on the data prediction and monitoring of these smart environments. Zhang et al. [57] proposed learning the inhabitant's activity patterns in a smart home environment to learn under uncertainty caused by sensor malfunctions. Alemdar et al. [3] proposed an uncertainty sampling-based active learning method that considers three different measures of uncertainty to select the most informative data points for activity recognition in smart homes. Several research works [4, 11, 12, 13, 20, 32] have been conducted on uncertainty in context-aware systems where the environment is well-defined. Various approaches have been proposed to model uncertainty, including semantic web [32], game theory [11], vector space

model [41], asymptotic equipartition property (AEP) [45], signal processing and information-theoretic techniques [52], and Moore finite state machine (FSM) [43]. Machado et al. [32] proposed a contextual reference model based on the semantic web to deal with uncertainty in a smart environment. Rocher et al. [43] proposed a framework for estimating behavior drift in smart-X systems at runtime. They leveraged Moore finite state machine (FSM) model combined with control theory and validated their approach based on a real dataset to ensure effectiveness and efficiency.

Unlike the cited works above, we consider multiple types of uncertainty and the uncertainty maximization technique in Subjective Logic (SL) [26] for monitoring data updates based on new evidence, thus maximizing monitoring quality while minimizing energy consumption.

III. Problem Statement

In this work, we aim to minimize the monitoring error rate (i.e., a gap between the sensed data aggregated from sensors and the ground truth; see Eq. (2)) and system overload (i.e., a mean fraction of the failed requests of all requests sent from low-energy sensors; see Eq. (3)) in a sensor network by identifying an *optimal policy*. An updated policy $\mathcal{P}_T = \{p_1, p_2, \ldots, p_T\}$ contains monitoring actions p_i, where $i \in [1, T]$, $p_i \in \mathcal{P}_T$ and $\mathcal{P}_T$ is a set of monitoring actions available to the sensor in every monitoring step. Given a dynamic sensor network $\mathcal{G}_T = \{g_1, \ldots, g_i, \ldots, g_T\}$, the objective function is defined by:

$$\begin{aligned} &\arg\max_{\mathcal{P}_T} \sum_{i=1}^{T} f(g_i(p_1, p_2, \ldots, p_i)), \\ &s.t. \quad \forall i \in [1, T], p_i \in \mathcal{P}_T, \end{aligned} \tag{1}$$

where $f(g)$ is based on the evaluation function $f : g \mapsto -\mathcal{ME}(g) - \mathcal{OL}(g)$, aiming to minimize the monitoring error rate $\mathcal{ME}$ and system overload $\mathcal{OL}$ defined as follows:

$$\mathcal{ME} = \frac{\sum_{t\in T}\sum_{x\in X} \mathrm{me}_t^x}{NT|X|}, \quad \mathrm{me}_t^x = \sum_{j=1}^{n} D(\mathrm{eo}_t^x(j), \mathrm{gt}_t^x(j)), \tag{2}$$

$$s.t. \quad D(a,b) = \begin{cases} 1 & \text{if } a \neq b; \\ 0 & \text{if } a = b. \end{cases}$$

Here T is the total monitoring time, N is the number of animals, me_t^x is the overall monitoring error rate of all N animals' conditions of attribute x at time t, $\mathrm{eo}_t^x(j)$, and $\mathrm{gt}_t^x(j)$ are the estimated and ground truth observation of animal j's condition in x attribute at time t, respectively.

$$\mathcal{OL} = \frac{1}{T}\sum_{t\in T} \frac{\mathrm{rq}_t^f}{\mathrm{rq}_t}, \tag{3}$$

where T is the total monitoring time, rq_t^f and rq_t are the numbers of failed requests and total requests at time t, respectively. Determining an optimal update policy to achieve multiple objectives is non-trivial, given the complexity of solving a multi-objective optimization problem [9]. This is detailed based on the experimental results as shown in Section VII.

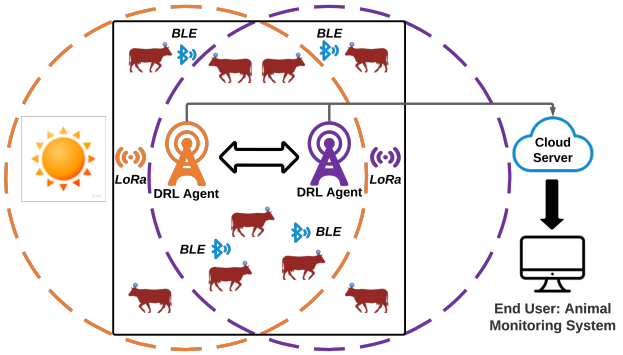


Fig. 1: Wireless solar sensor node-based smart farm network.

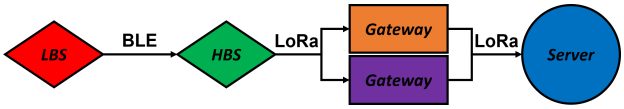


Fig. 2: The overall data flow of the smart farm network.

IV. System Model

This section discusses the network, node, and attack models.

A. Network Model

Our target WSN consists of solar sensors attached to the cattle that continuously measure the bio-metric information and transmit the sensed data to the LoRa gateway, which then aggregates and transmits the clustered data to the cloud. Given the relatively low cost of transmitting data over long ranges (LoRa) via the standard IP protocol for IoT devices, the LoRa gateways act as the optimal intermediary between the sensor nodes and the cloud server. In the given smart farm network (see Fig. 1), using BLE (Bluetooth Low Energy), each low-battery sensor (LBS) can relay its sensed data to one of the high-battery sensors (HBS). The high-battery sensor can then send the received sensed data and its own data to the LoRa gateway via LoRa. Fig. 2 shows the overall data flow. We assume that each sensor has a Microchip SAM R34/35 microcontroller with an embedded LoRa radio which dissipates 170 mW during transmission, while the microcontroller dissipates only 8 mW in active mode. The LoRa protocol is ideal for long-distance communication with a distance coverage of several kms and a data rate of 27 $kbps$. Contrarily, the BLE protocol is purposed for short-distance communication with a radius coverage of 100 $meters$ and a data rate of 2 $Mbps$. Furthermore, the BLE protocol drains considerably less power than the LoRa protocol. For example, only 11 mW of power is dissipated during transmission for a Texas Instruments CC2640R2F micro-controller chip with an embedded BLE radio [51]. Therefore, sending a single bit of data takes about 1,100 times less energy for the Texas Instruments CC2640R2F microcontroller chip when compared with the LoRa radio of the SAM R34/35 microcontroller chip. We assume that the initial battery level of each deployed sensor is 5 kWs, equivalent to a full charge. Outdoor solar has a power density of about 10 mW/cm^2 whereas indoor light has a power density of 0.1 mW/cm^2 [34]. For a solar panel of 5 cm, the maximum harvestable power for indoor light is about 2 mW, and outdoor light is about 200 mW. Fig. 1 describes

the considered network model in this work, describing a smart farm environment with solar sensors attached to the cattle.

To minimize the monitoring error rate and system overload, a DRL agent is deployed at every LoRa gateway to shortlist, select and prioritize which animal's sensed data is required at regular intervals. The process of identifying the important data by the DRL agent is described in Section V. We assume that for energy saving, there is no encryption when the sensors communicate with each other via BLE, and hence, malicious entities can intercept the data in transmission and modify/forge data. Additionally, an attacker can imitate a sensor by using its authentication key with the gateway and sending false data for the sensor itself as well as for other low-battery sensors. We assume that the LoRa gateway and the cloud server communication is secure and encrypted based on traditional secret cryptography. Note that the current LoRa technology uses a symmetric key cryptographic technique with AES-128 bit encryption [48]. As shown in Fig. 1, multiple LoRa gateways, each running a DRL agent, can collaborate with each other in sharing collected sensed data received from sensors.

B. Node Model

Sensor nodes in the smart environment are assumed solar-powered and deployed as implants and can transmit data on request. Depending upon the animal's movement and the varying weather condition from day to day, the battery levels of the sensor may fluctuate throughout the day. Therefore, it is essential to utilize the energy wisely for high availability, consistency, and sustenance. Each sensor node i is characterized by $\mathrm{sn}_i^t = [\mathrm{temp}_i^t, \mathrm{hb}_i^t, \mathrm{ma}_i^t, \mathrm{bl}_i^t]$, where temp_i^t is sensor node i's temperature at time t in Celsius, hb_i^t is the number of i's heartbeat at time t, ma_i^t is i's speed at time t, and bl_i^t is i's battery life at time t scaled in $[0, 100]$ in percent. Most sensors' energy will be used to transmit the sensed data to the LoRa gateway. In contrast, considerably less battery will be used for communication between the sensor and other nearby sensors via BLE as it consumes roughly 1,100 times less, as detailed in Section IV-A.

Utilizing the data reported by the sensors to the LoRa gateway, each DRL agent will try to maximize the monitoring quality by selecting what data is needed with priority to accurately estimate the well-being of all the animals on the farm. To this end, the sensor nodes in the WSN are categorized into high battery-level sensors (HBS) and low battery-level sensors (LBS) based on the recommended battery level T_M. Since we are only interested in transmissions from LBS to HBS, we model the sensor network as a directed bipartite graph. Section V describes the actions performed by the DRL agent running on every LoRa gateway. The end-user will get the efficient monitoring results of the smart farm from the cloud server, which aggregates data on individual animals from various LoRa gateways. This work aims to evaluate how the DRL agent on LoRa gateways can enhance the quality of animal monitoring in the presence of cyber attacks and fluctuating sensor energy levels.

C. Attack Model

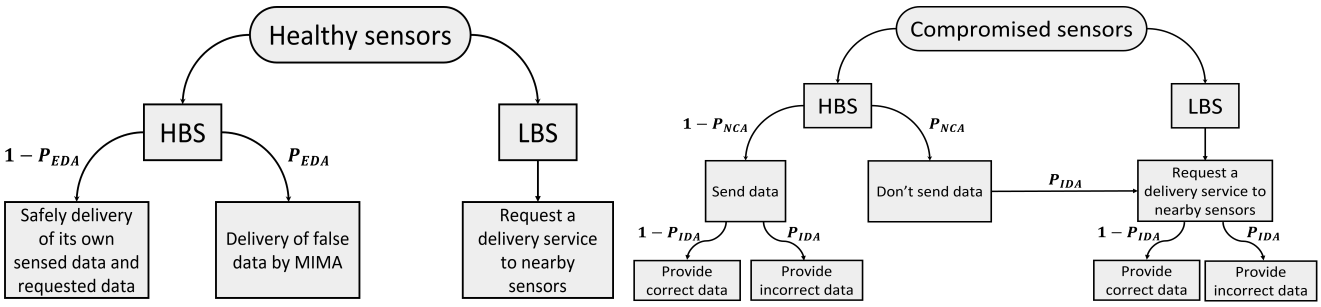
This work considers the two classes of attacks in terms of cyber attacks and adversarial examples. First, we consider the following **cyber attack** behaviors:

- *Non-compliance to the protocol*: A sensor node can be compromised and exhibit non-compliant behavior to the request by the DRL agents on LoRa gateways. For example, when an animal A's sensed data is requested by a DRL agent, the attacker can either not send A's data or send another sensor's data to the LoRa gateway. We model this with the attacker's non-compliance probability, P_{NCA}.
- *False data injection*: An attacker (e.g., a compromised sensor) can transmit forged/modified data or inject false data to gateways. In addition, man-in-the-middle attackers (MIMAs) can intercept data being transmitted in the middle and replace it with forged/modified data. The attackers can inject false data during the training phase (i.e., poisonous attacks) or the testing phase (i.e., evasion attacks). We call the compromised sensors *internal attackers* while calling the external attackers intercepting sensed data for forgery/modification or injecting false data *external attackers*. These attacks are modeled by the forging/modifying data attacks an internal or external attacker can launch, P_{IDA} and P_{EDA}, respectively.
- *Denial-of-Service (DoS)*: A compromised high-battery sensor can send a request to nearby sensors requesting them to send its fake sensed data. As it is a type of internal attack, we also model this DoS attack probability by P_{IDA}. This can drain other sensors' energy levels quickly but is wasted in sending false data. To avoid an infinite loop, we assume the attacker will request sending its fake sensed data to legitimate sensors.

We also consider the following **adversarial examples** to disrupt DRL agents' operations:

- *Fast Gradient Sign Method (FGSM)*: This state manipulation attack model is firstly proposed in [18] to generate adversarial examples in image classification tasks. To apply it in the context of DRL-based algorithm execution, we generalize DRL settings by considering actions as class labels. To make a fair comparison, we use the original loss function of each DRL algorithm to compute the gradients. Since we have multiple DRL agents in our smart farm system, we assume the attack will happen in both local agent states and global agent observations. We define the perturbation strength as the probability of performing an FGSM attack, P_{FGS}.
- *Momentum Iterative Method (MIM)* [16]: MIM is a gradient-based adversarial attack to boost the success rate of the generated adversarial examples. Unlike other gradient-based methods to maximize the loss function, MIM manipulates a velocity vector for the gradient direction in the loss function at each step as an iterative attack method. We apply the MIM attack to each DRL scheme generating adversarial states using the loss function. We model the perturbation strength by the probability of performing a MIM attack, P_{MIM}.
- *Projected Gradient Descent (PGD)* [33]: This attack is a multi-step variant FGSM and uses a negative loss function to maximize the inner part of the saddle point formulation.

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3287069



(a) Outsider attacks by false data injection

(b) Insider attacks by false data injection, DoS, or non-compliance

Fig. 3: Attack scenarios by both outsider and insider attackers. Recall that P_{EDA} is the probability of an external attacker performing false data injection attacks, P_{NCA} is the probability of an internal attacker performing non-compliance attacks, and P_{IDA} is the probability of an inside attacker performing false data injection attacks or DoS attacks.

For each DRL agent, we compute the gradient and generate corresponding adversarial examples using their loss functions. This attack strength is modeled by the probability of performing a PGM attack, P_{PGD}.

- *Basic Iterative Method (BIM)* [28]: This attack is a variant of the fast method to constrain each adversarial example in ϵ-neighborhood of the original example. The adversarial examples are computed by applying the fast method multiple times with a small step size, allowing the low computational cost to perform BIM. This attack is applied with adversarial states impacting the original loss function where the perturbation strength is modeled by the probability of performing a BIM attack, P_{BIM}.

We summarize the above seven types of attacks in Fig. 3.

V. DRL-based, Uncertainty-Aware Animal Monitoring

In this section, we provide a detailed description of our proposed uncertainty-aware DRL-based algorithms for smart farm animal monitoring.

A. Uncertainty-Aware Animal Monitoring

First, based on received data from sensors in the past, a gateway can estimate uncertainty in each animal's condition, such as heart rate, average temperature, minimum/maximum temperature, average activity, battery level of a sensor worn by the animal, and timestamp. Recall that each LBS will send its sensed data to an HBS within 100 meters. Hence, we distinguish direct sensed data from indirect sensed data in terms of whether a sensor sent its own sensed data or another sensor's sensed data. If the sensor with high energy transmitted other sensor's sensed data, it is treated as indirect sensed data. Otherwise, it is direct sensed data. The gateway periodically reports its collected data from sensors to the cloud server. Since the given network has multiple gateways, the corresponding multiple DRL agents will share information about sensed data and estimate each animal's conditions (see Table I) and associated confidence level on each observation item. A database on the gateway keeps recording all animals' condition data where sensed data by sensor i (i.e., ID) for an animal are stored.

Each observation item's condition (e.g., average temperature) will be reported as one of the K classes of the range, e.g., for the temperature reading, $K = 5$, meaning that there are 5 classes of ranges: 35 or below, 36-37, 38-39, 40-41, 42 or above. The end user can then easily determine if the temperature is normal based on the cloud server's received data. Since the gateway will periodically report average conditions for all animals to the cloud, it will aggregate sensed data from sensor nodes and measure their average with the probability of a condition being with K classes and multiple types of uncertainty values. To utilize the concept of uncertainty, we will apply SL [26] to compute an opinion on each animal's condition in a given attribute (e.g., temperature, heartbeats, activity, or battery level).

1) SL-based Formulation of a Multinomial Opinion: SL can explicitly express uncertainty caused by a lack of evidence, called *vacuity* in its opinion representation. In addition, SL can consider base rates as prior probabilities in a Bayesian way to formulate a second-order opinion and corresponding uncertainty estimates, where a second-order opinion is represented by Dirichlet distribution. We will use a Dirichlet probability density function (PDF) to model the distribution of class probabilities and corresponding uncertainty masses. In SL, given a random variable X and its sample space $\mathbb{X}$, a multinomial opinion of X is represented by $\omega_X = (\boldsymbol{b}_X, u_X, \boldsymbol{a}_X)$. We call $x \in \mathbb{X}$ as a belief mass, and $|\mathbb{X}|$ is the number of belief masses. The additivity requirement of ω_X holds $\sum_{x \in \mathbb{X}} \boldsymbol{b}_X(x) + u_X = 1$. Specifically, each parameter indicates,

- $\boldsymbol{b}_X$: *belief mass distribution* over $\mathbb{X}$;
- u_X: *uncertainty mass* representing *vacuity of evidence*;
- $\boldsymbol{a}_X$: *base rate distribution* over $\mathbb{X}$.

The projected probability distribution of multinomial opinions is given by:

$$\boldsymbol{P}_X(x) = \boldsymbol{b}_X(x) + \boldsymbol{a}_X(x) \cdot u_X, \quad \forall x \in \mathbb{X} \tag{4}$$

The base rate for belief $\boldsymbol{b}_X(x_i)$, which is $\boldsymbol{a}_X(x_i)$, means the prior preference over the x_i belief (e.g., a class). If no preference is given, we consider the base rate equally for each belief mass, i.e., $\boldsymbol{a}_X(x_i) = 1/|\mathbb{X}|$ for any x_i.

Given the amount of evidence supporting belief x_i is $\boldsymbol{r}(x_i)$, the observed evidence in the Dirichlet PDF can be mapped to the multinomial opinions as:

$$\boldsymbol{b}_X(x) = \frac{\boldsymbol{r}(x)}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \; u_X = \frac{W}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \tag{5}$$

where W refers to the amount of uncertain evidence commonly set as $W = |\mathbb{X}|$ initially.

2) Estimation of Multiple Types of Uncertainty: SL categorizes uncertainty into two primary sources [26]: (1) basic belief uncertainty derived from single belief masses, and (2) intra-belief uncertainty based on the relationships between different belief masses. These two sources of uncertainty can be categorized as the two uncertainty types, *vacuity* and *dissonance*, respectively, corresponding to vacuous beliefs and contradicting beliefs. In particular, the vacuity of an opinion ω_X is captured by uncertainty mass u_X while dissonance of an opinion, $\boldsymbol{b}_X^{\text{Diss}}$, is formulated by [25]:

$$\boldsymbol{b}_X^{\text{Diss}} = \sum_{x_i \in \mathbb{X}} \left(\frac{\boldsymbol{b}_X(x_i) \sum\limits_{x_j \in \mathbb{X} \setminus x_i} \boldsymbol{b}_X(x_j) \text{Bal}(x_j, x_i)}{\sum\limits_{x_j \in \mathbb{X} \setminus x_i} \boldsymbol{b}_X(x_j)} \right), \tag{6}$$

where the relative mass balance between a pair of belief masses $\boldsymbol{b}_X(x_j)$ and $\boldsymbol{b}_X(x_i)$ is expressed by:

$$\text{Bal}(x_j, x_i) = 1 - \frac{|\boldsymbol{b}_X(x_j) - \boldsymbol{b}_X(x_i)|}{\boldsymbol{b}_X(x_j) + \boldsymbol{b}_X(x_i)}. \tag{7}$$

The dissonance estimation is useful to measure the *inconclusiveness* of an opinion even under a large amount of evidence that almost equally supports each singleton belief.

In this work, we regard each reported data from sensor nodes to a gateway as evidence. For instance, in a temperature report, if 38 C is reported, b_2 (i.e., b_1 = lower than normal, b_2 = normal, b_3 = higher than normal) should be updated based on Eq. (5). When the uncertainty mass becomes zero, an opinion will not be updated anymore, which makes new evidence cannot be properly utilized in the latest opinion. To avoid this, we will deploy the *uncertainty maximization* technique [26] to reduce the impact of conflicting evidence while transforming the amount of conflicting evidence into the vacuity of an opinion.

Given opinion $\omega_X = (\boldsymbol{b}_X, u_X, \boldsymbol{a}_X)$ where $\boldsymbol{P}_X = \boldsymbol{b}_X + \boldsymbol{a}_X \cdot u_X$ for a domain X, the corresponding vacuity-maximized opinion is denoted by $\ddot{\omega}_X = (\ddot{\boldsymbol{b}}_X, \ddot{u}_X, \boldsymbol{a}_X)$ where $\ddot{u}_X$ and $\ddot{\boldsymbol{b}}_X$ are computed by:

$$\ddot{u}_X = \min_i \left[\frac{\boldsymbol{P}_X(x_i)}{\boldsymbol{a}_X(x_i)} \right], \tag{8}$$

$$\ddot{\boldsymbol{b}}_X(x_i) = \boldsymbol{P}_X(x_i) - \boldsymbol{a}_X(x_i) \cdot \ddot{u}, \;\; \text{for} \;\; x_i \in X.$$

We use a threshold ρ to trigger the vacuity maximization. That is, Eq. (8) above can be triggered only when $u_X < \rho$ where ρ is sufficiently low (e.g., 0.05). The purpose of updating ω_X to $\ddot{\omega}_X$ is to allow the opinion to be further updated by receiving new evidence or being combined with other opinions, which is possible only when $u_X > 0$.

In a given category X, the animal condition is estimated as an opinion, ω_X, where the corresponding uncertainty types, vacuity, and dissonance are estimated, respectively, at the gateways that aggregate sensed data and transmit the average condition value in a given category along with the belief masses and uncertainty masses associated with ω_X.

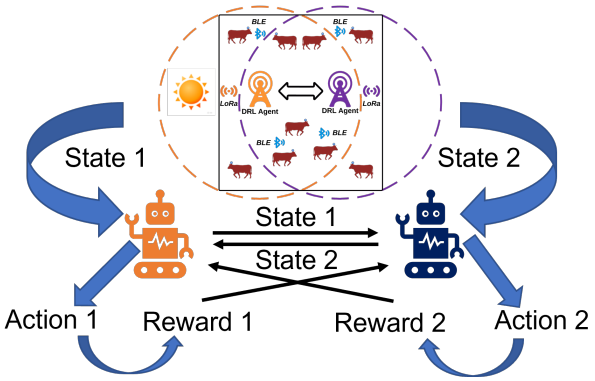


Fig. 4: The proposed Multi-Agent Deep Reinforcement Learning (MADRL) framework.

B. Data Aggregation at LoRa Gateways

Each sensor node will send its sensed data to LoRa gateways or a high-energy sensor close to it, as shown in Table I. After receiving the reports from all sensor nodes capable of transmitting their data, a LoRa gateway will compute an opinion based on the received data. The opinion is composed of belief and uncertainty in vacuity and dissonance masses. We define the opinion as a *monitoring opinion* (MO) and denote it as Δ below. Specifically, given time step t, for the agent i and animal j, we define the attribute sets $\mathbb{S}_{ij}^t = \{\text{temp}_{ij}^t, \text{hb}_{ij}^t, \text{ma}_{ij}^t\}$, where $\text{temp}_{ij}^t, \text{hb}_{ij}^t, \text{ma}_{ij}^t$ are temperature, the number of heartbeats, speed of animal j recorded by agent i at time t respectively. We let the random variable X_{ij}^t be the animal condition status of animal j recorded by agent i at time t and define the corresponding sample space $\mathbb{X}_{ij}^t = \{\text{below normal range}, \text{normal range}, \text{above normal range}\}$. Then for each sensor report x_{ij}^t about an attribute $s \in \mathbb{S}_{ij}$, $x_{ij}^t \in \mathbb{X}_{ij}$ and $\Delta(i,j,t) = \Delta(\omega_{X_{ij}^t}) = (u_{X_{ij}^t}, \boldsymbol{b}_{X_{ij}^t}^{\text{Diss}})$. When the amount of sensed data becomes large enough, the MO may not be effectively updated from new sensed data because vacuity approaches closely to zero based on Eq. (5). To update the MO properly from received new evidence, we will deploy the vacuity (uncertainty) maximization technique in Eq. (8). We use a threshold ρ (i.e., $0 < \rho < 1$) to determine when to update the MO based on Eq. (8). If $u_X < \rho$, this evidence will update an opinion based on Eq. (8). To evaluate the system monitoring error based on Eq. (2), we also introduce a system database to collect the latest monitoring opinions for each animal among all gateways.

C. DRL-based Monitoring Update

This section describes how the DRL agent's states, actions, and immediate reward are formulated in this work.

1) State Space ($\mathcal{S}_t$): We assume a partially observable environment where each DRL agent can only access information in the dataset of the local gateway it is running on. We formulate the state space of each agent i at time t indicating the total number of local reports for the duration of t, $\{\ell_{i,1}, \ell_{i,2}, \ldots, \ell_{i,t-1}, \ell_{i,t}\}$, where ℓ_{i,t^*} refers to the number of local reports in $[t^* - 1, t^*]$. To be specific, assume $t \in [0, T]$, the overall state space is given by $\mathcal{S}_t = \{s_{1,t}, \ldots, s_{2,t}, \ldots, s_{m,t}\}$, where m is the number of DRL agents (i.e., LoRa gateways) and $s_{i,t}$ is the state space for agent i at time t, which is given by $s_{i,t} = \{\ell_{i,1}, \ell_{i,2}, \ldots, \ell_{i,t}, 0, \ldots, 0\}$.

2) Action Space ($\mathcal{A}_t$): For each DRL agent, it will choose k animals whose data is more helpful in improving monitoring quality and reducing system overload. Note that a certain amount of redundant information is desired since there is a possible situation that sensors fail to transmit data due to limitations of their energy level or topology. For the agent i and animal j, the utility of animal j is given by:

$$\text{utility}_{ij} = (1 - u_{X_{ij}^t}) + (1 - \boldsymbol{b}^{\text{Diss}}_{X_{ij}^t}) + \text{fr}_{ij}^t + f(\text{bl}_{ij}^t), \quad (9)$$

where $u_{X_{ij}^t}$, $\boldsymbol{b}^{\text{Diss}}_{X_{ij}^t}$, and fr_{ij}^t are vacuity, dissonance, and degree of freshness of animal j's sensed data at time t by DRL agent i. fr_{ij}^t is formulated by $\text{fr}_{ij}^t = e^{-\phi \mathbb{T}(t)}$, where $\mathbb{T}(t)$ is the time elapsed from the last update and ϕ is a constant to normalize the freshness. $f(x)$ is defined by $f(x) = -(x - T_M)^2$ where x is set to bl_{ij}^t, the battery life of sensor j at time t by DRL agent i. By scaling $u_{X_{ij}^t}, \boldsymbol{b}^{\text{Diss}}_{X_{ij}^t}, \text{bl}_{ij}^t$, and fr_{ij}^t in $[0, 1]$, we set each component of utility_{ij} to $[0, 1]$ as a real number. Here T_M denotes the recommended level that the battery of a sensor node should be maintained. A list of animal IDs will be calculated based on Eq. (9) in ascending order so that each agent will request data for the top k animals. The action space has three actions selecting the first k animal IDs such that $k \in \{0, \lfloor \frac{nl}{2} \rfloor, nl\}$, where nl is how many LBS nodes are in the current environment. In this way, the size of action space is not dependent on nl (i.e., 3), which is able to reduce the computation load raised by infinite action spaces and make this monitoring system possible for applying to larger-scale sensor networks as a generalization. For a lower k, it may impact the monitoring quality. At the same time, a higher k will obtain a larger amount of unnecessary data transmission and result in a system overload. Therefore, the DRL agent aims to identify the best action, which is the optimal k for this setting.

3) Immediate Reward (r_t): This is formulated by $r_t^i = f(g_t^i(k_1^i, k_2^i, \ldots, k_t^i), g_t)$ based on $f(g_t^i, g_t) = -\mathcal{ME}(g_t^i) - \mathcal{OL}(g_t)$ given in Eq. (1), where g_t^i and k_t^i are the local sensor network and action with respect to gateway i at time step t. Note that we use the accurate reward r_t in the DRL performance evaluations and use the noisy reward r_t^* in the DRL training due to observational errors. Here, we assume that r_t^* follows the normal distribution $\mathcal{N}(\mu, \sigma^2)$ where $\mu = r_t$ and $\sigma = 0.3 * r_t$.

In this work, we consider a cooperative framework where multiple DRL agents share their states and rewards to obtain the global observation of the whole farm area. Fig. 4 provides an overview of the proposed Multi-Agent Deep Reinforcement Learning (MADRL) framework.

D. Mathematical Proof of Effectiveness Using Uncertainty Maximization

In this section, we formally prove the effectiveness of the *uncertainty maximization* technique [26] by mathematical proof. We observe that uncertainty mass and the monitoring error rate can be viewed as functions of evidence. Based on Eq. (5), the uncertainty (*vacuity*) drops when the amount of received evidence increases. Furthermore, given Eq. (6), *dissonance* solely depends on the distribution of belief masses without vacuity being involved. Thus, the dissonance can be viewed as a constant when there is enough evidence from the same distribution.

The uncertainty (vacuity) maximization in Eq. (8) reinitializes the vacuity by transforming previous evidence from the belief masses to the uncertainty mass. Given the following [26],

$$j = \arg\min_i \Big[\frac{\boldsymbol{P}_X(x_i)}{\boldsymbol{a}_X(x_i)}\Big], \quad (10)$$

where $\boldsymbol{P}_X(x_i)$ is the projected probability of having x_i and $\boldsymbol{a}_X(x_i)$ is the base rate (i.e., prior belief) that supports a belief mass x_i. Then, we have the updated belief masses and the uncertainty (vacuity) mass given by:

$$\ddot{\boldsymbol{b}}_X(x_k) = \frac{\boldsymbol{r}(x_k) - \boldsymbol{r}(x_j)}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \ddot{u}_X = \frac{W + K\boldsymbol{r}(x_j)}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \quad (11)$$

where W is the amount of non-informed evidence (i.e., uncertain evidence), K is the number of belief masses (e.g., classes), and $\boldsymbol{r}(x)$ is the amount of evidence supporting belief mass x. Since the amount of non-informed evidence, W, increases to $W + K\boldsymbol{r}(x_j)$, we replace W with $W + K\boldsymbol{r}(x_j)$ in the denominators of both $\ddot{\boldsymbol{b}}_X(x_k)$ and $\ddot{u}_X$ as

$$\ddot{\boldsymbol{b}}_X(x_k) = \frac{\boldsymbol{r}(x_k) - \boldsymbol{r}(x_j)}{(W + K\boldsymbol{r}(x_j)) + \sum_{x_i \in \mathbb{X}} (\boldsymbol{r}(x_i) - \boldsymbol{r}(x_j))}, \quad (12)$$

$$\ddot{u}_X = \frac{W + K\boldsymbol{r}(x_j)}{(W + K\boldsymbol{r}(x_j)) + \sum_{x_i \in \mathbb{X}} (\boldsymbol{r}(x_i) - \boldsymbol{r}(x_j))}. \quad (13)$$

The above implies that the updated vacuity only considers partial history evidence, indicating more recent evidence than past evidence.

As shown in Eq. (2), the monitoring error rate is closely related to the amount of evidence. Assume that at each time step $t \in [0, T]$, the information of n_t animal is being updated and each animal j's information is updated m_j times in total.

TABLE I: EVD DATASET DESCRIPTION

Metric	Description
serial	A unique animal identifier
HR	Heart Rate of the animal
average-temperature	Average body temperature in Celsius
min-temperature	Minimum temperature in Celsius
max-temperature	Maximum temperature in Celsius
average-activity	Average activity recorded by the number of steps taken
battery-level	Residual battery life
timestamp	Date and time of transmission

Hence, we have the expected monitoring error rate given by,

$$\begin{aligned} E(\mathcal{ME}) &= \frac{E(\sum_{t\in T}\sum_{x\in X}\mathrm{me}_t^x)}{NT|X|} \\ &= \frac{\sum_{t\in T}\sum_{x\in X}E(\sum_{j=1}^{n}D(\mathrm{eo}_t^x(j),\mathrm{gt}_t^x(j)))}{NT|X|} \\ &= \frac{\sum_{x\in X,t\in T}E(\sum_{j=1}^{n_t}0+\sum_{j=n_t+1}^{n}1)}{NT|X|} \\ &= \frac{\sum_{x\in X}\sum_{t\in T}(N-n_t)}{NT|X|} \\ &= 1-\frac{\sum_{x\in X}\sum_{j=1}^{n}m_j}{NT|X|}. \end{aligned} \tag{14}$$

Here T is the total monitoring time, N is the number of solar sensors attached to cows, me_t^x is the overall monitoring error rate of all N cows' conditions of attribute x at time t, $\mathrm{eo}_t^x(j)$ and $\mathrm{gt}_t^x(j)$ is the estimated and ground truth observation of cow j's condition in x attribute at time t, respectively. The above derivation proves that $E(\mathcal{ME})$ is only dependent upon n_t and it increases when n_t decreases. In addition, each animal j's monitoring error rate is only related to the amount of corresponding evidence, m_j. Therefore, we prove that the order of any pair of sensors remains invariant under the partial order relations of vacuity and monitoring error rate.

VI. EXPERIMENTAL SETUP

This section describes the datasets, parameterization, DRL schemes, and performance metrics used for our experiments.

A. Datasets

At Virginia Tech, we have a collection of interconnected data collection and analysis hubs called the *Smart Farm Innovation Network* (TM), which is designed to facilitate the testing and demonstration of emerging technologies throughout the state. From the smart farm, we obtained sample datasets collected from four different sensors, namely, EmbediVet Implantable Temperature Device (EVD), Halter Sensor, Heart Rate Sensor, and Implantable Temperature Sensor. The dataset from the EVD consists of 8 components as described in Table I. We consider 6 components out of them, except a serial number and timestamp, as the sensed data to represent the physical conditions of animals. The temperature and the heart rate sensor provide us with temperature in Celsius and heart rate in beats per minute (bpm), respectively. The Halter sensor could identify each animal's geolocation and assess its motion and posture to report its activity level. Since the existing dataset obtained from Virginia Tech's smart farm does not include any data compromised by attackers, we designed a framework where each sensor could generate synthetic datasets similar to real datasets, and some compromised sensors are modeled as attackers following the attack model described in Section IV-C.

B. Parameterization

We consider 20 cows within 40 acres ($\sim$ 160K square meters) square farm area (A) with 402 meters in length (a). To determine the number of implemented gateways, we consider the communication distance, costs, and latency. Note that a lack of gateways could introduce communication overhead and latency due to the overwhelming communication requests and the long communication distance. An excessive amount of gateways would introduce the implementation costs for devices and computation costs for DRL agents. Thus, in this work, we consider two gateways with the same circular coverage. We further assume that these gateways could cover the farm area and each of them is covered by the other. In general, for a given number of m gateways with the same radius r, we aim to find the minimum radius r_m such that the farm area is fully covered by the total gateway coverage and each gateway is covered by other gateways to enable mutual communications. To this end, we solve the following optimization problem to identify the minimum radius r_m:

$$\begin{aligned} r_m = \min\{r : \forall i \in [1,m], \exists P_i = (x_i,y_i) \in \mathbb{R}^2, \\ s.t.\ \forall P=(x,y)\in[-a/2,a/2]^2, \min_{1\le i\le m} d(P,P_i) \le r \\ \wedge \forall (j,k)\in[1,m]^2,\ \ \mathrm{d}(P_j,P_k)\le r\}, \end{aligned} \tag{15}$$

where a is the length of farm side, P_i is the center of gateway, and $\mathrm{d}(\cdot,\cdot)$ is the Euclidean distance function. We only consider the case when $m=2$, where gateways locate in $(-\frac{a}{4},0)$ and $(\frac{a}{4},0)$ respectively with the same radius $\frac{\sqrt{5}a}{4}$. Fig. 1 shows how two gateways are optimally deployed with the corresponding wireless radio ranges used in our smart farm network environment. To further scale the current design to more gateways in a larger farm, we can identify the optimal number of gateways m^* in the following optimization problem:

$$m^* = \min\{m : r_m \le r^*\} \tag{16}$$

where r^* is a given upper bound of gateway coverage, r_m is defined in Eq. (15).

To model the availability of solar energy based on the sun's movement in a day, we define a charging probability distribution $P(x,y,t)$ over the farm area as the probability of being charged if a sensor is located at (x,y) at time t. For simplicity, we assume that $P(x,y,t)$ has a quadratic form at time t and is represented by $P(x,y,t) = \max\{0, -\frac{1}{6}(t-t_{xy})^2+1\}$, where t_{xy} is a function of location (x,y) based on the farm's direction. We consider a square farm with its center at the origin and x axis towards the west. Thus, t_{xy} is formulated as $t_{xy} = \frac{t_0}{a}\times(x-\frac{a}{2})+12$, where t_0 is a hyper-parameter. In general, to model different weather conditions, we can use a

weight α to discount the charging probability as $\alpha P(x, y, t)$ with $0 \leq \alpha \leq 1$.

Each cow's attributes are collected by an attached solar-powered sensor. We adopt normal distributions $\mathcal{N}(38, 1^2)$ and $\mathcal{N}(1.5, 0.1^2)$ to describe a cow's temperature [14] and velocity [30] respectively. The cow's heartbeat is modeled as two uniform distributions [1]: $\mathcal{U}(60, 84)$ when it moves or $\mathcal{U}(48, 60)$ when it does not move [14]. We use P_i^{mv} for cow i's moving probability.

For an opinion about a cow's attributes, we will simply categorize based on three beliefs, i.e., lower than normal, normal, and higher than normal. The normal ranges of a cow's temperature, heart rate, and moving activity are given $[37.8, 39.2]$ Celsius, $[48, 84]$ number of beats per min., and $[1, 2]$ meters per sec., respectively. We consider the number of uncertain evidence to be three where each belief mass has the same base rate (i.e., 1/3) [26].

We consider 24 consecutive hours as the total monitoring session. For every $T_a = 60$ sec, each gateway would take an action to identify an optimal monitoring strategy. We assume 5 HBS with a full initial battery level and 15 LBS with random initial battery levels based on T_L. When the battery level is below T_M, LBS could only broadcast their own data to HBS via BLE. Each sensor can broadcast at most two sets of sensed data to each LoRa gateway within the wireless range per $T_u = 30$ sec. This way allows each HBS to send its own data and another set of data requested by the LBS. The monitoring system would derive the consolidated priority list of update lists from gateways. Then the system would leverage the Hopcroft–Karp algorithm [21] to solve the maximum matching problem in bipartite sensor networks. This way enables the system to ensure the maximum number of transmissions being executed.

As for energy consumption, message transmissions need 170 mW per sec and the sleep mode costs 2 mW per sec. We assume a sensor is only activated for message transmissions. A sensor can be charged by the outdoor solar with 200 mW per sec. In this way, a sensor can be charged 200 $mW \times 6$ h = 4.32 kWs under 6 hours of the sun.

We set all attack probabilities to P_A. For inside attackers, we initially pick them among the total number of sensors at random. For outside attackers (i.e., MIMAs), we pick a set of nodes at random transmitting messages intercepted by MIMAs with $P_{EDA}(P_A)$. The attackers launch attacks based on Fig. 3. Finally, we consider FGSM as a state manipulation attack to disturb the DRL training phase. Table II summarizes the key design parameters, their meanings, and default values.

TABLE II: KEY DESIGN PARAMETERS, THEIR MEANINGS, AND DEFAULT VALUES

Param.	Meaning	Value
m	The number of gateways	2
N	The number of sensors(cows)	20
T_M	A minimum battery level to transmit sensed data by a sensor	30%
LBS/HBS	Low/High battery level sensors	/
P_i^{mv}	Cow i's probability to move	$[0.3, 0.7]$
P_A	Probability for an attacker or a compromised node to perform a cyber attack (i.e., P_{NCA}, P_{IDA}, P_{EDA})	0.3
P_{ADV}	Probability for an attacker to perform an adversarial attack (i.e., P_{FGS}, P_{MIM}, P_{PGD}, P_{BIM})	0
A	Area of a given smart farm	40 acres
a	length of a given smart farm	402 m
ρ	Uncertainty maximization threshold	0.05
ϕ	A constant to normalize the freshness	0.01
t_0	Hyper-parameter used in sun model	0.2
T_u	Time interval for a sensor to send sensed data	30 s
T_a	Time interval for a gateway to take an action to adjust k	60 s
T_L	Initial battery level for low battery level sensors	30%
α	The probability for a cow to be exposed to the sun depending on its position when the sun is available	1

C. Comparing Schemes

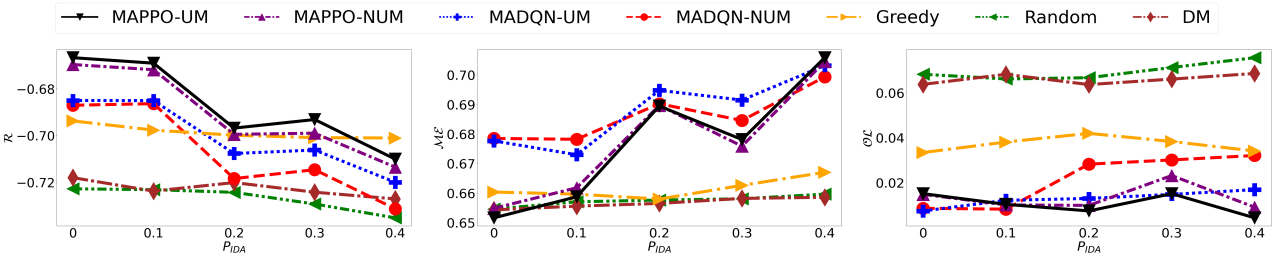
We develop two uncertainty-aware DRL algorithms (MADQN and MAPPO) whose performance is compared against three baseline schemes (Greedy, Random, and DM), described as follows:

- **Multi-Agent Deep Q-Learning (MADQN)** [35]: DRL agents learn a state-value Q function to select the optimal actions. In a multi-agent scenario, we extend DQN to MADQN, where each DRL agent learns an independent local Q function. We consider two variants of MADQN using UM or not and name them MADQN-UM and MADQN-NUM, respectively.
- **Multi-Agent Proximal Policy Optimization (MAPPO)**: MAPPO extends the PPO [46] to a multi-agent environment to mitigate non-stationarity by adopting a global critic value function to guide each local actor value function. We consider two variants of MAPPO with and without using uncertainty maximization. We name them MAPPO-UM and MAPPO-NUM, respectively.
- **Greedy Algorithm**: DRL agents make heuristic choices by enumerating all actions at each step and choosing the one with the optimal reward. Note that this approach is not a feasible solution in practice under resource-constrained environments, such as WSNs considered in this work. This approach is used to indicate how much our proposed approach can perform comparably to this algorithm.
- **Random**: DRL agents randomly select an action, e.g., k animal IDs to receive their sensed data.
- **Data Mitigation (DM) [24]**: DRL agents randomly select an action and reduce the redundant requests by restricting the communications between sensors and gateways. Each LBS is allowed to communicate to one gateway at one time.

D. Metrics

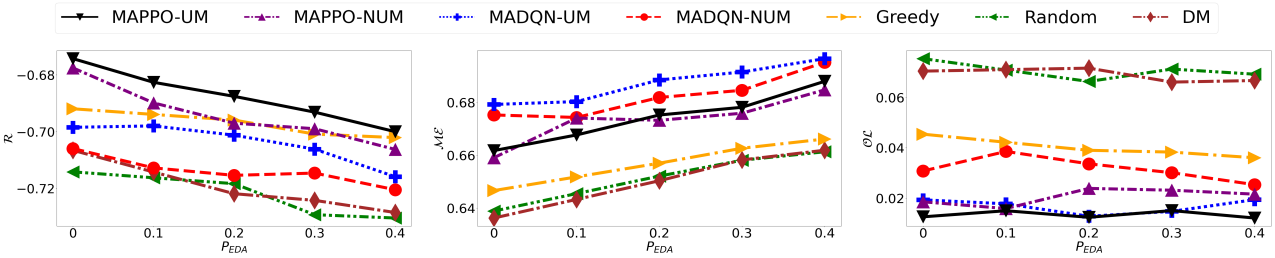
We consider the following metrics to evaluate the performance of the four DRL schemes described in Section VI-C:

- **Accumulated reward** ($\mathcal{R}$): This metric represents the mean of the sum of immediate rewards for all DRL agents through all simulation runs. The system goal is to minimize the monitoring error and system overload simultaneously as shown in the objective function Eq. (1) in our problem statement. In Eq. (1), we defined the evaluation function



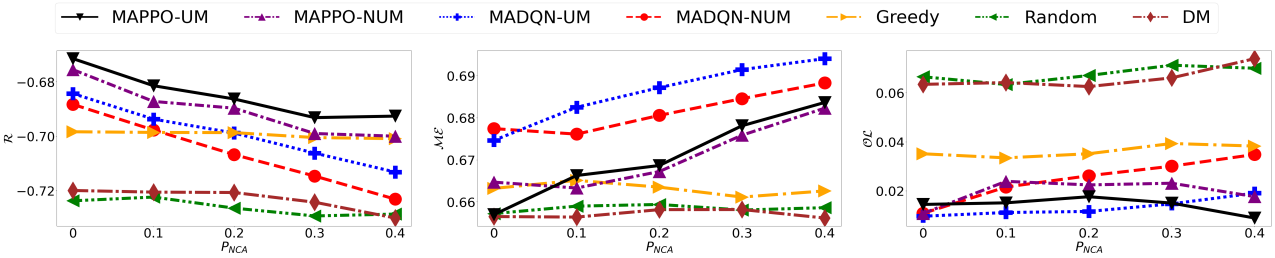
(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 5: Comparative performance with respect to varying the internal attack probability (P_{IDA}).



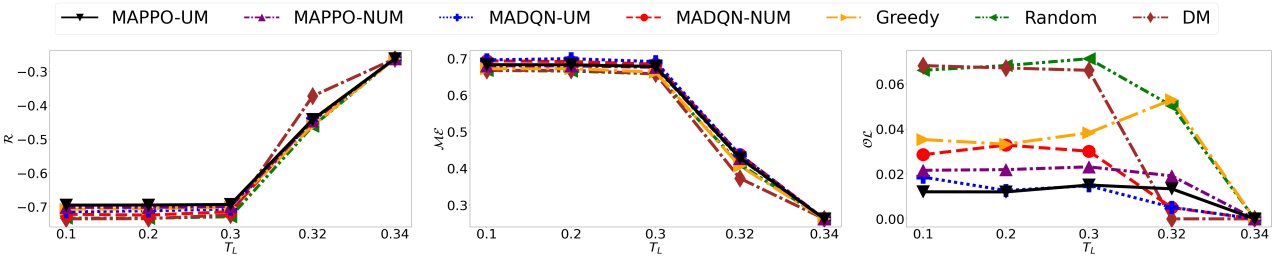
(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 6: Comparative performance with respect to varying the external attack probability (P_{EDA}).



(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 7: Comparative performance with respect to varying the attacker's non-compliance probability (P_{NCA}).



(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 8: Comparative performance with respect to varying the initial low battery level (T_L).

$f : g \mapsto -\mathcal{ME}(g) - \mathcal{OL}(g)$. The output of f is also the immediate reward of each DRL agent defined in Section V-C3. Thus, the accumulated reward defined in Section VI.D can be used to evaluate the overall performance of our comparing schemes.

- **Monitoring error rate** ($\mathcal{ME}$): This metric is measured based on the mean difference between the latest data of each animal's condition from all gateways and the ground truth data of the corresponding animal's condition. We measure $\mathcal{ME}$ by Eq. 2 introduced in Section III. There are other error rates metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). However, we measure the number of errors because small value differences may not indicate low error times due to the heterogeneous data distributions.
- **Overload** ($\mathcal{OL}$): This metric evaluates the system overload by the mean fraction of the failed requests over all sent requests from LBS. We measure $\mathcal{OL}$ following Eq. 3 in Section III.

VII. Experimental Results and Analysis

A. Algorithmic Complexity Analysis

We first analyze the algorithmic complexity of the five DRL schemes described in Section VI-C. Table III shows

TABLE III: ASYMPTOTIC COMPLEXITY ANALYSIS OF THE CONSIDERED SCHEMES

Scheme	Complexity
MADQN/MAPPO	$O(n_e \times t_{train})$
Greedy	$O(n_{action})$
Random	$O(1)$
DM	$O(1)$

TABLE IV: EFFECT OF ADVERSARIAL EXAMPLE ATTACKS ON THE REDUCTION OF THE ACCUMULATED REWARD IN %

P_{ADV} \ Attack	FGS	MIM	PGD	BIM
0.1	0.2	0.2	0.8	0.3
0.2	0.7	0.9	1.4	0.9
0.3	1.2	0.8	1.6	0.8
0.4	1.3	1.3	2.1	0.9

the asymptotic complexities in Big-O notation for the five schemes discussed in Section VI-C. We notice that the cost of MADQN/MAPPO only depends on the training episode n_e and training time per episode t_{train}. Greedy needs to enumerate the total action space and thus its complexity depends on the action space size n_{action}. When the action space is large enough, greedy can incur more cost than MADQN/MAPPO. Table III shows that the Random and DM are the most efficient algorithms among all while showing the worst performance (to be discussed further in the next section).

B. Sensitivity Analyses

Below, we conduct in-depth sensitivity analyses of the two baseline schemes (Greedy and Random) and the two uncertainty-aware DRL schemes (i.e., MADQN, MAPPO) with uncertainty maximization (i.e., MADQN-UM, MAPPO-UM) vs. without uncertainty maximization (i.e., MADQN-NUM, MAPPO-NUM) over a wide range of the cyber attack probability, P_A (i.e., P_{NCA}, P_{IDA}, P_{EDA}), the initial low battery level T_L, the number of cows (sensors) N, and the chance for a cow to be exposed to the sun, α. We also conduct the sensitivity analyses of MAPPO-UM with respect to four different adversarial attacks.

1) Effect of Varying the Adversarial Attack Severity: Table IV shows the effect of varying the adversarial attack probability, P_{ADV}, on the performance of MAPPO-UM in terms of the reduction in the accumulated reward. For a clear comparison, we show the relative performance decline in % with respect to the default setting where $P_{ADV} = 0$. We observe that MAPPO-UM shows resilience towards all four attacks as they only decrease performance slightly. Among all considered attacks, PGD is the strongest, while BIM is the weakest.

2) Effect of Varying the Cyber Attack Severity: Fig. 5–7 show the effect of varying the cyber attack probability, P_A, on the performance of the seven schemes in terms of the three metrics in the network. We observe that increasing P_A decreases $\mathcal{R}$ while increasing $\mathcal{ME}$ and $\mathcal{OL}$. When P_A increases, the monitoring system becomes severely compromised and accordingly the payoff per monitoring update would drop. MAPPO can successfully identify this change in the payoff and achieve better performance than other schemes. The overall performance order with respect to the three metrics is MAPPO-UM $\geq$ MAPPO-NUM $\geq$ Greedy $\geq$ MADQN-UM $\geq$ MADQN-NUM $\geq$ DM $\geq$ Random.

3) Effect of Varying the Initial Battery Levels (T_L): Fig. 8 shows the effect of varying the initial battery level assigned to low-battery level sensors (LBS), T_L, on the performance of the seven schemes in terms of the three metrics in the network. We observe that increasing attack probability (T_L) increases the accumulated reward ($\mathcal{R}$) while decreasing the monitoring error rate ($\mathcal{ME}$) and the degree of overload ($\mathcal{OL}$). We also observe that when T_L increases, all three metrics converge to one point due to the decreased number of LBS. Monitoring policies only apply to LBS and thus, negligible differences are observed under high T_L. Note that DM performs the best when $T_L = 0.32$ since it can only effectively save LBS nodes' energy with a relatively high number of LBS nodes and high initial battery level of LBS nodes. This results in more successful reports of LBS nodes and, thus less monitoring error. Overall, our proposed MAPPO-based schemes can achieve a low monitoring error rate with the lowest overload.

4) Effect of Varying the Node Density (N): Fig. 9 shows the effect of varying the number of solar sensors, N, on the performance of the seven schemes in the three metrics in the network. We observe that increasing the number of sensor nodes (N) decreases the accumulated reward ($\mathcal{R}$) while introducing higher $\mathcal{ME}$ and $\mathcal{OL}$. When N increases, there are not enough HBS (high battery level sensors) to transmit data for LBS. Consequently, the update requests from LBS may mostly fail. Our proposed MAPPO-UM scheme achieves the lowest monitoring error rate when N is low and the lowest overload when N is high, revealing the tradeoff that a lower monitoring error rate can incur a higher system overload.

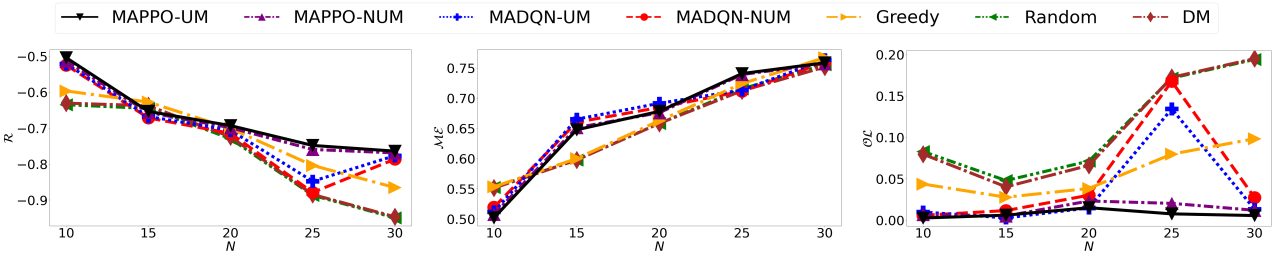
5) Effect of Varying the Degree of Sun Exposure (α): Fig. 10 shows the effect of α (the probability for a cow to be exposed to the sun for energy harvesting) on the performance of the seven schemes. We observe that a higher α contributes to boosting $\mathcal{R}$ while reducing $\mathcal{ME}$. Moreover, $\mathcal{OL}$ is not sensitive to varying α because the energy harvesting speed exceeds the battery consumption speed. We also observe that $\mathcal{ME}$ is equally reduced for every monitoring policy. Thus, when α is high, small changes in monitoring policies lead to insensitive $\mathcal{OL}$ while decreasing $\mathcal{ME}$.

Summarizing above, MAPPO-UM performs the best among all schemes, the effect of which is especially pronounced when the system is under high-stress situations, such as high attack severity, low initial battery energy, low node density, and low sun exposure. The reason is that MAPPO-UM applies not only an actor-critic framework (from *Deep Reinforcement Learning* [15]) but also uncertainty maximization (from *Subjective Logic* [25]) to derive stable monitoring policy updates based on new evidence, thus maximizing monitoring quality in terms of $\mathcal{R}$ and $\mathcal{ME}$ while reducing energy consumption in terms of $\mathcal{OL}$. This allows LoRa gateways to obtain more accurate sensed data from solar sensors having limited and fluctuating energy and to maximize uncertainty-aware monitoring quality.

VIII. CONCLUSIONS & FUTURE RESEARCH

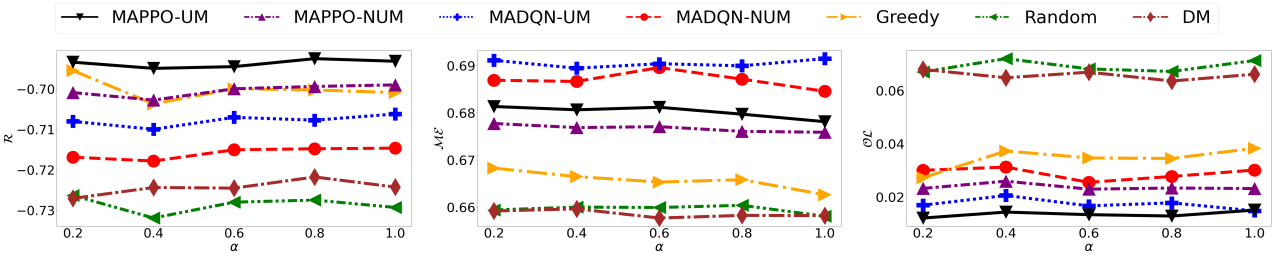
The proposed monitoring system for smart farms is the first that proposed technologies to support an energy-adaptive

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3287069



(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 9: Comparative performance with respect to varying the number of solar sensors (N) attached to cows.



(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 10: Comparative performance with respect to the different levels of sun exposure (α).

monitoring system properly operating even in the presence of various adversarial attacks, including false data injection, DoS, and state manipulation (i.e., poisoning datasets in deep learning models) attacks. Unlike existing works mainly focused on energy-aware approaches, our work achieved energy-adaptiveness and data security under energy-fluctuating, adversarial, and dynamic IoT environments. In addition, we introduced uncertainty-aware data aggregation and updated approaches to enhance the monitoring quality of the proposed smart farm system without the system being overloaded. We validated this approach via mathematical proof and extensive experiments using real datasets. We also considered multiple deep reinforcement learning agents to identify optimal settings to maximize the monitoring quality of smart farms with solar-powered sensors. This design allowed high sustainability and scalability. Moreover, collaborative learning results in high performance in monitoring quality and system overload.

From this study, the proposed MAPPO-UM showed the following performance:

- The system overload does not always increase the monitoring error rate. MAPPO-UM can identify monitoring policies minimizing monitoring errors and system overload.
- The payoffs to monitoring updates are vastly different under different scenarios. This implies that different optimal monitoring policies can be applied under different scenarios.
- MAPPO-UM generates an acceptable time complexity where the major complexity comes from the training time and the size of the action space. MAPPO-UM outperformed other counterparts by lowering about 4% in monitoring error rate and the system overload.
- Among all schemes compared, MAPPO-UM can best adapt to different scenarios and identify the best monitoring policies minimizing monitoring error rate and system overload.
- MAPPO-UM showed strong robustness, particularly under harsh environments, based on our extensive performance analyses.

We also plan to conduct the following **future research**:

- We will use more than two gateways to introduce more DRL agents for the proposed smart farm system to be applicable to larger-scale networks.
- We will leverage *transfer learning* algorithms to further improve the speed of learning convergence.
- We will identify an optimal energy level that can be used for low-energy solar sensors to request data transmission to nearby high-energy sensors.

ACKNOWLEDGEMENT

This work is partly funded by NSF Grant 2106987 and 2107450, the Commonwealth Cyber Initiative (CCI), and Virginia Tech's ICTAS EFO Opportunity Seed Investment Grant.

REFERENCES

[1] "Merck veterinary manual," Jul. 2016. [Online]. Available: https://www.merckvetmanual.com/multimedia/table/resting-heart-rates

[2] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the limits of lorawan," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, 2017.

[3] H. Alemdar, T. L. M. van Kasteren, and C. Ersoy, "Active learning with uncertainty sampling for large scale activity recognition in smart homes," *Journal of Ambient Intelligence and Smart Environments*, vol. 9, no. 2, pp. 209–223, Jan. 2017. [Online]. Available: https://content.iospress.com/articles/journal-of-ambient-intelligence-and-smart-environments/ais427

[4] A. Almeida and D. L. de Ipiña, "An approach to more reliable context-aware systems by assessing ambiguity - taking into account indeterminations and vagueness in smart environments," in *Proceedings of the 2nd International Conference on Pervasive Embedded Computing and Communication Systems - Volume 1: PECCS,*, INSTICC. SciTePress, 2012, pp. 233–236.

[5] R. S. Alonso, I. Sittón-Candanedo, R. Casado-Vara, J. Prieto, and J. M. Corchado, "Deep reinforcement learning for the management of software-defined networks in smart farming," in *2020 International Conference on Omni-layer Intelligent Systems (COINS)*. IEEE, 2020, pp. 1–6.

[6] A. D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, and S. K. Goudos, "Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: A comprehensive review," *Internet of Things*, vol. 18, p. 100187, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660520300238

[7] G. Chen, Y. Zhan, Y. Chen, L. Xiao, Y. Wang, and N. An, "Reinforcement learning based power control for in-body sensors in WBANs against jamming," *IEEE Access*, vol. 6, pp. 37 403–37 412, 2018.

[8] G. Chen, Y. Zhan, G. Sheng, L. Xiao, and Y. Wang, "Reinforcement learning-based sensor access control for wbans," *IEEE Access*, vol. 7, pp. 8483–8494, 2019.

[9] J. Cho, et al., "A survey on modeling and optimizing multi-objective systems," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1867–1901, 2017.

[10] M. Colezea, G. Musat, F. Pop, C. Negru, A. Dumitrascu, and M. Mocanu, "CLUeFARM: Integrated web-service platform for smart farms," *Computers and Electronics in Agriculture*, vol. 154, pp. 134–154, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169917305112

[11] S. K. Das and N. Roy, "Learning, prediction and mediation of context uncertainty in smart pervasive environments," in *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, R. Meersman, Z. Tari, and P. Herrero, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 820–829.

[12] A. De Paola, P. Ferraro, S. Gaglio, and G. Lo Re, "Context-awareness for multi-sensor data fusion in smart environments," in *AI*IA 2016 Advances in Artificial Intelligence*, G. Adorni, S. Cagnoni, M. Gori, and M. Maratea, Eds. Cham: Springer International Publishing, 2016, pp. 377–391.

[13] A. De Paola, P. Ferraro, S. Gaglio, G. L. Re, and S. K. Das, "An adaptive bayesian system for context-aware data fusion in smart environments," *IEEE Transactions on Mobile Computing*, vol. 16, no. 6, pp. 1502–1515, 2017.

[14] M. Denwood, "Clinican examination of the cow," Oct. 2012. [Online]. Available: https://www.gla.ac.uk/t4/~vet/files/teaching/clinicalexam/examination/info/temperatures.html

[15] H. Dong, Z. Ding, and S. Zhang, Eds., *Deep Reinforcement Learning Fundamentals, Research and Applications*. Springer, 2020.

[16] Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li, "Boosting adversarial attacks with momentum," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9185–9193.

[17] FAO. (2020) The food and agriculture organization (fao) of the united nations. [Online]. Available: http://www.fao.org/home/en/

[18] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *International Conference on Learning Representations*, 2015. [Online]. Available: http://arxiv.org/abs/1412.6572

[19] M. Gupta, M. Abdelsalam, S. Khorsandroo, and S. Mittal, "Security and privacy in smart farming: Challenges and opportunities," *IEEE Access*, vol. 8, pp. 34 564–34 584, 2020.

[20] Y. Hajjaji, W. Boulila, I. R. Farah, I. Romdhani, and A. Hussain, "Big data and iot-based applications in smart environments: A systematic review," *Computer Science Review*, vol. 39, p. 100318, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013720304184

[21] J. E. Hopcroft and R. M. Karp, "An n^5/2 algorithm for maximum matchings in bipartite graphs," *SIAM Journal on computing*, vol. 2, no. 4, pp. 225–231, 1973.

[22] S. Igder, S. Bhattacharya, and J. M. H. Elmirghani, "Energy efficient fog servers for internet of things information piece delivery (iotipd) in a smart city vehicular environment," in *2016 10th International Conference on Next Generation Mobile Applications, Security and Technologies (NGMAST)*, 2016, pp. 99–104.

[23] A. Izaddoost, E. Ogodo, and S. Prasai, "Enhanced data transmission platform in smart farms," in *ACM Proceedings of the International Conference on Omni-Layer Intelligent Systems (COINS)*, New York, NY, USA, 2019, pp. 58–61.

[24] H. M. Jawad, R. Nordin, S. K. Gharghan, A. M. Jawad, and M. Ismail, "Energy-efficient wireless sensor networks for precision agriculture: A review," *Sensors*, vol. 17, no. 8, p. 1781, 2017.

[25] A. Jøsang, J. Cho, and F. Chen, "Uncertainty characteristics of subjective opinions," in *2018 21st International Conference on Information Fusion (FUSION)*, July 2018, pp. 1998–2005.

[26] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer Publishing Company, 2016.

[27] F. Kiani, "Reinforcement learning based routing protocol for wireless body sensor networks," in *2017 IEEE 7th International Symposium on Cloud and Service Computing (SC2)*, 2017, pp. 71–78.

[28] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," 2017.

[29] B. P. L. Lau, N. Wijerathne, B. K. K. Ng, and C. Yuen, "Sensor fusion for public space utilization monitoring in a smart city," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 473–481, 2018.

[30] P. Laube and R. S. Purves, "How fast is a cow? cross-scale analysis of movement data," *Transactions in GIS*, vol. 15, no. 3, pp. 401–418, 2011.

[31] F. Li, X. Song, H. Chen, X. Li, and Y. Wang, "Hierarchical routing for vehicular ad hoc networks via reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1852–1865, 2019.

[32] A. Machado, V. Maran, I. Augustin, J. a. C. Lima, L. K. Wives, and J. P. M. de Oliveira, "Reasoning on uncertainty in smart environments," in *Proceedings of the 18th International Conference on Enterprise Information Systems*, ser. ICEIS 2016. Setubal, PRT: SCITEPRESS - Science and Technology Publications, Lda, 2016, p. 240–250. [Online]. Available: https://doi.org/10.5220/0005866502400250

[33] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," 2019.

[34] C. O. Mathuna, T. O'Donnell, R. V. Martinez-Catala, J. Rohan, and B. O'Flynn, "Energy scavenging for long-term deployable wireless sensor networks," *Talanta*, vol. 75, no. 3, pp. 613–623, 2008.

[35] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[36] A. Modarresi and J. Symons, "Modeling technological interdependency in iot - a multidimensional and multilayer network model for smart environments," in *2019 11th International Workshop on Resilient Networks Design and Modeling (RNDM)*, 2019, pp. 1–7.

[37] N. Naderializadeh, J. Sydir, M. Simsek, and H. Nikopour, "Resource Management in Wireless Networks via Multi-Agent Deep Reinforcement Learning," *arXiv e-prints*, p. arXiv:2002.06215, Feb. 2020.

[38] A. C. Nguyen, T. Pamuklu, A. Syed, W. S. Kennedy, and M. Erol-Kantarci, "Deep reinforcement learning for task offloading in uav-aided smart farm networks," *arXiv preprint arXiv:2209.07367*, 2022.

[39] J. O, D. Noh, and Y. Sohn, "Empirical test of Wi-Fi environment stability for smart farm platform," in *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, 2017, pp. 1–5.

[40] D. Popa, F. Pop, C. Serbanescu, and A. Castiglione, "Deep learning model for home automation and energy reduction in a smart home environment platform," *Neural Computing and Applications*, vol. 31, 05 2019.

[41] D. Preuveneers and Y. Berbers, "Architectural backpropagation support for managing ambiguous context in smart environments," in *Universal Access in Human-Computer Interaction. Ambient Interaction*, C. Stephanidis, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 178–187.

[42] N. Qi, K. Dai, F. Yi, X. Wang, Z. You, and J. Zhao, "An adaptive energy management strategy to extend battery lifetime of solar powered wireless sensor nodes," *IEEE Access*, vol. 7, pp. 88 289–88 300, 2019.

[43] G. Rocher, J.-Y. Tigli, and S. Lavirotte, "Probabilistic models toward controlling smart-* environments," *IEEE Access*, vol. 5, pp. 12 338–12 352, 2017.

[44] M. Roser. (2020) Future population growth. [Online]. Available: https://ourworldindata.org/future-population-growth

[45] N. Saxena, A. Roy, and J. Shin, "Chase: Context-aware heterogenous adaptive smart environments using optimal tracking for resident's comfort," in *Ubiquitous Intelligence and Computing*, J. Indulska, J. Ma, L. T. Yang, T. Ungerer, and J. Cao, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 133–142.

[46] J. Schulman, et al., "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[47] H. Sun, Q. Zhu, J. Ren, D. Barclay, and W. Thomson, "Combining image analysis and smart data mining for precision agriculture in livestock farming," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2017, pp. 1065–1069.

[48] J. P. S. Sundaram, W. Du, and Z. Zhao, "A survey on lora networking: Research problems, current solutions, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 371–388, 2019.

[49] Technical Marketing Workgroup, "A technical overview of lora® and lorawan™," LoRa Alliance®, Fremont, CA, USA, Nov. 2015.

[50] P. Tehrani, F. Restuccia, and M. Levorato, "Federated deep reinforcement

learning for the distributed control of nextg wireless networks," 2021. [Online]. Available: https://arxiv.org/abs/2112.03465

[51] *CC2640R2F SimpleLink™ Bluetooth® 5.1 Low Energy Wireless MCU*, Texas Instruments, 2016, rev. C. [Online]. Available: https://www.ti.com/product/CC2640R2F

[52] N. Twomey, T. Diethe, I. Craddock, and P. Flach, "Unsupervised learning of sensor topologies for improving activity recognition in smart environments," *Neurocomputing*, vol. 234, pp. 93–106, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231216315740

[53] M. Woolley, "The bluetooth low energy primer," Bluetooth SIG, Kirkland, WA, USA, Jun. 2022.

[54] K. Yang, C. Shen, and T. Liu, "Deep reinforcement learning based wireless network optimization: A comparative study," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2020, pp. 1248–1253.

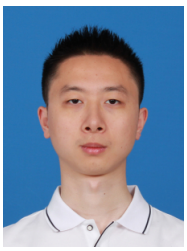
[55] C. Yoon, M. Huh, S.-G. Kang, J. Park, and C. Lee, "Implement smart farm with iot technology," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 2018, pp. 749–752.

[56] Q. Zhang, Y. Mahajan, I. Chen, D. Ha, and J. Cho, "An attack-resilient and energy-adaptive monitoring system for smart farms," in *IEEE Global Communications Conference (GLOBECOM)*, 2022, accepted.

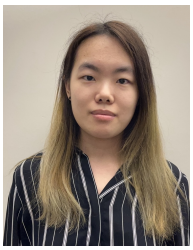
[57] S. Zhang, S. McClean, B. Scotney, and C. Nugent, "Learning under uncertainty in smart home environments," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 2083–2086.

[58] Z. Zhang, P. Wu, W. Han, and X. Yu, "Remote monitoring system for agricultural information based on wireless sensor network," *Journal of the Chinese Institute of Engineers*, vol. 40, no. 1, pp. 75–81, 2017.

[59] H. H. Zhuo, W. Feng, Q. Xu, Q. Yang, and Y. Lin, "Federated reinforcement learning," *CoRR*, vol. abs/1901.08277, 2019. [Online]. Available: http://arxiv.org/abs/1901.08277



Qisheng Zhang is currently a Ph.D. student in the Department Computer Science at Virginia Tech since 2019. He received the B.S. degree in mathematics from Shandong University in 2017 and the M.S. degree in mathematics from the University of Warwick in 2018. His research interests include network security and network science.



Dian Chen is currently a Ph.D. student in the Department of Computer Science at Virginia Tech since 2022. She received a B.S. degree in Computer Science from Worcester Polytechnic Institute in 2020 and an M.S. degree in Computer Science from Northwestern University in 2022. Her research interests include network security in cyber-physical systems.

Yash Mahajan received the M.S. degree in computer science and application from Virginia Tech in 2021 and the B.Tech. degree from Vellore Insititute of Technology (VIT) in 2019. He worked as a Research Assistant with the Trustworthy Cyberspace Laboratory, supervised by Dr. Jin-Hee Cho, from 2019 - 2021. He is currently a Software Engineer at Oracle Corporation. His research interest includes network security and social network analysis.

Ing-Ray Chen received a BS degree from the National Taiwan University and MS and Ph.D. degrees in computer science from the University of Houston, University Park. He is a professor at the Department of Computer Science at Virginia Tech. His research interests are primarily in trust, network, and service management as well as reliability, security, and performance analysis of mobile systems and wireless networks, including the Internet of Things, wireless sensor networks, service-oriented peer-to-peer networks, ad hoc networks, mobile social networks, mobile web services, mobile cloud services, and cyber-physical systems. Dr. Chen has published over 120 journal papers, with more than one-third of them appearing in IEEE/ACM Transactions journals. He is a recipient of the IEEE Communications Society William R. Bennett Prize in the field of Communications Networking, The 2023 IEEE ComSoc Network Operations & Management (CNOM) Test of Time Paper Award, and the U.S. Army Research Laboratory (ARL) Publication Award. Dr. Chen is a member of the IEEE and ACM.

Dong Sam Ha (Life Fellow) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, in 1974, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Iowa, Iowa City, IA, USA, in 1984 and 1986, respectively. Since Fall 1986, he has been a Faculty Member with The Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA. He is currently a Professor and the Founding Director of the Multifunctional Integrated Circuits and Systems (MICS) Group, composed of five faculty members and about 30 graduate students. His research interests include power management circuits for energy harvesting, intelligent analog and RF circuits and systems, wireless sensor nodes for smart farms, and high-temperature RF circuits and systems for downhole communications, jet engine monitoring, and extreme environment sensing.

Jin-Hee Cho (M'09; SM'14) is currently an Associate Professor in the Department of Computer Science at Virginia Tech since Aug. 2018 and a director of the Trustworthy Cyberspace Lab. Prior to joining Virginia Tech, she worked as a computer scientist at the U.S. Army Research Laboratory (USARL), Adelphi, Maryland, since 2009. Dr. Cho has published peer-reviewed technical papers in leading journals and conferences in the areas of cybersecurity, decision-making under uncertainty, and network science. She received the best paper awards in IEEE TrustCom'2009, BRIMS'2013, IEEE GLOBECOM'2017, 2017 ARL's publication award, and IEEE CogSima 2018. She is a winner of The 2015 IEEE Communications Society William R. Bennett Prize in the Field of Communications Networking and The 2023 IEEE ComSoc Network Operations & Management (CNOM) Test of Time Paper Award. Dr. Cho was selected for the 2013 Presidential Early Career Award for Scientists and Engineers (PECASE), the highest honor bestowed by the U.S. government on outstanding scientists and engineers in the early stages of their independent research careers. She is also a recipient of the 2022 Faculty Fellow Award from the College of Engineering at Virginia Tech. Dr. Cho earned a Ph.D. degree in computer science from Virginia Tech in 2008. She is currently serving on the editorial board as an associate editor in IEEE Transactions on Network and Service Management, IEEE Transactions on Services Computing, and The Computer Journal. She is a senior member of the IEEE and a member of ACM.