An Attack-Resilient and Energy-Adaptive Monitoring System for Smart Farms

Qisheng Zhang[†], Yash Mahajan[†], Ing-Ray Chen[†], Dong Sam Ha[‡], and Jin-Hee Cho[†]
[†]Department of Computer Science, Virginia Tech, Falls Church, VA, USA
[‡]Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA
{qishengz19, yashmahajan, irchen, dha, jicho}@vt.edu

Abstract—**In this work, we propose an energy-adaptive monitoring system for a solar sensor-based smart animal farm (e.g., cattle). The proposed smart farm system aims to maintain high-quality monitoring services by solar sensors with limited and fluctuating energy against a full set of cyberattack behaviors including false data injection, message dropping, or protocol non-compliance. We leverage *Subjective Logic* (SL) as the belief model to consider different types of uncertainties in opinions about sensed data. We develop two Deep Reinforcement Learning (DRL) schemes leveraging the design concept of uncertainty maximization in SL for DRL agents running on gateways to collect high-quality sensed data with low uncertainty and high freshness. We assess the performance of the proposed energy-adaptive smart farm system in terms of accumulated reward, monitoring error, system overload, and battery maintenance level. We compare the performance of the two DRL schemes developed (i.e., multi-agent deep Q-learning, MADQN, and multi-agent proximal policy optimization, MAPPO) with greedy and random baseline schemes in choosing the set of sensed data to be updated to collect high-quality sensed data to achieve resilience against attacks. Our experiments demonstrate that MAPPO with the uncertainty maximization technique outperforms its counterparts.**

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

I. Introduction

The smart farm research has been conducted to enhance monitoring animal welfare, or to support farmers' decisions for sensing data and environmental controls. However, there has been a lack of efforts to develop security-aware smart farm technologies for energy limited sensor networks. Food contamination due to bacteria, viruses, toxins, or chemicals has made hundreds of million people sick and nearly half a million people die according to the head of the World Health Organization (WHO) [6]. This can be increased exponentially if cyberattacks aiming at disrupting the process of correct data are not properly defended in farms, transportation systems, or food processing industrial control systems (ICSs). In particular, monitoring livestock in smart farms [14] plays a critical role toward increasing a farmer's revenue. If false or misleading data are received for the status of monitored animals, this will lead to improper actions, such as spread of disease or provision of wrong information to potential customers of the livestock [6].

In this work, we concern how accurate monitoring of cattle in a smart farm can be achieved in the presence of cyberattacks forging, modifying, dropping sensed data, or injecting false data from sensors to gateways or edge devices. Since most sensors for cattle are powered by batteries and attached to collars, they are incapable of measuring biometrics. Moreover, sensors using three AA-size batteries last only two to three days. Replacing or recharging batteries of sensors in every few days is laborious and not cost-effective for a typical farm. To address the problem, we consider a sensor attached to an animal's ear and powered by solar energy harvesting. However, the amount of harvested energy is small due to a small size solar panel. In addition, the harvested energy level fluctuates drastically as the animal and its ears with sensors move. Therefore, this calls for an energy-adaptive monitoring system for smart farms.

In this work, we make the following **key contributions**:

- We propose an energy-adaptive monitoring system for smart farms with solar sensors attached to cattle. This is the first work that aims at achieving high monitoring quality and energy maintenance of smart farms in the presence of high uncertainty and adversarial attacks.
- We leverage the merits of both deep reinforcement learning (DRL) [5] and a belief model, called *Subjective Logic* [8] (SL), to achieve our research goal. We develop an uncertainty maximization (UM) technique derived from SL for DRL agents running on gateways to collect high-quality sensed data with low uncertainty and high freshness for building an attack-resilient and energy-adaptive smart farm. Specifically, we develop two DRL-based schemes incorporated with the UM technique for energy-adaptive monitoring of smart farms. We demonstrate the effectiveness of the developed DRL-based schemes in terms of monitoring quality, system overload, and energy consumption in smart farm environments against baseline models.
- We show the robustness of the proposed smart farm monitoring system against a full set of cyberattack behaviors (i.e., protocol non-compliance, false data injection, and denial of service) that can happen to a smart farm.

II. Related Work

Energy-aware algorithms for wireless sensor networks (WSNs) have been proposed in various applications. A cluster-based routing protocol, called *QL-Cluster* [10], was proposed using Q-learning to continuously and efficiently monitor a patient's health. An adaptive energy management strategy was proposed for a solar-powered WSN with hybrid storage [13].

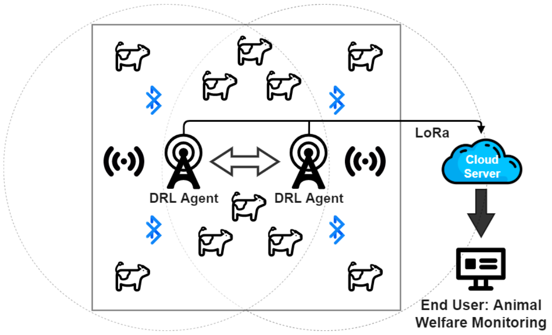


Fig. 1. Wireless Sensor Node-based Smart Farm Environment.

A sleep scheduling algorithm was proposed for rechargeable sensors based on a DRL algorithm [3]. Q-learning was also leveraged to control power for communications to build jamming attack-resistant, healthcare applications [1]. A sensor access control using DRL is proposed to adjust the access time and transmit power of the sensor based on the state of the sensor [2]. Relative to the above cited works which applied DRL to develop energy-efficient WSNs, our work develops an uncertainty maximization technique derived from SL for DRL agents running on gateways to collect high-quality sensed data with low uncertainty and high freshness for building an attack-resilient and energy-adaptive smart farm with sensors having limited and fluctuating energy.

III. Problem Statement

In this work, we are interested in identifying an optimal monitoring policy to minimize the monitoring error and system overload in a sensor network. Here, an update policy $P = \{p_1, p_2, \ldots, p_T\}$ consists T monitoring actions p_i with a total monitoring step T, where $p_i \in \mathcal{P}$ and $\mathcal{P}$ is a set of available monitoring actions in each monitoring step. When a dynamic sensor network $G = \{g_1, g_2, \ldots, g_T\}$ is given, we define the objective function as follows:

$$\underset{P=\{p_1,p_2,\ldots,p_T\}}{\arg\max} \sum_{i=1}^{T} f(g_i(p_1, p_2, \ldots, p_i)), \qquad (1)$$
$$s.t. \quad \forall i \in [1, T], p_i \in \mathcal{P},$$

where $f(g)$ returns the values depending on the evaluation function $f : g \mapsto -\mathcal{ME}(g) - \mathcal{OL}(g)$, aiming to minimize the monitoring error $\mathcal{ME}$ and system overload $\mathcal{OL}$, which are detailed in Section VI. It is a non-trivial task for the DRL agent to identify an optimal update policy that can meet multiple objectives based on the inherent difficulty of solving multi-objective optimization [4]. This is discussed with more details based on experimental results shown in Section VII.

IV. System Model

Network Model: The network consists of solar-powered wireless sensor nodes attached to cattle where the sensors transmit sensed data to LoRa gateways, which send the data to a cloud server. The LoRa gateways act as intermediaries between the sensors and the cloud server, allowing inexpensive, long-range (LoRa) connectivity for Internet-of-Things (IoT) devices via the standard IP protocol. In the given WSN-based smart farm environment, a low energy sensor can transmit its sensed data to a nearby sensor with excess energy via BLE (Bluetooth Low Energy), allowing transmitting the received data as well as its own sensed data to LoRa gateways via LoRa. 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 itself dissipates only 8 mW in active mode. The BLE protocol is intended for short distance communications with a maximum distance of 100 $meters$ and a data rate of 2 $Mbps$. In contrast, the LoRa protocol is for long distance communications with a distance of several km and a data rate of 27 $kbps$. BLE dissipates much less power than LoRA. For example, a Texas Instruments CC2640R2F microcontroller chip with an embedded BLE radio dissipates 11 mW during transmission [18]. Hence, the amount of energy to send one bit of data for the BLE radio is about 1,100 times smaller when compared with the LoRa radio of SAM R34/35 microcontroller chip. We assume each sensor is deployed with full charge of 5 kWs as an initial energy level. The power density for outdoor solar is about 10 mW/cm^2 and 0.1 mW/cm^2 for indoor light [11]. As such, the maximum harvestable power for outdoor solar is about 200 mW for a solar panel with the diameter of 5 cm, and it is about 2 mW for indoor light. Fig. 1 describes the high level overview of the considered network in this work.

A DRL agent is deployed on each LoRa gateway to identify which animal's sensed data are more needed than others to enhance the overall monitoring quality. We describe how the DRL agent identifies the sensed data in need in Section V. We assume that sensors communicate with each other via BLE without encryption considering its limited and fluctuating energy. Hence, attackers can intercept data in transmission and freely modify/forge data or inject false data. In addition, if a sensor is compromised by obtaining the sensor's key to be authenticated with gateways, the attacker can impersonate the sensor and transmit false data for other sensors with low energy and itself to a LoRa gateway. We assume that the gateways and the cloud server are trusted and will use secure communication channels based on existing security technologies. We leave the performance and security concerns between LoRa gateways and a cloud server for our future work. As shown in Fig. 1, multiple LoRa gateways each running a DRL agent can collaborate to each other in sharing collected sensed data received from sensors.

Node Model: Sensor nodes in a given smart environment are assumed solar-powered and deployed as implants and can transmit data on request. The energy levels of the sensor nodes vary throughout the day upon animals moving and from day to day. Hence, it is necessary to use energy efficiently for ubiquitous, steady, and persistent use. Each sensor node i is characterized by $\mathrm{sn}_t^i = [\mathrm{temp}_t^i, \mathrm{hb}_t^i, \mathrm{ma}_t^i, \mathrm{bl}_t^i]$, where temp_t^i refers to sensor node i's temperature at time t in Celsius, hb_t^i is the number of i's heart beat at time t, ma_t^i is i's speed at time t and bl_t^i is i's battery life at time t scaled in $[0, 100]$ in percent. A sensor's battery will be consumed mostly for data transmission to a LoRa gateway while communications between the sensor and other nearby sensors via BLE will consume about 1000 times less.

Based on the reported data by sensors to LoRa gateways, each DRL agent will identify what data is needed with high priority to achieve high monitoring quality in sensing accurate conditions of animals in the farm. To this end, we separate sensor nodes into low or high battery level sensors, denoted by LBS and HBS, respectively, by a recommended battery level T_M. In this way, sensor networks can be modeled as directed bipartite graphs as we are only interested in the transmissions from LBS to HBS. We will discuss the operations performed by DRL agents running on LoRa gateways in Section V. A cloud server will collect the sensed data of animals from multiple LoRa gateways and provide the aggregated results to end users. The scope of this work is to examine how the DRL agent on LoRa gateways can contribute to enhancing the quality of animal monitoring in the presence of cyberattacks and fluctuating energy levels.

Attack Model: We consider the following attacks:

- *Protocol non-compliance*: A compromised sensor node may not be compliant to the request by the DRL agents on LoRa gateways with a probability P_{NCA}. That is, when a DRL agent requests animal A's sensed data, the attacker may send another animal's sensed data or may not send A's data.
- *False data injection*: A compromised sensor can transmit forged/modified data to gateways or inject false data. In addition, man-in-the-middle attackers (MIMAs) can intercept data being transmitted in the middle and insert modified/forged data. We model these with the probabilities of data change by internal or external attackers, denoted by P_{IDA} and P_{EDA}, respectively.
- *Denial-of-Service (DoS)*: An attacker can send a request to nearby sensors requesting them to forward its data, thus exhausting their energy. We model this by probability P_{IDA}.

V. DRL-based Animal Monitoring

A. Uncertainty-Aware Animal Monitoring

Multiple gateways receive sensed data from solar sensors attached to cattle. They periodically report their collected sensed data to the cloud server. The multiple DRL agents sitting on the gateways will share knowledge on sensed data and associated information, including each animal's condition and associated uncertainty levels. The gateway will maintain a database of all animals' reported data.

Each observation item's condition (see Table I) will be reported based on the range of condition in $K = 3$ classes, which can be easily diagnosed as normal by the end user's diagnosis based on the received data from the cloud server. Since the gateway will report all animals' average conditions periodically to the cloud, it will collect sensed data from sensors and estimate their average values with the probability of each class and their corresponding uncertainty levels. To realize this, we will leverage Subjective Logic (SL) [9] to formulate an opinion on an animal's condition in a given category. Due to the space constraint, interested readers can refer to [9] for formulating a multinomial opinion and [8] for estimating two types of uncertainties (i.e., vacuity and dissonance) considered in this work.

In this work, we consider each report by sensors to a gateway as evidence. For example, 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) will be updated based on the mapping rule in SL [9]. SL stops updating an opinion when uncertainty becomes zero. This will prevent new reports to be effectively applied in a latest opinion. To avoid this, we will leverage the uncertainty (vacuity) maximization technique [9] to offset conflicting evidence while the amount of conflicting evidence can be transformed to vacuity of an opinion. Due to space constraint, readers can be referred to [9] for the uncertainty (vacuity) maximization function.

The animal condition in a given category X is estimated as an opinion, $\omega_X = (\boldsymbol{b}_X, u_X, \boldsymbol{a}_X)$ where $\boldsymbol{b}_X$ is a vector of belief masses, u_X is vacuity, and $\boldsymbol{a}_X$ is a vector of base rates (i.e., prior belief) to the corresponding belief masses. Note that dissonance can be estimated based on the belief masses where its formulation is given in [8]. The aforementioned opinion and uncertainties are estimated at the gateways based on the received sensed data from solar sensors.

B. DRL-based Monitoring Update

State space ($\mathcal{S}_t$): The state space $\mathcal{S}_t$ of the proposed smart farm monitoring system at time t is the sensor network g_t represented by the history action sequence. We evaluate the state space in four aspects: mean vacuity ($\hat{\text{vac}}_t^i$), mean dissonance ($\hat{\text{diss}}_t^i$), mean degree of freshness ($\hat{\text{fr}}_t^i$) [16], and mean battery life ($\hat{\text{bl}}_t^i$). Each component of the state at time t by agent i is the average value of each measurement by:

$$\hat{\text{vac}}_t^i = \frac{\sum_{j=1}^{n} \text{vac}_t^{ij}}{n}, \quad \hat{\text{diss}}_t^i = \frac{\sum_{j=1}^{n} \text{diss}_t^{ij}}{n},$$
$$\hat{\text{fr}}_t^i = \frac{\sum_{j=1}^{n} \text{fr}_t^{ij}}{n}, \quad \hat{\text{bl}}_t^i = \frac{\sum_{j=1}^{n} \text{bl}_t^{ij}}{n}, \tag{2}$$

where each component measures the mean value of the measurements for n animals, which is in the range of $[0, 1]$. The fr_t^{ij} is formulated by $\text{fr}_t^{ij} = e^{-\phi t}$, where t is time elapsed from the last update and ϕ is a constant to normalize the freshness. Note that we need at least one update to calculate fr_t^{ij}. Here $\text{vac}_t^{ij}, \text{diss}_t^{ij}, \text{bl}_t^{ij}$, and fr_t^{ij} are all scaled in $[0, 1]$. Note that we use each report (i.e., sensed data of an animal) as evidence to support a categorical class (i.e., below normal range, normal range, above normal range). An opinion towards a given animal is initialized with one evidence (i.e., $\mathbf{r}(x_i) = 1$) for each class i and $K = 3$.

Action space ($\mathcal{A}_t$): To minimize the monitoring error and system overload, each DRL agent will select k animals whose reports should be received with high priority. Note that a certain degree of information redundancy is needed since sensors may not be able to send their data due to battery or topology constraints. The utility of animal j is given by:

$$\text{utility}_{ij} = (1 - \text{vac}_t^{ij}) + (1 - \text{diss}_t^{ij}) + \text{fr}_t^{ij} + f(\text{bl}_t^{ij}), \tag{3}$$

TABLE I
EVD Dataset Description

Metric	Description
Serial	A unique animal identifier
Heart rate	Heart bits per min.
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

where $f(x)$ is defined by $f(x) = -(x - T_M)^2$ where x is set to bl_t^{ij}. Here T_M indicates the recommended level of a battery life to be maintained in a sensor node not to be depleted or overcharged under sun. Each agent will send out a list of top k animal IDs based on the ranks by Eq. (3) in ascending order. In this way, we consider three discrete actions to select top k animal IDs such that $k \in [0, \lfloor \frac{n}{2} \rfloor, n]$, where n is the total number of LBS. Therefore, the action space size is independent from n, which allows us to mitigate the overhead caused by unbounded action spaces and generalize the proposed framework to large sensor networks. Higher k will increase the number of unnecessary requests and cause the system overload, while lower k could hurt the monitoring quality. Thus, an action is to determine the optimal value of k in this context.

Immediate reward (r_t): This is formulated by $r_t^i = f(g_t(k_1, k_2, \ldots, k_t))$ based on $f(g_t) = -\mathcal{ME}(g_t) - \mathcal{OL}(g_t)$ given in Eq. (1) where k_i is an action taken in step i.

C. Data Aggregation at LoRa Gateways

Each sensor sends its sensed data consisting of the components in Table I. After a LoRa gateway receives the sensed data from each sensor, it will calculate an opinion about the sensed data received. The opinion consists of belief and uncertainty in two dimensions (i.e., vacuity, and dissonance masses). We call the opinion a *monitoring opinion* (MO) hereafter, of measured heart beats, temperature, and activity during the observation time period Δ. The Δ is determined based on the time elapsed since the last reported time. When the update is not made due to limited energy or a node cannot find any nearby sensor which can send its MO on behalf of itself, the MO may not be updated. This would not increase belief masses of the MO. In addition, if a sufficient number of sensed data is received for a certain animal, vacuity becomes close to zero, which makes an opinion stop being updated in each sensor in SL [8]. To receive new evidence and update the MO accordingly, we will use the uncertainty (vacuity) maximization (UM) technique with a threshold ρ (i.e., $0 < \rho < 1$) [9]. That is, if $u_X < \rho$, the MO will be updated based on the UM. A DRL agent at each gateway will calculate the average condition of the animal based on a set of sensed data and MOs received from multiple sensors at time t. We will use the multinominal multiplication technique in SL [9] to compute joint opinions where each opinion is independent to each other.

VI. Experimental Setup

Datasets: We utilized Virginia Tech (VT)'s *SmartFarm Innovation Network* (TM), an interconnected data collection and analysis hub throughout the state of Virginia to facilitate testing and demonstration of emerging technologies. We obtained sample datasets from a smart farm managed by VT's College of Agriculture and Life Sciences, collected from EmbediVet Implantable Temperature Devices (EVD), Halter Sensors, Heart Rate Sensors, and Implantable Temperature Sensors, as an example shown in Table I.

To realize adversarial attacks, we generated synthetic datasets using the sample datasets where each sensor attached to a cow is modeled to generate similar datasets while some of the sensors are compromised or external attackers exist.

Parameterization: We consider 20 cows moving around in the square farm area (A) of 40 acres (~ 160K square meters) and length (a) of 402 meters. We assume the farm area is fully covered by two gateways where each gateway is within the other's coverage. Furthermore, we assume both gateways have the same circular coverage. Specifically, they have locations as $(-\frac{a}{4}, 0)$ and $(\frac{a}{4}, 0)$ and the same radius as $\frac{\sqrt{5}a}{4}$. It can be proven that this setting allows the minimum coverage of the farm area for each agent. Fig. 1 describes the farm setting as our network model. We model the availability of solar energy based on sun's movement in a day by defining a probability distribution P over the farm area, where $P(x, y, t)$ indicates the probability of being charged if a sensor locates in (x, y) at time t in hour. For simplicity, we assume that $P(x, y, t)$ has a quadratic form at time t and can be written as $P(x, y, t) = \max\{0, -\frac{1}{6}(t - t_{xy})^2 + 1\}$, 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 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.

A solar sensor is attached to each cow's ear to capture its attributes. A cow's temperature follows a normal distribution with mean being 38 C and standard deviation being 1 C. The cow's heart beat is randomly ranged in $[60, 84]$ or $[48, 60]$ when it moves or not respectively. We use P_i^{mv} for cow i's moving probability. We assume random movements of cows and a normal distribution of their speeds with average 1.5m/s and standard deviation 0.1 m/s.

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 $[37.8, 39.2]$ in Celsius, $[48, 84]$ number of beats per min., and $[1, 2]$ meters per sec., respectively. We consider the number of uncertain evidence being three where each belief mass has the same base rate (i.e., 1/3) [9].

We consider the whole monitoring session as 24 consecutive hours. Each gateway takes an action to identify an optimal k with the interval $T_a = 60$ sec. We assume 5 HBS with initial battery level 1 and 15 LBS with random initial battery levels in $[0, T_M]$. All HBS have an update interval $T_u = 30$ sec. to send

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

Param.	Meaning	Value
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 certain attack (e.g., P_{NCA}, P_{IDA}, P_{EDA}, P_{MDA})	0.1
n	Total number of cows (sensors)	20
A	Area of a given smart farm	40 acres
a	length of a given smart farm	402 m
ρ	Uncertainty maximization threshold	0.05
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
ϕ	A constant factor to normalize freshness of a received sensed data	0.01

their sensed data to gateways. All 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. In this way, HBS can send its own data and another sensor's data if they receive a request from LBS. Specifically, the gateway agents would firstly calculate the consolidated priority list of received update lists from gateways within its wireless range. Then the agents would solve the maximum matching problem in bipartite sensor networks by the Hopcroft–Karp algorithm [7] to ensure the maximum number of transmissions being executed.

A sensor will consume about 2 mW in sleep mode and 170 $mW \times T_s$ upon sending a packet with sensed data where T_s refers to transmission time in sec. A sensor will be charged with 200 mW for outdoor solar and 2 mW for indoor light. Considering maximum 6 hours under direct sun, the sensor can be charged 200 $mW \times 6\ h = 4.32\ kWs$.

For simplicity, we set all attack probabilities as P_A. For inside attackers, we initially pick them among the total number of sensors at random. For outside attackers (i.e., MIMAs), we first pick a set of nodes at random transmitting messages intercepted by MIMAs with $P_{EDA}(P_A)$. We consider false data injection attacks by outside attackers which insert false temperature (e.g., $[30, 50]$ in Celsius), heart beats (e.g., $[30, 100]$), and moving activities (e.g., $[0, 5]$ meters per sec.), where the false information exhibits deviated ranges from the normal conditions. We summarize the key design parameters, their meanings, and default values used in Table II.

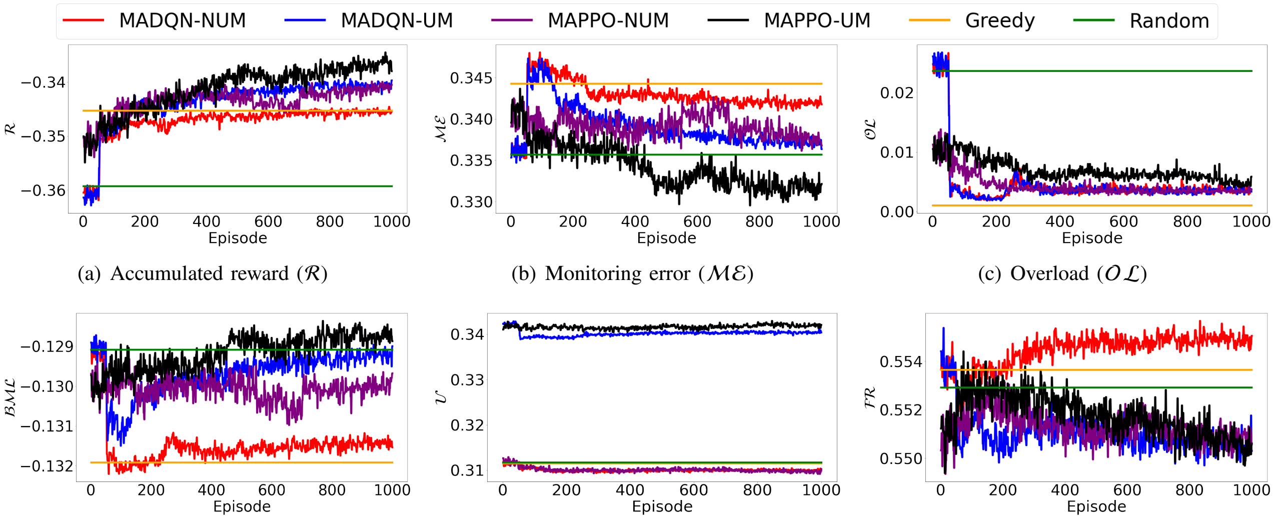
Comparing Schemes: (1) **Multi-Agent Deep Q-Learning (MADQN)** [17]: DQN [12] utilizes neural networks parameterized by $\boldsymbol{\theta}$ to represent an action-value function (i.e., Q-function). It assumes that the agent can fully observe the environment. By assigning a local Q-function to each agent, we can easily extend DQN to a Multi-Agent DRL algorithm, namely MADQN. We consider two types of MADQN with and without using uncertainty maximization (UM). We name them MADQN-UM and MADQN-NUM, respectively; (2) **Multi-Agent Proximal Policy Optimization (MAPPO)**: MAPPO extends the PPO [15] to a multi-agent environment to mitigate non-stationarity (i.e., uncertainty) from the environments when multiple agents share their actions and rewards to minimize the changes in the policy. This method deploys a central critic value function with stochastic policies. We consider two variants of MAPPO using UM or not and name them MAPPO-UM and MAPPO-NUM, respectively; (3) **Greedy**: It makes greedy choices by looking one-step ahead at each step and choosing k returning a maximum reward; and (4) **Random**: It is a baseline model where agents select optimal k at random.

Metrics: (1) **Accumulated reward ($\mathcal{R}$)** is the average value of the final accumulated rewards observed in each DRL agent at the end of the simulation; (2) **Monitoring error ($\mathcal{ME}$)** estimates the mean difference between the aggregated data of each animal's condition at each gateway and the ground truth data of the corresponding animal's condition; (3) **Overload ($\mathcal{OL}$)** evaluates the system overload by the fraction of the failed requests over all sent requests from LBS; (4) **Battery maintenance level** ($\mathcal{BML}$) as given by $f(\hat{\mathrm{bl}}_t^i)$ in Eq. (2) measures the difference between the recommended battery (T_M) and the current battery (bl_t^i) levels; (5) **Uncertainty ($\mathcal{U}$)** is the average uncertainties (i.e., vacuity and dissonance) of sensed data of all animals' conditions; and (6) **Freshness** ($\mathcal{FR}$) as given by $\hat{\mathrm{fr}}_t^i$ in Eq. (2) measures sensor data freshness.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

We conducted all experiments with 100 simulations on randomly generated farm environments based on Section VI where each data point represents the average of the simulation runs. For MADQN, we use 500 as batch size, 0.02 as learning rate. For MAPPO, we use 500 as batch size, 0.08 and 0.008 as learning rates for critic and actor networks, respectively.

Fig. 2 compares the six schemes with respect to DRL training episodes. Greedy and random schemes have flat learning curves since they are non-DRL schemes. We observe that MAPPO-UM and MADQN-UM outperform their corresponding NUM counterparts with respect to the accumulated reward ($\mathcal{R}$), monitoring error ($\mathcal{ME}$), and overload ($\mathcal{OL}$). This is because uncertainty maximization (UM) can update the uncertainty information from time to time, which reflects the sensor network status in a timely manner. As a result, MAPPO-UM and MADQN-UM also have the highest uncertainties ($\mathcal{U}$) in Fig. 2 (e). Our proposed uncertainty measures can estimate the update priority by considering the number of history updates for each sensor in a sensor network with acceptable dynamics. The proposed MADQN-based and greedy schemes fail to learn the effective monitoring policies. This is due to the non-stationary environment in the training phase of each agent [17]. Our proposed MAPPO-UM leverages uncertainty information in the stationary decision process and achieves the best performance among all comparing schemes. Note that MAPPO-UM also achieves the best battery efficiency ($\mathcal{BML}$) compared to other schemes in Fig. 2 (d). The overall performance order of the considered schemes is: MAPPO-UM $\geq$ MADQN-UM $\approx$ MAPPO-NUM $\geq$ MADQN-NUM$\geq$ Greedy $\geq$ Random.



(a) Accumulated reward ($\mathcal{R}$)

(b) Monitoring error ($\mathcal{ME}$)

(c) Overload ($\mathcal{OL}$)

(d) Battery maintenance level ($\mathcal{BML}$)

(e) Uncertainty ($\mathcal{U}$)

(f) Freshness ($\mathcal{FR}$)

Fig. 2. Performance of comparing schemes with respect to training episodes.

Since our multi-objective function has two conflict goals, different schemes could have very different policies. For example, MADQN-NUM and Greedy choose to minimize overload as their primary goals, as shown in Fig. 2 (b) and (c). Thus, they have the lowest overloads and highest monitoring error. Since low freshness records only come from LBS, they also have the highest freshness ($\mathcal{FR}$) in Fig. 2 (f) due to minimum transmissions from LBS to HBS, as in Fig. 2 (d).

VIII. Conclusions

From this study, we obtained the following **key findings**: (1) our proposed MAPPO-UM shows its strong resilience against attacks by achieving the best monitoring quality and minimum system overload; (2) our proposed MAPPO-UM intelligently leverages the uncertainty information and achieves the best energy maintenance level; and (3) the uncertainty maximization technique greatly enhances performance as it effectively updates the overall system uncertainty allowing the system to better balance the monitoring energy and monitoring quality.

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] G. Chen and et al., "Reinforcement learning based power control for in-body sensors in WBANs against jamming," *IEEE Access*, vol. 6, pp. 37 403–37 412, 2018.

[2] 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.

[3] H. Chen, et al., "A reinforcement learning-based sleep scheduling algorithm for desired area coverage in solar-powered wireless sensor networks," *IEEE Sensors Jour.*, vol. 16, no. 8, pp. 2763–2774, 2016.

[4] 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.

[5] H. Dong, et al., Ed., *Deep Reinforcement Learning Fundamentals, Research and Applications*. Springer, 2020.

[6] 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.

[7] J. E. Hopcroft and R. M. Karp, "An nˆ5/2 algorithm for maximum matchings in bipartite graphs," *SIAM Journal on computing*, vol. 2, no. 4, pp. 225–231, 1973.

[8] 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.

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

[10] F. Kiani, "Reinforcement learning based routing protocol for wireless body sensor networks," in *2017 IEEE 7th Int'l Symp. Cloud and Service Computing (SC2)*, 2017, pp. 71–78.

[11] C. O. Mathuna, et al., "Energy scavenging for long-term deployable wireless sensor networks," *Talanta*, vol. 75, no. 3, pp. 613–623, 2008.

[12] V. Mnih, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[13] N. Qi, et al., "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.

[14] J. R. Rosell-Polo, F. Auat Cheein, E. Gregorio, D. Andùjar, L. Puigdomènech, J. Masip, and A. Escolà, "Chapter three - advances in structured light sensors applications in precision agriculture and livestock farming," ser. Advances in Agronomy, D. L. Sparks, Ed. Academic Press, 2015, vol. 133, pp. 71 – 112. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0065211315001078

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

[16] R. Talak, et al., "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 15–28, 2019.

[17] A. Tampuu, T. Matiisen, D. Kodelja, I. Kuzovkin, K. Korjus, J. Aru, J. Aru, and R. Vicente, "Multiagent cooperation and competition with deep reinforcement learning," *PloS one*, vol. 12, no. 4, p. e0172395, 2017.

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