THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | LIBRARY

HKUST SPD - INSTITUTIONAL REPOSITORY

Title	Online System Identification and Optimal Control for Mission-critical IoT Systems over MIMO Fading Channels
Authors	Tang, Minjie; Cai, Songfu; Lau, Kin Nang
Source	IEEE Internet of Thing Journal, May 2022
Version	Accepted Version
DOI	10.1109/JIOT.2022.3175965
Publisher	IEEE
Copyright	© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This version is available at HKUST SPD - Institutional Repository (https://repository.ust.hk)

If it is the author's pre-published version, changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published version.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet of Things Journal

Online System Identification and Optimal Control for Mission-critical IoT Systems over MIMO Fading Channels

Minjie Tang, *Graduate Student Member, IEEE,* Songfu Cai, *Member, IEEE,* and Vincent K. N. Lau, *Fellow, IEEE*

***Abstract*—With the rapid development of mobile computing, mission-critical internet-of-things (IoT) systems have become popular. Typical mission-critical IoT systems may contain complicated unknown and unstable elements and it is of particular importance to identify and stabilize them as unstable systems may experience catastrophic consequences. We consider the identification and optimal control for a mission-critical IoT system over multiple-input multiple-output (MIMO) fading channels. First, we focus on the optimal control of the mission-critical IoT system, assuming that the system dynamics are known, and propose a novel stochastic-approximation-based algorithm to learn the optimal control solution for the IoT controller in an online manner. Second, we extend the optimal control framework to deal with the unknown mission-critical IoT system and propose a novel normalized-stochastic-gradient-descent-based algorithm to simultaneously identify and control the system in an online manner. Using Lyapunov stability analysis, we theoretically show the asymptotic optimality of the proposed learning algorithms. Numerical results are analyzed for our proposed scheme and for several state-of-the-art learning schemes in terms of the computational complexity, the convergence and the stability performance. Specifically, the proposed scheme can be implemented more than 50% faster than the state-of-the-art learning schemes. Moreover, the system identification performance of the proposed scheme can achieve a normalized system identification mean square error (MSE) of around 0.01 in 100 iterations. This is a substantial improvement compared to the baseline algorithms, where the normalized system identification MSE diverges.**

***Index Terms*—Mission-critical IoT system, online system identification, optimal control, Markov decision process, Lyapunov stability analysis.**

I. Introduction

A. Background

With the rapid development of mobile communication technologies, such as 5G, in recent years, the connection between the physical world and the network world has become increasingly close [1]–[4]. Under this development situation, an emerging communication paradigm, namely IoT, envisions a near future [5]–[7]. The IoT paradigm enables easy access and interaction among various devices and hence it boosts the development of a range of applications such as smart transportation in supply chains [8], smart homes for home automation [9], healthcare applications [10], agriculture [11], autonomous driving [12] and industrial control systems [13], that benefit from the framework.

Minjie Tang, Songfu Cai, and Vincent K. N. Lau are with the Department of ECE, The Hong Kong University of Science and Technology (HKUST). Songfu Cai is also with the Shenzhen Research Institute, HKUST, Shenzhen 518000, China (e-mail: mtangad@connect.ust.hk; scaiae@connect.ust.hk; eeknlau@ust.hk). The corresponding author is Songfu Cai.

Copyright (c) 2022 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

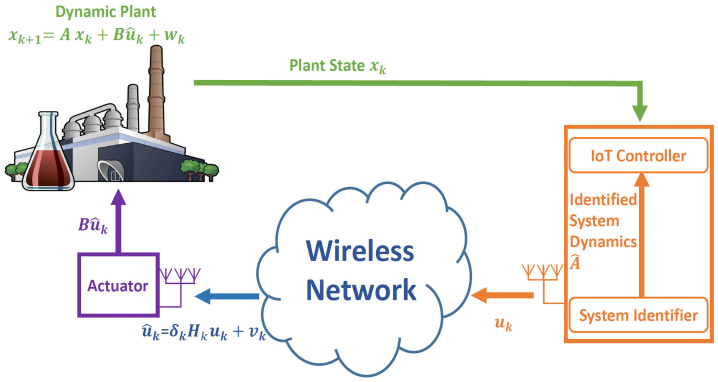


Fig. 1: Illustration of the architecture of the mission-critical IoT system.

B. Motivation and Challenges

In the aforementioned applications, the IoT systems may contain complicated unknown and unstable elements in the IoT network, and it is of a particular importance to identify and stabilize the systems as unstable systems may experience costly catastrophic consequences. For example, an unstable flock of aircraft may crash [14], and unstable power grids may explode [15]. Such kinds of IoT systems are considered to be mission-critical IoT systems. A typical mission-critical IoT system is comprised of an unknown and unstable *dynamic plant*, an *IoT controller* and an *actuator* collocated with the *dynamic plant*, as illustrated in Fig. 1. Specifically, the IoT controller receives the instantaneous plant state, identifies the system dynamics of the dynamic plant, and generates the control signal at each timeslot. The real-time control command will be delivered to the actuator over an unreliable wireless network to neutralize the instability of the system. The presence of the wireless network in between the IoT controller and the actuator will induce various impairments, such as fading and packet loss, and hence it will severely jeopardize the stability of the system.

System identification and optimal control for mission-critical IoT systems has been widely designed in an offline manner [16]–[20]. However, the required number of system state samples for the identification algorithm will grow with the increase of the system dimensions, which leads to an exponentially large sample memory and computational complexity

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

for large-scale IoT systems. Moreover, when there is a wireless network in between the IoT controller and the actuator, the offline control solutions obtained via these approaches cannot capture the transmission opportunities induced by the time-varying wireless channel, and the system will be unstable. Instead, the online system identification and optimal control algorithm aims at generating the system model and optimal control solutions using real-time system data, and hence it enables low computational complexity at each timeslot. Moreover, the system is likely to be stabilized in the presence of the wireless network between the IoT controller and the actuator as the obtained online optimal control solutions can fully capture the transmission opportunities induced by the time-varying wireless channels in between the IoT controller and the actuator. However, the online system identification and optimal control algorithm design for mission-critical IoT systems over a wireless network is extremely challenging, as summarized in the following.

Challenge 1: Optimality Condition. It is important to analyze the existence condition for the optimal control solutions for mission-critical IoT systems over the wireless network as it provides sufficient requirements under which the unstable systems are likely to be stabilized by optimal design of the control solutions. In conventional mission-critical IoT systems over static channels, such a requirement can be guaranteed by the well-known *controllability assumption* for the dynamic systems [21]. However, due to the time-varying wireless network between the IoT controller and the actuator, the systems are unlikely to be controllable at each timeslot, and hence the existence analysis for the optimal control solutions becomes challenging.

Challenge 2: Online Optimal Control Algorithm Design. It is challenging to design an optimal control algorithm for mission-critical IoT systems in an online manner when there is a wireless network between the IoT controller and the actuator because the time-varying wireless environment between the IoT controller and the actuator may jeopardize the stability performance of the control algorithm. When the dynamic plant is unknown, the situation becomes even worse as the IoT controller has no prior information on the internal behavior of the dynamic plant that it targets to stabilize. However, such information is critical for the optimal controller design.

Challenge 3: Online Identification Algorithm Design. It is challenging to design the online identification algorithm for mission-critical IoT systems in the presence of the wireless network in between the IoT controller and the actuator. This is because the random time-varying fading channels as well as the additive noise induced by the wireless network will result in the noisy plant states received at the IoT controller and this will jeopardize the identification performance. When the dynamic plant is unstable, the situation becomes even worse in the sense that the plant state might drift away from its desired value significantly throughout its sample path, which may further reduce the identification accuracy.

Challenge 4: Asymptotic Optimal Convergence Analysis. It is also important to theoretically analyze the asymptotic optimal convergence performance of the proposed online identification and control algorithm as it provides the stability guarantee for mission-critical IoT systems over the wireless network based on the proposed online identification and control algorithm. However, such a task is challenging due to the tight coupling between the identification and control algorithm in an online manner.

C. Related Works

Category 1: System Identification. System identification is important for an IoT-based communication framework. For example, in [22] and [23], the IoT devices are identified using federated learning for detection of the problematic IoT devices in an IoT network for security and privacy. For mission-critical IoT systems, identification of the unstable dynamic plants is also important as it provides the critical information for optimal control design to stabilize the systems. There exist some works that identify the system dynamics of the mission-critical IoT systems in an offline manner using a least-square-based approach [16], maximum-likelihood-based approach [17] and sampling-based algorithm [18]. However, these algorithms require an exponentially large sample memory for identification and hence they are not applicable for large-scale mission-critical IoT systems.

Online system identification for mission-critical IoT systems over static channels has been proposed using a recursive least square approach [24], [25], projected online identification algorithm [26], and recursive maximum-likelihood (ML) algorithm [27], [28]. However, brute-force applications of the above algorithm over the wireless network between the IoT controller and the actuator will lead to poor identification performance because of the random channel noise. Gradient-descent-based algorithms [29], [30] have been widely applied in learning and identification problems under a random wireless environment. However, the standard stochastic-gradient descent-based algorithms will not converge in the case of unstable mission-critical IoT systems because the boundness of the increment in the stochastic gradient descent (SGD) update is not guaranteed. Different from the above works, we propose a novel normalized-gradient-descent-based algorithm to identify the system dynamics of the unknown mission-critical IoT system over the wireless MIMO fading channels. Our proposed identification scheme can track the true system dynamics even in the presence of the unstable dynamic plant and wireless network due to the normalization operator in the SGD update.

Category 2: Optimal Control. Optimal control algorithms for mission-critical IoT systems have recently been reported in [31]–[33]. Specifically, in [31] and [32], the potential-learning-based policy iteration and value iteration algorithms are developed for adaptive optimal control. In [33], the Q-learning-based algorithm is proposed to solve the linear optimal state-feedback control problems. Note that in the above works, the static channels between the IoT controller and the actuator are assumed and brute-force applications of the control solutions in [31]–[33] under the random wireless network will lead to the "curse of dimensionality" [34] issue induced by the extended time-varying wireless channel state with infinitely many possible fading realizations at each timeslot. As a result,

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

the control solutions via [31]–[33] will deviate from the optimal control solution when a wireless network is considered, and the systems will be unstable. Different from above works, we consider the online data-driven optimal control design for mission-critical IoT systems in the presence of a wireless network between the IoT controller and the actuator. Specifically, we exploit the i.i.d. properties of the wireless channel state and learn the control solution by learning the equivalent reduced-state value function (where the wireless channel state is reduced in the value function) via a stochastic-approximation-based online approach. The control algorithm via the proposed approach learns the optimal control solution even in the presence of the wireless network since there is no "curse of dimensionality" issue in our proposed approach due to the state reduction.

D. Contributions and Organization

We propose a novel online approach for simultaneous identification and optimal control of an unstable mission-critical IoT system over MIMO fading channels. The following summarizes the key contributions of this work.

- **Closed-form Optimality Condition.** We provide the closed-form characterization of the sufficient condition for the existence of the optimal control solution over the wireless MIMO fading channels by analyzing the equivalent reduced-state optimality equation via positive-semidefinite cone decomposition.
- **Design of the Online Optimal Control Algorithm.** We propose a novel online optimal control approach for the mission-critical IoT system with both known and unknown system dynamics over wireless MIMO fading channels via the stochastic-approximation-based algorithm.
- **Design of the Online Identification Algorithm.** We also provide a novel normalized-gradient-descent-based algorithm to identify the system dynamics of the unknown mission-critical IoT system over the wireless MIMO fading channels.
- **Asymptotic Optimal Convergence Analysis.** Using the Lyapunov analysis method, we theoretically show that the control solution via the proposed control algorithm for the IoT controller will converge to the optimal control solution, and the identified system dynamics via the proposed identification algorithm will converge to the true system dynamics asymptotically under the consideration of the wireless MIMO channels between the IoT controller and the actuator.

The remainder of this paper is organized as follows. In section II, we outline the key components of the mission-critical IoT system. In Section III, we focus on the optimal control of the mission-critical IoT system with known system dynamics, and we analyze the existing conditions for the optimal control solution over static and MIMO fading channels. Based on this, we propose a novel stochastic-approximation-based algorithm to learn the optimal control solution for the IoT controller in an online manner. In Section IV, we extend the optimal control framework proposed in Section III to deal with the unknown mission-critical IoT system, and propose a novel normalized-stochastic-gradient-descent-based algorithm to simultaneously identify and control the IoT system. Simulation results on the performance are discussed in Section V, followed by the concluding remarks in Section VI.

Notation: Uppercase and lowercase boldface denotes matrices and vectors, respectively. The operator $(\cdot)^T$ and $\mathrm{Tr}(\cdot)$ is the transpose, and trace of a matrix, respectively. $\mathrm{Diag}(a, b, ...)$ is a diagonal matrix with the diagonal elements being $\{a, b, ...\}$. $\mathbb{R}^{m\times n}$ and $\mathbb{S}_+^a$ denotes the set of $m \times n$ dimensional real matrices and the set of $a \times a$ dimensional positive definite matrices, respectively. $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ is the spectral norm of a matrix $\mathbf{A}$ and the Frobenius norm of a matrix $\mathbf{A}$, respectively. $\mathbf{1}_{\{a\geq 0\}} \in \{0, 1\}$ is the indicator function and $\mathbf{1}_{\{a\geq 0\}} = 1$ if and only if the statement $a \geq 0$ holds true.

II. Mission-critical IoT System Model

In this section, we introduce the architecture of the mission-critical IoT system, which is composed of dynamic plant model, wireless MIMO fading channel model and the stability metric of the mission-critical IoT system.

A. Dynamic Plant Model

We consider a time-slotted mission-critical IoT system with S state variables. We assume the IoT controller and the system identifier are equipped with N_t transmission antennas and the actuator is equipped with N_r receiving antennas. The physical dynamic plant of the mission-critical IoT system is modelled by a set of first order coupled linear difference equations representing the evolution of the system state, as follows:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\widehat{\mathbf{u}}_k + \mathbf{w}_k, \quad k = 0, 1, 2, ..., \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^{S\times 1}$ is the system state variable, $\mathbf{A} \in \mathbb{R}^{S\times S}$ is the plant dynamics, $\mathbf{B} \in \mathbb{R}^{S\times N_r}$ is the control input matrix, $\widehat{\mathbf{u}}_k \in \mathbb{R}^{N_r\times 1}$ is the received control signal from the IoT controller and $\mathbf{w}_k \in \mathbb{R}^{S\times 1}$ is the system noise with zero mean and finite noise covariance matrix $\mathbf{W} \in \mathbb{S}_+^S$. We assume the dynamic evolution (1) is potentially unstable[1] and we have the following assumption on the dynamic plant model (1).

Assumption 1: (Controllability of Dynamic Plant) The dynamic plant (1) is controllable, i.e., the matrix $(\mathbf{A}, \mathbf{B}) \triangleq [\mathbf{B}, \mathbf{AB}, \mathbf{A}^2\mathbf{B}, ..., \mathbf{A}^{S-1}\mathbf{B}] \in \mathbb{R}^{S\times SN_r}$ has a full row rank ($\mathrm{rank}((\mathbf{A}, \mathbf{B})) = S$) [21]. ■

The above controllability assumption is a common prerequisite assumption in the study of closed-loop control systems [35]–[37]. This because an unstable linear time-invariant (LTI) system can be stabilized by feedback control only if it is controllable. In other words, if the LTI system is not controllable, there does not exist any control action that can stabilize the system. As a result, our work focuses on analyzing the existence of and finding the optimal control action for the controllable LTI systems. Therefore, Assumption 1 of the controllability will not affect the novelty of the proposed methods.

[1]"The dynamic evolution (1) is potentially unstable" means that plant dynamics $\mathbf{A}$ contains possibly unstable eigenvalues, i.e., $\|\mathbf{A}\| > 1$.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet of Things Journal

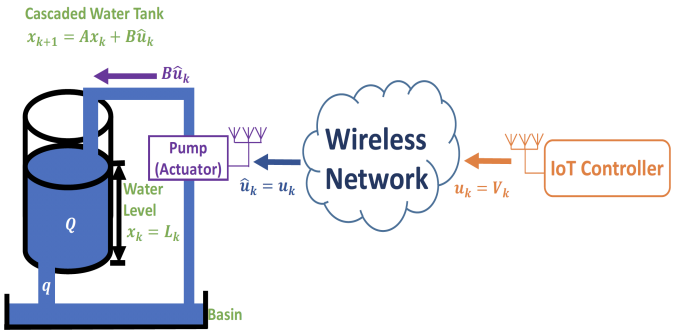


Fig. 2: Illustration of a cascaded water tank IoT system.

The dynamic plant model in (1) can embrace many practical scenarios of mission-critical IoT systems, such as the temperature and pressure control of a chemical factory or the velocity of vehicles in a platooning system. As an illustrative example, we consider a controlled cascaded water tank IoT system as shown in Fig. 2. The IoT system consists of a water tank equipped with orifices and a pump. The pump of the water tank is connected to the IoT controller via a wireless network. K_p is the pump constant and g is the gravitational acceleration. Q is the cross-sectional area of the water tank and q is the cross-sectional area of the outflow orifice at the bottom of the water tank. The system samples the water level of the tank once per timeslot with slot duration τ. At the k-th slot, L_k and V_k denotes the water level of the tank and the voltage applied to the pump, respectively. The system state is $x_k = L_k$, which represents the variations of the tank water level. The control action u_k denotes the desired control signal for variations of the voltages applied to the water tank generated by the IoT controller. $\widehat{u}_k$ is the received signal at the pump for variations of the voltages applied to the water tank. The evolution of the system state x_k is characterized by $x_{k+1} = A\mathbf{x}_k + B\widehat{u}_k$, where $B = \frac{(\ln A)^{-1}(A-1)K_p}{Q}$ and $\widehat{u}_k = u_k$.

B. Wireless MIMO Fading Channel Model

We model the wireless communication channels from the IoT controller and the system identifier to the actuator as wireless MIMO fading channels, as illustrated in Fig. 1. At the k-th time slot, the received control signal $\widehat{\mathbf{u}}_k \in \mathbb{R}^{N_r \times 1}$ at the actuator is given by

$$\widehat{\mathbf{u}}_k = \delta_k \mathbf{H}_k \mathbf{u}_k + \mathbf{v}_k, \tag{2}$$

where $\mathbf{u}_k \in \mathbb{R}^{N_t \times 1}$ is the control action of the IoT controller, and $\delta_k \in \{0, 1\}$ is the i.i.d. random access variable for the IoT controller with $\Pr(\delta_k = 1) = p$. $\mathbf{H}_k \in \mathbb{R}^{N_r \times N_t}$ is the wireless MIMO fading matrix from the IoT controller and the system identifier to the actuator, and $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_{N_r})$ is the additive Gaussian noise. We have the following assumption on $\mathbf{H}_k$.

Assumption 2: (Wireless MIMO Fading Channel Model) The random wireless MIMO channel realization $\mathbf{H}_k$ from the IoT controller and the system identifier to the actuator is assumed to be an i.i.d. Gaussian random process with zero mean and unit variance [38]. ■

The wireless MIMO fading channel model in Assumption 2 is widely adopted in existing literature on wireless communications [39]–[41]. Specifically, let us first consider a single-input single-output (SISO) communication system. When propagating through a wireless medium, a radio frequency signal suffers from the following effects [42]: 1) path loss, i.e., the reduction in power density (at tenuation) of an electromagnetic wave as it propagates through space; 2) shadowing due to the absorption, scattering and reflection of the electromagnetic wave; and 3) fading, i.e., the phase cancellation or reinforcement due to multipaths. Since the path loss and shadowing typically change slowly with time and are compensated via automatic gain control and power control, the input–output relationship of the SISO communication channel is represented as $y_k = h_k x_k + v_k$, where $y_k \in \mathbb{R}$ is the received signal, $x_k \in \mathbb{R}$ is the transmitted symbol and $h_k \in \mathbb{R}$ is the random SISO channel coefficient, which is i.i.d. Gaussian distributed with zero mean and unit variance and models the channel fading. $v_k \in \mathbb{R}$ is the additive channel noise. The MIMO communication channel model in Assumption 2 is the generalization of the SISO communication channel via equipping the transmitter and the receiver with multiple antennas. Specifically, a MIMO channel is composed of a collection of SISO channels. As such, the MIMO channel in Assumption 2 with N_t transmit antennas and N_r receive antennas is modeled as an i.i.d. Gaussian random matrix $\mathbf{H}_k \in \mathbb{R}^{N_r \times N_t}$ with each element being i.i.d. Gaussian distributed with zero mean and unit variance. Our work focuses on the system identification and optimal control design over the Gaussian MIMO fading channels. As a result, the correctness and novelty of the proposed methods will not be affected by Assumption 2.

C. Stability Metric of the Mission-critical IoT System

The goal of the mission-critical IoT system is to stabilize the potentially unstable plant (1) with limited wireless communication resources [43], [44]. Note that it is important to maintain stability of the mission-critical IoT system. This is because the instability of such a system will lead to catastrophic consequences. Such consequences include the explosion of chemical factories in industrial IoT systems [45] and traffic accidents in autonomous vehicle control IoT systems [46]. Specifically, we have the following metric on the stability of the mission-critical IoT system [43].

Definition 1: (Stability Metric of the Mission-critical IoT System) The mission-critical IoT system with dynamic evolution (1) is stable if

$$\limsup_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left\{\|\mathbf{x}_k\|^2\right\} < \infty, \tag{3}$$

where the expectation is taken with respect to the randomness of the plant noise $\mathbf{w}_k$, the channel noise $\mathbf{v}_k$, the wireless MIMO fading matrix $\mathbf{H}_k$, and the random access of the IoT controller δ_k. ■

To maintain stability of the mission-critical IoT system, the IoT controller should be involved and the optimal control policy may be considered. Note that the optimal control policy for the IoT controller will be different under different wireless communication channels between the IoT controller and the actuator. This is because the optimal control gain

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

for the IoT controller should adapt to not only the dynamic system state realization but also the realization of the wireless fading between the IoT controller and the actuator to capture the dynamic urgency of the control and good transmission opportunities induced by the fading channels. In the following sections, we first introduce the optimal control solution for the IoT controller under static and wireless MIMO fading channels between the IoT controller and the actuator when the system dynamics $\mathbf{A}$ is known. We provide the associated stability conditions for the mission-critical IoT system. After that, we further extend the optimal control framework to deal with a mission-critical IoT system with unknown unstable system dynamics. We provide an efficient online learning algorithm to simultaneously identify the system dynamics and learn the optimal control solutions for the mission-critical IoT system in the presence of the MIMO fading channels between the IoT controller and the actuator.

III. Online Optimal Control for the Mission-critical IoT System over the Wireless Channels with Known System Dynamics

In this section, we first focus on the optimal control solutions of an IoT controller for a mission-critical IoT system under the static and MIMO fading channels when the system dynamics $\mathbf{A}$ is known.

A. Optimal Control Solution of the Mission-critical IoT System over Static Channels

Optimal control for mission-critical IoT systems over static channels has been widely studied as optimal control for LTI systems in the existing literature [47]–[49]. The system dynamics of the mission-critical IoT system can be considered as a special case of system dynamics (1) in our case by restricting $\widehat{\mathbf{u}}_k = \mathbf{H}\mathbf{u}_k, 1 \leq k \leq K$, where $\mathbf{H} \in \mathbb{R}^{N_r \times N_t}$ is a constant matrix. Specifically, a control policy π for the IoT controller consists of a sequence of mappings $\pi = \{\Omega^0, \Omega^1, ...\}$. The mapping $\Omega^k : \mathbb{R}^{S\times 1} \rightarrow \mathbb{R}^{N_t \times 1}$ at the k-th timeslot is a mapping from the system state $\mathbf{x}_k$ to the control action of the IoT controller $\mathbf{u}_k$, i.e., $\mathbf{u}_k = \Omega^k(\mathbf{x}_k)$. $r(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{x}_k^T\mathbf{Q}\mathbf{x}_k + \mathbf{u}_k^T\mathbf{R}\mathbf{u}_k$ is the per-stage cost reflecting the quadratic cost of state $\mathbf{x}_k^T\mathbf{Q}\mathbf{x}_k$, and the control cost $\mathbf{u}_k^T\mathbf{R}\mathbf{u}_k$, $\mathbf{Q} \in \mathbb{S}_+^S$ and $\mathbf{R} \in \mathbb{S}_+^{N_r}$ are the weighting matrices.

The optimal control for noisy plant can be formulated using the infinite horizon ergodic control formulation given by [50].

Problem 1: (Optimal Control Problem for a Mission-critical IoT System over Static Channels)

$$\begin{aligned} \min_{\pi} \mathcal{J}^{\pi} &= \limsup_{K\rightarrow\infty} \frac{1}{K}\mathbb{E}_{\mathbf{w}_k,\mathbf{v}_k}\left\{\sum_{k=0}^{K} r(\mathbf{x}_k, \mathbf{u}_k)\right\} \\ \text{subject to} \quad & \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{H}\mathbf{u}_k + \mathbf{B}\mathbf{v}_k + \mathbf{w}_k, \end{aligned} \tag{4}$$

where the expectation in the objective function of Problem 1 is w.r.t. the random non-state variables $\mathbf{w}_k$ and $\mathbf{v}_k$.

The sufficient condition for the existence of the solution to Problem 1 is given by

Lemma 1: (Sufficient Condition for the Existence of the Solution to Problem 1) Problem 1 has a solution (i.e., $\limsup_{K\rightarrow\infty} \frac{1}{K}\mathbb{E}[\sum_{k=0}^{K} r(\mathbf{x}_k, \mathbf{u}_k)] < \infty$) if $(\mathbf{A}, \mathbf{BH}) \triangleq [\mathbf{B}, \mathbf{ABH}, \mathbf{A}^2\mathbf{BH}, ..., \mathbf{A}^{S-1}\mathbf{BH}] \in \mathbb{R}^{S\times SN_t}$ and $(\mathbf{A}^T, \sqrt{\mathbf{Q}}^T) \triangleq [\sqrt{\mathbf{Q}}^T, \mathbf{A}^T\sqrt{\mathbf{Q}}^T, ..., [\mathbf{A}^{S-1}]^T\sqrt{\mathbf{Q}}^T] \in \mathbf{R}^{S\times S^2}$ has a full row rank, i.e., rank$((\mathbf{A}, \mathbf{BH}))$ = rank $((\mathbf{A}^T, \sqrt{\mathbf{Q}}^T)) = S$.

Proof: Please refer to [50]. ■

The optimal control solution to Problem 1 can be obtained via the solution of the Bellman optimality equation for Problem 1 given by [51]:

$$\theta^* + V^*(\mathbf{x}_k) = \min_{\mathbf{u}_k}[r(\mathbf{x}_k, \mathbf{u}_k) + \mathbb{E}_{\mathbf{w}_k\mathbf{v}_k}[V^*(\mathbf{x}_{k+1})|\mathbf{x}_k, \mathbf{u}_k)]], \tag{5}$$

where the value function $V^*(\mathbf{x}_k) = \mathbf{x}_k^T\mathbf{P}\mathbf{x}_k$, $\mathbf{P} \in \mathbb{S}_+^S$ is a positive definite matrix and the optimal average cost $\theta^* = \mathrm{Tr}(\mathbf{PW} + \mathbf{B}^T\mathbf{PB})$.

Using the structure properties of the value function $V^*(\mathbf{x}_k)$ as well as the optimal average cost θ^* in (5), and applying (1) to (5), the closed-form optimal control solution to (5) has a linear state-feedback form given by

$$\begin{aligned} \mathbf{u}_k^* &= \mathrm{argmin}_{\mathbf{u}_k}[r(\mathbf{x}_k, \mathbf{u}_k) + \mathbb{E}_{\mathbf{w}_k,\mathbf{v}_k}[V^*(\mathbf{x}_{k+1})|\mathbf{x}_k, \mathbf{u}_k]] \\ &= -(\mathbf{R} + \mathbf{H}^T\mathbf{B}^T\mathbf{PBH})^{-1}\mathbf{H}^T\mathbf{B}^T\mathbf{PA}\mathbf{x}_k. \end{aligned} \tag{6}$$

Note that when the wireless channel matrix $\mathbf{H}_k$ and the random access variable δ_k are random and i.i.d. in each time slot, brute-force application of the control solution (6) will lead to poor stability performance. For instance, the system dynamics in (1) becomes

$$\mathbf{x}_{k+1} = (\mathbf{A} - \delta_k\mathbf{B}\mathbf{H}_k(\mathbf{R} + \mathbf{H}^T\mathbf{B}^T\mathbf{PBH})^{-1}\mathbf{H}^T\mathbf{B}^T\mathbf{PA})\mathbf{x}_k + \mathbf{w}_k. \tag{7}$$

Note that if the system dynamics $\mathbf{A}$ is unstable, i.e., $\|\mathbf{A}\| > 1$ and the wireless fading $\mathbf{H}_k$ follows an i.i.d. Gaussian random process with zero mean and unit variance, $\limsup_{k\rightarrow\infty} \mathbb{E}\|\mathbf{x}_k\|^2 = \infty$ and the mission-critical IoT system will be unstable.

> Challenge 1: Optimal Control Solution of the IoT controller for a Mission-critical IoT System over MIMO Fading Channels.

In the following, we shall extend the solution to ergodic optimal control over general random fading channels between the IoT controller and the actuator.

B. Optimal Control Solution of the Mission-critical IoT System over the MIMO Fading Channels

Note that the instability of the mission-critical IoT system over the MIMO fading channels under the optimal control solution $\mathbf{u}_k^*$ to Problem 1 is induced by the independency between $\mathbf{u}_k^*$ and the wireless channel state $\delta_k\mathbf{H}_k$. As a result, to properly formulate the optimal control problem over the MIMO fading channels in the sense that the optimal control solution is likely to stabilize the system, the objective function in the optimal control problem over the MIMO fading channels should incorporate the wireless channel state $\delta_k\mathbf{H}_k$. Since the non-state random variables will be averaged out in the objective function of the optimal control problem, it is desirable to extend the state space $\mathbf{S}_k = (\mathbf{x}_k, \delta_k\mathbf{H}_k) \in \mathcal{S}$ to include both the system state $\mathbf{x}_k$ and the channel state $\delta_k\mathbf{H}_k$

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

in the formulation of the optimal control problem over MIMO fading channels. Correspondingly, the dynamic control policy Ω_u is extended into a mapping from $\mathcal{S} \to \mathcal{U}$, so that the control action of the IoT controller can be adaptive to both the realizations of the system state $\mathbf{x}_k$ (reflecting the *urgency of the control*) and the channel state $\delta_k \mathbf{H}_k$ (revealing *the transmission opportunities in the wireless channels*). Furthermore, the per-stage cost should include both the state error $\mathbf{x}_k^T \mathbf{Q} \mathbf{x}_k$, the actuator cost $\mathbb{E}_{\mathbf{v}_k}[\widehat{\mathbf{u}}_k^T \mathbf{M} \widehat{\mathbf{u}}_k]$ and the transmission cost $\mathbf{u}_k^T \mathbf{R} \mathbf{u}_k$. As a result, the per-stage cost is given by

$$r(\mathbf{S}_k, \mathbf{u}_k) = \mathbf{x}_k^T \mathbf{Q} \mathbf{x}_k + \mathbf{u}_k^T \mathbf{R} \mathbf{u}_k + \mathbb{E}_{\mathbf{v}_k}[\widehat{\mathbf{u}}_k^T \mathbf{M} \widehat{\mathbf{u}}_k]. \quad (8)$$

In addition, the extended state sequence $\mathbf{S}_k$ is a *controlled Markov process* [52] with the transition kernel given by

$$\begin{aligned} &\Pr[\mathbf{S}_{k+1}|\mathbf{S}_k, \mathbf{u}_k] \\ &= \Pr[\mathbf{H}_{k+1}|\mathbf{S}_k, \mathbf{u}_k] \Pr[\mathbf{x}_{k+1}|\mathbf{S}_k, \mathbf{u}_k] \\ &= \Pr[\mathbf{H}_{k+1}] \Pr[\mathbf{x}_{k+1}|\mathbf{S}_k, \mathbf{u}_k]. \end{aligned} \quad (9)$$

Substitute (2) into (1), we have the equivalent linear system dynamics given by

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \delta_k \mathbf{B}\mathbf{H}_k \mathbf{u}_k + \mathbf{B}\mathbf{v}_k + \mathbf{w}_k. \quad (10)$$

Based on these, the optimal control problem for the mission-critical IoT system over the MIMO fading channels can be formulated as an infinite horizon ergodic control problem over the extended state space $\mathcal{S}$.

Problem 2: (Optimal Control Problem for a Mission-critical IoT System over MIMO Fading Channels))

$$\begin{aligned} &\min_{\mathcal{U}=\{\mathbf{u}_0, \mathbf{u}_1, \ldots\}} \limsup_{K\to\infty} \frac{1}{K} \mathbb{E}_{\mathbf{w}_k, \mathbf{v}_k} \left\{ \sum_{k=0}^{K} r(\mathbf{S}_k, \mathbf{u}_k) \right\} \\ &\text{subject to} \quad \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \delta_k \mathbf{B}\mathbf{H}_k \mathbf{u}_k + \mathbf{B}\mathbf{v}_k + \mathbf{w}_k. \end{aligned} \quad (11)$$

We first provide the sufficient condition for the existence of the solution to Problem 2 as follows.

Theorem 1: (Sufficient Condition for the Existence and Uniqueness of the Solution to Problem 2) Let the SVD of $\delta_k \mathbf{B}\mathbf{H}_k \left(\mathbf{H}_k^T \mathbf{M} \mathbf{H}_k + \mathbf{R}\right)^{-1} \mathbf{H}_k^T \mathbf{B}^T$ be

$$\delta_k \mathbf{B}\mathbf{H}_k \left(\mathbf{H}_k^T \mathbf{M} \mathbf{H}_k + \mathbf{R}\right)^{-1} \mathbf{H}_k^T \mathbf{B}^T = \mathbf{V}_k^T \zeta_k \mathbf{V}_k, \quad (12)$$

with the diagonal elements of ζ_k in descending order. Let $\operatorname{rank}\left(\delta_k \mathbf{B}\mathbf{H}_k \mathbf{H}_k^T \mathbf{B}^T\right) = \gamma_k$ and $\mathbf{\Pi}_k = \begin{bmatrix} \mathbf{I}_{\gamma_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{S\times S}$. Problem 2 has a solution if the following condition (13) is satisfied:

$$\left\| \mathbb{E}\left[\mathbf{A}^T \mathbf{V}_k^T \left(\mathbf{I}_S - \mathbf{\Pi}_k\right) \mathbf{V}_k \mathbf{A} \right] \right\| < 1. \quad (13)$$

Proof: Please see Appendix A. ■

The optimality condition (13) in Theorem 1 delivers some key system design insights. Specifically,

- **Impact of Dynamic Plant.** Less unstable dynamic plant, i.e., $\|\mathbf{A}\|$ is small, is more favorable for the optimality condition (13) because a less unstable dynamic plant will lead to the L.H.S. of (13) being smaller and the optimality condition (13) easier to satisfy. This physically means that a less unstable mission-critical IoT system is easier to be controlled compared to an unstable one.
- **Impact of the Random Access of the Controller.** A larger activation probability of the controller p is favorable for the optimality condition (13) because a larger p will lead the rank of the term $\mathbf{\Pi}_k$ in the L.H.S. of (13) to being larger statistically, which makes the value of the L.H.S. of (13) smaller and the optimality condition (13) more likely to be satisfied. This physically means that a larger activation probability of the IoT controller would benefit the stability of the mission-critical IoT system.
- **Impact of Communication Antennas.** A larger number of transmission antennas N_t and receiving antennas N_r is favorable for the optimality condition (13) because a larger N_t and N_r would lead the term $\mathbf{I}_S - \mathbf{\Pi}_k$ in the L.H.S. of (13) to be smaller. This would further lead to a smaller value of the L.H.S. of (13), and hence the optimality condition (13) would be easier to satisfy. This physically means that more communication antennas will enhance the control ability for the mission-critical IoT system.

The optimal control solution to Problem 2 can be obtained via the solution of the Bellman optimality equation for Problem 2, as summarized in the following Theorem.

Theorem 2: (Bellman Optimality Equation for Problem 2) Under the condition (13) in Theorem 1, the optimal solution to Problem 2 is equivalent to the solution of the Bellman optimality equation given by

$$\theta^* + V^*(\mathbf{S}_k) = \min_{\mathbf{u}_k} [r(\mathbf{S}_k, \mathbf{u}_k) + \mathbb{E}_{\mathbf{w}_k, \mathbf{v}_k}[V^*(\mathbf{S}_{k+1})|\mathbf{S}_k, \mathbf{u}_k]], \quad (14)$$

where

- a) $\theta^* = J^{\Omega_u^*} = \inf_{\Omega_u} J^{\Omega_u}$ is the optimal average cost in Problem 2;
- b) $V^*(\mathbf{S}_k)$ is the optimal value function of the extended state $\mathbf{S}_k$.

Proof: Please see Appendix B. ■

There are various standard techniques, such as value iteration [32] or Q-learning [53], that can be used to solve the Bellman equation (14). However, there is a challenge to solving (14) due to the curse of dimensionality in the extended state space $\mathcal{S}$. Specifically, the total dimensions of the extended state space are $S + N_r \times N_t \times 2$, which can be huge when N_r and N_t are large. If we adopt the standard Q-learning approach, which is model-free [53], the domain of the Q-function to be learned has $\frac{(S+N_t+N_t\times N_r\times 2)\times(S+N_t+N_t\times N_r\times 2+1)}{2}$ variables, which is huge. As a result, the learning process will take a very long time. If we learn the value function [32], the domain of the value function involves $\frac{(S+N_t\times N_r\times 2)\times(S+N_t\times N_r\times 2+1)}{2}$ variables, which is also huge.

Challenge 2: Huge Dimension of Variables Involved in the Value Function $V^*(\mathbf{S}_k)$ or the Q-function $Q(\mathbf{S}_k, \mathbf{u}_k)$.

To address Challenge 2, we exploit the special structure in the transition kernel (9) and propose an equivalent *reduced-state Bellman optimality equation*.

Theorem 3: (Reduced-State Bellman Optimality Equation) Under the condition (13) in Theorem 1, the optimal solution

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

to Problem 2 is equivalent to the solution of the equivalent reduced-state Bellman optimality equation given by

$$\widetilde{\theta}^* + \widetilde{V}^*(\mathbf{x}_k) = \mathbb{E}_{\delta_k \mathbf{H}_k}[\min_{\mathbf{u}_k}[r(\mathbf{S}_k, \mathbf{u}_k) + \mathbb{E}_{\mathbf{w}_k, \mathbf{v}_k}[V^*(\mathbf{S}_{k+1})|\mathbf{S}_k, \mathbf{u}_k]], \tag{15}$$

where

- a) $\widetilde{V}^*(\mathbf{x}_k) = \mathbb{E}[V^*(\mathbf{x}_k, \delta_k \mathbf{H}_k)|\mathbf{x}_k] = \mathbf{x}_k^T \mathbf{P} \mathbf{x}_k$ is the reduced-state value function and $\mathbf{P} \in \mathbb{S}_+^S$ is a positive definite matrix;
- b) The optimal average cost $\widetilde{\theta}^* = \theta^* = J^{\Omega_u^*} = \inf_{\Omega_u} J^{\Omega_u} = \mathrm{Tr}(\mathbf{M} + \mathbf{PW} + \mathbf{B}^T \mathbf{PB})$;
- c) The optimal control policy $\Omega_u^* = \{\mathbf{u}_k^*, \forall k\}$, where $\mathbf{u}_k^*$ is the solution to (15) and Problem 2. Furthermore, the optimal control solution of the IoT controller $\mathbf{u}_k^*$ has a linear state-feedback form given by

$$\mathbf{u}_k^* = -(\mathbf{R} + \mathbf{H}_k^T (\mathbf{B}^T \mathbf{PB} + \mathbf{M}) \mathbf{H}_k)^{-1} \delta_k \mathbf{H}_k^T \mathbf{B}^T \mathbf{PA} \mathbf{x}_k. \tag{16}$$

Proof: Please see Appendix C. ■

Compared to the control policy over static channels in (6), the control rule (16) is adaptive to both the system state $\mathbf{x}_k$ and channel state $\delta_k \mathbf{H}_k$. As such, both the the dynamic urgency of the control and the transmission opportunities induced by the fading channels can be captured by the control rule (16). In order to obtain the optimal control solution of the IoT controller $\mathbf{u}_k^*$ in (16), the reduced-state value function $\widetilde{V}^*(\mathbf{x}_k)$ should be learned. Note that the reduced-state value function $\widetilde{V}^*(\mathbf{x}_k)$ is a function of the system state $\mathbf{x}_k$ only. The number of variables in the domain of $\widetilde{V}^*(\mathbf{x}_k)$ is reduced to $\frac{S(S+1)}{2}$. As such, learning the reduced-state value function $\widetilde{V}^*(\mathbf{x}_k)$ would be much easier compared to learning the original value function. Specifically, using the structure properties of the reduced-state value function $\widetilde{V}^*(\mathbf{x}_k)$, the optimal average cost $\widetilde{\theta}^*$ as well as the optimal control solution of the IoT controller $\mathbf{u}_k^*$ in Theorem 3, the Bellman optimality equation (15) can be written as

$$\mathbf{x}_k^T \mathbf{P} \mathbf{x}_k = \mathbf{x}_k^T (\mathbb{E}_{\delta_k \mathbf{H}_k}[\mathbf{Q} + \mathbf{A}^T \mathbf{PA} - \delta_k \mathbf{A}^T \mathbf{PB} \mathbf{H}_k (\mathbf{H}_k^T (\mathbf{B}^T \mathbf{PB} + \mathbf{M}) \mathbf{H}_k + \mathbf{R})^{-1} \mathbf{H}_k^T \mathbf{B}^T \mathbf{PA}]) \mathbf{x}_k. \tag{17}$$

Note that Equation (17) is an algebraic equation with an unknown variable $\mathbf{P}$. Thus, we shall utilize stochastic approximation theory to construct an online learning algorithm to learn $\mathbf{P}$ based on the algebraic equation (17). The learned $\mathbf{P}$ can then be applied to obtain the reduced-state value function $\widetilde{V}^*(\mathbf{x}_k)$ and the optimal control solution for the IoT controller $\mathbf{u}_k^*$.

Specifically, we first rewrite (17) into the standard form $f(\mathbf{P}) = 0$, where $f(\mathbf{P})$ is given by

$$f(\mathbf{P}) = \mathbb{E}_{\delta_k \mathbf{H}_k}[\mathbf{Q} + \mathbf{A}^T \mathbf{PA} - \delta_k \mathbf{A}^T \mathbf{PB} \mathbf{H}_k (\mathbf{H}_k^T (\mathbf{B}^T \mathbf{PB} + \mathbf{M}) \mathbf{H}_k + \mathbf{R})^{-1} \mathbf{H}_k^T \mathbf{B}^T \mathbf{PA}] - \mathbf{P}. \tag{18}$$

To obtain the roots of $f(\mathbf{P}) = 0$, we can apply the stochastic approximation algorithm as shown in Algorithm 1.

In Step 2 and Step 3 of Algorithm 1, realization of the channel state $\delta_k \mathbf{H}_k$ will be required. This can be obtained by standard *channel estimation* at the actuator based on the received pilot symbol from the controller and channel feedback to the controller[2].

Note that Algorithm 1 requires knowledge of the system dynamics $\mathbf{A}$. In the following section, we shall extend this to deal with optimal control for a mission-critical IoT system with unknown system dynamics $\mathbf{A}$ in the presence of the wireless MIMO fading channel between the IoT controller and the actuator. Specifically, we propose an efficient online algorithm to simultaneously identify the system dynamics $\mathbf{A}$ and learn the optimal control solution for the IoT controller $\mathbf{u}_k^*$ in (16) over the MIMO fading channel between the IoT controller and the actuator.

Challenge 3: Simultaneous Online Identification and Optimal Control for the Mission-critical IoT System.

Algorithm 1 Online learning Algorithm for the Optimal Control Solution of the IoT Controller for a Mission-critical IoT System over the MIMO Fading Channels.

Initialization: Given a feasible initial value $\mathbf{P}_0 = \mathbf{P}^0 \in \mathbb{S}_+^S$, the initial estimated reduced-state value function is given by

$$V_0(\mathbf{x}_0) = \mathbf{x}_0^T \mathbf{P}_0 \mathbf{x}_0, \tag{19}$$

and the estimated optimal control solution at the initial timeslot is given by

$$\mathbf{u}_0 = -(\mathbf{R} + \mathbf{H}_0^T (\mathbf{B}^T \mathbf{P}_0 \mathbf{B} + \mathbf{M}) \mathbf{H}_0)^{-1} \delta_0 \mathbf{H}_0^T \mathbf{B}^T \mathbf{P}_0 \mathbf{A} \mathbf{x}_0. \tag{20}$$

For $k = 1, 2, 3, ...$

Step 1: (Update reduced-state value function) Using $\mathbf{P}_k$ updated at the $(k-1)$-th timeslot, the estimated reduced-state value function at the k-th timeslot is given by

$$V_k(\mathbf{x}_k) = \mathbf{x}_k^T \mathbf{P}_k \mathbf{x}_k. \tag{21}$$

Step 2: (Update the optimal control solution) Using $\mathbf{P}_k$ updated at the $(k-1)$-th timeslot, the estimated optimal control solution at the k-th timeslot is given by

$$\mathbf{u}_k = -(\mathbf{R} + \mathbf{H}_k^T (\mathbf{B}^T \mathbf{P}_k \mathbf{B} + \mathbf{M}) \mathbf{H}_k)^{-1} \delta_k \mathbf{H}_k^T \mathbf{B}^T \mathbf{P}_k \mathbf{A} \mathbf{x}_k. \tag{22}$$

Step 3: (Update the unknown value $\mathbf{P}$ in (17)) $\mathbf{P}_{k+1}$ is updated using $\mathbf{P}_k$ and $\mathbf{H}_k$ given by

$$\mathbf{P}_{k+1} = \mathbf{P}_k + \alpha_k (\mathbf{Q} + \mathbf{A}^T \mathbf{P}_k \mathbf{A} - \delta_k \mathbf{A}^T \mathbf{P}_k \mathbf{B} \mathbf{H}_k (\mathbf{H}_k^T (\mathbf{B}^T \mathbf{P}_k \mathbf{B} + \mathbf{M}) \mathbf{H}_k + \mathbf{R})^{-1} \mathbf{H}_k^T \mathbf{B}^T \mathbf{P}_k \mathbf{A} - \mathbf{P}_k), \tag{23}$$

where $\alpha_k > 0$ is the learning stepsize at k-th timeslot.

End

[2]In the existing LTE framework, the actuator transmits the pilot symbol $\mathbf{T} \in \mathbb{R}^{N_t \times N_t}$ to the IoT controller at each timeslot, and the received pilot signal at the IoT controller $\mathbf{y}_k^p = \delta_k \mathbf{H}_k \mathbf{T} + \mathbf{v}_k^p$ can be utilized to obtain the realizations of the channel state $\delta_k \mathbf{H}_k$ via the least square approach.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

IV. Online Identification and Optimal Control for the Mission-critical IoT System over the MIMO Fading Channel with Unknown System Dynamics

In this section, we extend the online optimal control framework in Section III to deal with the unknown mission-critical IoT system in the sense that the system dynamics as well as the optimal control solutions for the IoT controller can be learned simultaneously in an online manner. Specifically, we first formulate the online identification problem for the system dynamics of the mission-critical IoT system. After that, we propose a novel online learning algorithm to simultaneously identify the system dynamics and learn the optimal control solutions for the mission-critical IoT system.

A. Problem Formulation for System Identification of the Mission-critical IoT System over the MIMO Fading Channel

The online system identification of the mission-critical IoT system over the wireless MIMO fading channel between the IoT controller and the actuator can be formulated as an optimization problem as follows.

Problem 3: (Identification of System Dynamics $\mathbf{A}$*)*

$$\min_{\widehat{\mathbf{A}}} \quad \mathbb{E}\left[\left\|\mathbf{x}_{k+1}-\mathbf{A}\mathbf{x}_k-\delta_k\mathbf{B}\mathbf{H}_k\mathbf{u}_k\right\|^2\right], \forall k\geq 1. \tag{24}$$

Note that $\widehat{\mathbf{A}} = \mathbf{A}$ is obviously the global optimal solution of (24) because the objective function $\mathbb{E}\left[\left\|\mathbf{x}_{k+1}-\mathbf{A}\mathbf{x}_k-\delta_k\mathbf{B}\mathbf{H}_k\mathbf{u}_k\right\|^2\right]$ achieves its minimum value $\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^T+\mathbf{W}\right)$ if and only if $\widehat{\mathbf{A}}=\mathbf{A}$. Since the objective function in (24) is a convex function, one may consider utilizing the stochastic gradient descent (SGD) algorithm to obtain the solution to Problem 3 [15], [16], as summarized in the following Algorithm 2.

Algorithm 2 Online Identification for System Dynamics via the Stochastic Gradient Decent Algorithm [15].

Initialization: Set the initial value of $\widehat{\mathbf{A}}$ as $\widehat{\mathbf{A}}_0=\mathbf{A}^0$, where $\mathbf{A}^0$ is an $S\times S$ dimensional constant matrix.

Step 1 (Update of the identified A): At the k-th timeslot, $\forall k>0$, the identified system dynamics $\widehat{\mathbf{A}}_k$ is updated using the system state $\mathbf{x}_k,\mathbf{x}_{k-1}$, control action of the IoT controller $\mathbf{u}_{k-1}$, and the channel state $\delta_{k-1}\mathbf{H}_{k-1}$ as follows:

$$\widehat{\mathbf{A}}_k=\widehat{\mathbf{A}}_{k-1}+\alpha_k(\mathbf{x}_k-\widehat{\mathbf{A}}_{k-1}\mathbf{x}_{k-1}-\delta_{k-1}\mathbf{B}\mathbf{H}_{k-1}\mathbf{u}_{k-1})\mathbf{x}_{k-1}^T, \tag{25}$$

where $\alpha_k>0$ is the learning stepsize at the k-th timeslot.

Step 2 (Termination): If $\|\widehat{\mathbf{A}}_k-\mathbf{A}\|^2<\epsilon$, where $\epsilon>0$ is an arbitrary small value, then obtain the identified system dynamics $\widehat{\mathbf{A}}_k$. Otherwise, go to Step 1.

Note that the convergence of the SGD-based system identification algorithm, Algorithm 2, depends heavily on the statistical boundness of the system state $\mathbf{x}_k$, as summarized in the following Theorem 4.

Theorem 4: (Convergence Conditions of Algorithm 2) If the following two conditions are satisfied:

- *Stepsize Condition*: The stepsize sequence $\{\alpha_k, k\geq 0\}$ obeys

$$\sum_{k=1}^{\infty}\alpha_k=\infty, \sum_{k=1}^{\infty}\alpha_k^2<\infty, \tag{26}$$

- *Bounded Conditional Variance*: There exist two bounded constants $\eta>0$ and $\mu>0$ such that the conditional variance $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]$ is bounded and satisfies

$$\eta\mathbf{I}_S<\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]<\mu\mathbf{I}_S, \forall k\geq 1, \tag{27}$$

then $\widehat{\mathbf{A}}_k$ in Algorithm 2 converges to the true system dynamics $\mathbf{A}$ almost surely.

Proof: Please see Appendix D. ■

Note that when $\widehat{\mathbf{A}}^*=\mathbf{A}$ is stable, $\widehat{\mathbf{A}}_k$ obtained by the SGD update in (25) can converge to $\widehat{\mathbf{A}}^*$ for arbitrary bounded control sequence $\{\mathbf{u}_k, k\geq 0\}$ because the closed-loop system is mean-square stable, which implies the conditional variance $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]$ is also bounded. As such, online system identification can be considered separately from the control as in [16]-[17]. However, when $\widehat{\mathbf{A}}^*$ is unstable, the control sequence $\{\mathbf{u}_k, k\geq 0\}$ cannot be an arbitrary bounded control sequence and it plays a critical role in maintaining the conditional variance $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]$ to be bounded in the identification process. Moreover, the design of control $\mathbf{u}_k$ and $\widehat{\mathbf{A}}_k$ will be coupled. Brute-force application of $\mathbf{u}_k$ in (22) in Algorithm 1 cannot achieve bounded $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]$ because $\mathbf{u}_k$ in (22) is not a stabilizing control solution when $\widehat{\mathbf{A}}_k$ deviates from $\widehat{\mathbf{A}}^*$, which will result in closed-loop instability.

> Challenge 4: Simultaneous Learning of System Dynamics and Optimal Control for the Mission-critical IoT System with an Unstable Dynamic Plant.

In order to address above challenge, in the following, we modify the SGD algorithm in (25) and propose a novel normalized stochastic gradient descent algorithm (NSGD) to simultaneously learn the system dynamics $\mathbf{A}$ and the optimal control $\mathbf{u}_k^*$ for the mission-critical IoT system in an online manner.

B. Simultaneous Learning of the System Dynamics and the Optimal Control Solution for the Mission-critical IoT System over the MIMO Fading Channel

Since the divergence of Algorithm 2 is due to the unbounded conditional variance of the system state $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]$, we propose a novel normalization method to solve Challenge 4, where the learning stepsize is dynamically normalized based on the realization of the system state at each timeslot. We summarize the proposed normalized stochastic gradient descent algorithm in the following Algorithm 3.

Compared to Algorithm 2, the conditional variance of the system state $\mathbf{x}_k$ is changed from $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]$ to

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

Algorithm 3 Online System Identification and Optimal Control for the System Dynamics via the **Normalized Stochastic Gradient Decent Algorithm**.

Initialization: Set the initial value of $\widehat{\mathbf{A}}$ as $\widehat{\mathbf{A}}_0 = \mathbf{A}^0$, where $\mathbf{A}^0$ is an $S \times S$ dimensional constant matrix.

Step 1 (Update of the learned A): At the k-th time slot, $\forall k > 0$, the learned system dynamics $\widehat{\mathbf{A}}_k$ is obtained using the system states $\mathbf{x}_k, \mathbf{x}_{k-1}$, control action $\mathbf{u}_{k-1}$ and the channel state $\delta_{k-1}\mathbf{H}_{k-1}$ as follows:

$$\widehat{\mathbf{A}}_k = \widehat{\mathbf{A}}_{k-1} + \widehat{\alpha}_k \Big(\mathbf{x}_k - \widehat{\mathbf{A}}_{k-1}\mathbf{x}_{k-1} - \delta_{k-1}\mathbf{B}\mathbf{H}_{k-1} \mathbf{u}_{k-1}\Big)\mathbf{x}_{k-1}^T, \tag{28}$$

where $\widehat{\alpha}_k > 0$ is the normalized learning stepsize, given by

$$\widehat{\alpha}_k = \begin{cases} \alpha_k, & if \quad \|\mathbf{x}_{k-1}\|^2 < 1; \\ \frac{\alpha_k}{\|\mathbf{x}_{k-1}\|^2}, & otherwise, \end{cases} \tag{29}$$

and $\{\alpha_k, k \geq 0\}$ is the stepsize sequence.

Step 2 (Update of the control $\mathbf{u}_k$): At the k-th time slot, $\forall k > 0$, the control action $\mathbf{u}_k$ is updated using plant state $\mathbf{x}_k$, $\mathbf{P}_k$ and the channel state $\delta_k\mathbf{H}_k$ according to (22), and $\mathbf{P}_k$ is updated as

$$\mathbf{P}_{k+1} = \mathbf{P}_k + \alpha_k(\widehat{\mathbf{A}}_k^T\mathbf{P}_k\widehat{\mathbf{A}}_k - \delta_k\widehat{\mathbf{A}}_k^T\mathbf{P}_k\mathbf{B}\mathbf{H}_k(\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k\mathbf{B}\mathbf{H}_k + \mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{R})^{-1}\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k\widehat{\mathbf{A}}_k - \mathbf{P}_k + \mathbf{Q}). \tag{30}$$

Step 3 (Termination): If $\|\widehat{\mathbf{A}}_k - \mathbf{A}\|^2 + \|\mathbf{P}_{k+1} - \mathbf{P}\|^2 < \epsilon$, where $\epsilon > 0$ is an arbitrary small value, then obtain the identified system dynamics $\widehat{\mathbf{A}}_k$ and the unknown $\mathbf{P}_{k+1}$. The optimal control solution for the IoT controller $\mathbf{u}_k$ is given by (22). Otherwise, go to Step 1.

$\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\mathbf{1}_{\{\|\mathbf{x}_k\|^2<1\}} + \frac{\mathbf{x}_k\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2}\mathbf{1}_{\{\|\mathbf{x}_k\|^2\geq 1\}}\middle|\widehat{\mathbf{A}}_k\right]$, which is upper bounded by $\mathbf{I}_S$ due to the normalization factor $\|\mathbf{x}_k\|^2$ in the normalized stepsize $\widehat{\alpha}_k$ (29). As a result, the convergence of the proposed online learning algorithm, Algorithm 3, can be guaranteed.

C. Convergence Analysis

Since the learned control solution for the IoT controller $\mathbf{u}_k$ in Step 2 of the proposed online learning Algorithm 3 is obtained based on the successive update of $\widehat{\mathbf{A}}_k$ and $\mathbf{P}_k$, the convergence analysis for the learned control solution $\mathbf{u}_k$ can be obtained by analyzing the convergence of $\widehat{\mathbf{A}}_k$ and $\mathbf{P}_k$. The convergence of the online system identification in Step 1 of Algorithm 3 is summarized below.

Theorem 5: (Convergence of Online System Identification) If the learning stepsize sequence $\{\alpha_k, k \geq 1\}$ satisfies Condition (26), then $\widehat{\mathbf{A}}_k$ obtained by (28) in Step 1 of Algorithm 3 converges to the true system dynamics $\mathbf{A}$ almost surely, i.e.,

$$\Pr\left(\lim_{k\to\infty}\widehat{\mathbf{A}}_k = \mathbf{A}\right) = 1. \tag{31}$$

Proof: Please see Appendix E. ■

Due to the convergence of $\widehat{\mathbf{A}}_k$ to $\mathbf{A}$, if $\mathbf{P}_k$ converges to $\mathbf{P}$, then the limiting convergent point $\mathbf{P}$ must be the root of $f(\mathbf{P}) = \mathbf{0}$. Furthermore, if $\mathbf{P}_k$ in (30) converges, $\mathbf{u}_k$ in (22) will also converge to the optimal control action $\mathbf{u}_k^*$ in (16). The convergence results of $\mathbf{P}_k$ in (30) and $\mathbf{u}_k$ in (22) are formally summarized in the following theorem.

Theorem 6: (Almost Sure Convergence of Online Learning of the Value Function and Control Action) Let $\mathbf{P}$ be the unique root of $f(\mathbf{P}) = \mathbf{0}$. If the sufficient condition (13) in Theorem 1 is satisfied and the stepsize sequence $\{\alpha_k, k > 0\}$ satisfies the condition (26), then

- *Convergence of* $\mathbf{P}_k$: $\mathbf{P}_k$ obtained by (30) in Step 2 of the proposed online learning Algorithm 3 converges to $\mathbf{P}$ almost surely, i.e., $\Pr\left(\lim_{k\to\infty}\mathbf{P}_k = \mathbf{P}\right) = 1$.
- *Convergence of the Value Function and Control Action*: The learned value function $\widetilde{V}_k(\mathbf{x}_k)$ in (21) converges to the optimal value function $\widetilde{V}(\mathbf{x}_k) = \mathbf{x}_k^T\mathbf{P}\mathbf{x}_k$ in Theorem 3 almost surely, i.e.,

$$\Pr\left(\lim_{k\to\infty}\widetilde{V}_k(\mathbf{x}_k) = \widetilde{V}(\mathbf{x}_k)\right) = 1. \tag{32}$$

Moreover, the learned control action $\mathbf{u}_k$ in Step 2 of the proposed online learning Algorithm 3 converges to the optimal control action $\mathbf{u}_k^* = \Omega_u^*(\mathbf{S}_k)$ in Theorem 3 almost surely, i.e.,

$$\Pr\left(\lim_{k\to\infty}\mathbf{u}_k = \mathbf{u}_k^*\right) = 1. \tag{33}$$

Proof: Please see Appendix F. ■

The convergence results in Theorem 6 reveal the fact that less unstable dynamic plant (i.e., smaller $\|\mathbf{A}\|$) and more communication resources (i.e., smaller $\left\|\mathbb{E}\left[\mathbf{A}^T\mathbf{V}_k^T(\mathbf{I}_S - \mathbf{\Pi}_k)\mathbf{V}_k\mathbf{A}\right]\right\|$) are more favorable for the convergence of the proposed online optimal control algorithm, Algorithm 3.

V. Numerical results

In this section, we shall verify the performance advantages of the proposed learning algorithms for the optimal control solution of the IoT controller as well as the system dynamics for the mission-critical IoT system. Specifically, we compare the proposed scheme with various baselines.

- **Baseline 1** *(Known System Dynamics and Known Optimal Control Solution [54]):* The system dynamics $\mathbf{A}$ and the optimal control solution for the mission-critical IoT system over the MIMO fading channels in (16) are known. Based on the received system state $\mathbf{x}_k$ at each k-th timeslot, the IoT controller generates the optimal control signal via (16).
- **Baseline 2** *(Q-learning-based Control with Known System Dynamics [49]):* The system dynamics $\mathbf{A}$ is known. At each k-th timeslot, the dynamic plant transmits the system state $\mathbf{x}_k$ to the IoT controller. The IoT controller generates the control signal $\mathbf{u}_k$ via solving the unknown kernel value of the Q function with the knowledge of the system dynamics $\mathbf{A}$.
- **Baseline 3** *(Q-learning-based Control with Identified System Dynamics via the Least-square-based Approach*

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

[16]): At each k-th timeslot, the dynamic plant transmits the system state $\mathbf{x}_k$ to the IoT controller. The system dynamics $\mathbf{A}$ is identified via the least-square-based approach using the system state trajectory $\{\mathbf{x}_0, \mathbf{x}_1, ...\mathbf{x}_k\}$. The IoT controller generates the control signal $\mathbf{u}_k$ via solving the unknown kernel value of the Q function based on the identified system dynamics $\widehat{\mathbf{A}}$ for $\mathbf{A}$.

- **Baseline 4** *(Q-learning-based Control with Identified System Dynamics via the SGD-based Approach [55]):* At each k-th timeslot, the dynamic plant transmits the system state $\mathbf{x}_k$ to the IoT controller. The system dynamics $\mathbf{A}$ is learned via the SGD-based algorithm using the system state trajectory $\{\mathbf{x}_0, \mathbf{x}_1, ...\mathbf{x}_k\}$, as shown in Algorithm 2. The IoT controller generates the control signal $\mathbf{u}_k$ via solving the unknown kernel value of the Q function based on the identified system dynamics $\widehat{\mathbf{A}}$ for $\mathbf{A}$.
- **Baseline 5** *(Potential-learning-based Control with Known System Dynamics [32]):* The system dynamics $\mathbf{A}$ is known. At each k-th timeslot, the dynamic plant transmits the system state $\mathbf{x}_k$ to the IoT controller. The IoT controller generates the control signal $\mathbf{u}_k$ via solving the unknown kernel value of the value function with the knowledge of the system dynamics $\mathbf{A}$.

We consider an end-to-end mission-critical IoT system in the presence of a MIMO fading channel between the IoT controller and the actuator of the dynamic plant. Specifically, we consider both the stable and unstable dynamic plant with system dynamics given by $\mathbf{A}_1 = \begin{bmatrix} 0.0470 & 0.0172 & -0.0054 & 0.0019 \\ 0.0117 & 0.0267 & -0.0004 & 0.0066 \\ -0.0132 & 0.0096 & 0.0021 & -0.0113 \\ -0.0002 & 0.0022 & -0.0151 & 0.0099 \end{bmatrix}$ and $\mathbf{A}_2 = \begin{bmatrix} 0.9970 & 0.0172 & -0.0054 & 0.0019 \\ 0.0117 & 1.0167 & -0.0004 & 0.0066 \\ -0.0132 & 0.0096 & 0.9920 & -0.0113 \\ -0.0002 & 0.0022 & -0.0151 & 0.9941 \end{bmatrix}$, respectively. The control input matrix $\mathbf{B} = \begin{bmatrix} 0.3661 & 0.3920 \\ -0.3717 & 0.5731 \\ -1.2367 & 0.2821 \\ -1.2242 & 0.4940 \end{bmatrix}$. The system noise variance for all the simulation results is $\mathbf{W} = \mathbf{I}_4 \in \mathbb{S}_+^4$.

A. CPU Computational Time Analysis

The CPU computational time versus the dimension of system state S, the number of transmission antennas N_t and the number of receiving antennas N_r is illustrated in Figure 3, 4 and 5, respectively. As shown in the figures, the CPU computational time for 10^4 simulation runs of the proposed scheme is substantially less than that of Baseline 2-5. This is because the Q-learning-based control algorithm in Baseline 2-4 involves computation of the Q-function with $\frac{(S+N_t+N_t\times N_r\times 2)\times(S+N_t+N_t\times N_r\times 2+1)}{2}$ unknown variables. The potential-learning-based control algorithm in Baseline 5 involves computation of the value function with $\frac{(S+N_t\times N_r\times 2)\times(S+N_t\times N_r\times 2+1)}{2}$ unknown variables. Differently, the proposed scheme only requires computation for the reduced-state value function with $\frac{S(S+1)}{2}$ unknown variables, which are strictly smaller then those in each of Baseline 2-5, and hence the computational complexity can be significantly reduced. We also observe that Baseline 1 has

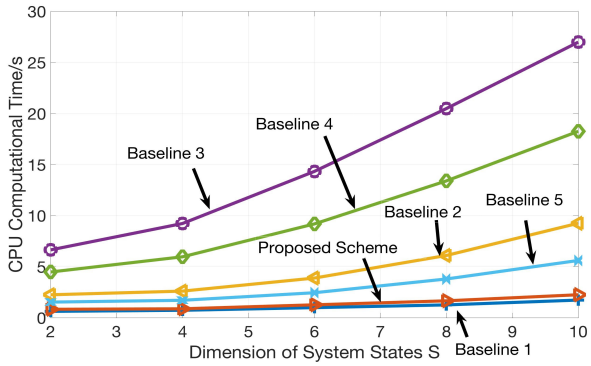


Fig. 3: CPU computational time versus the dimension of system states for 10^4 iterations. The number of transmission antennas $N_t = 4$ and the number of receiving antennas $N_r = 4$. The dynamic plant $\mathbf{A} \in \mathbb{R}^{S\times S}$ and the control input matrix $\mathbf{B} \in \mathbb{R}^{S\times 4}$ are randomly generated. Specifically, each element in matrix $\mathbf{A}$ and $\mathbf{B}$ are i.i.d. generated following a Gaussian distribution with zero mean and unit variance.

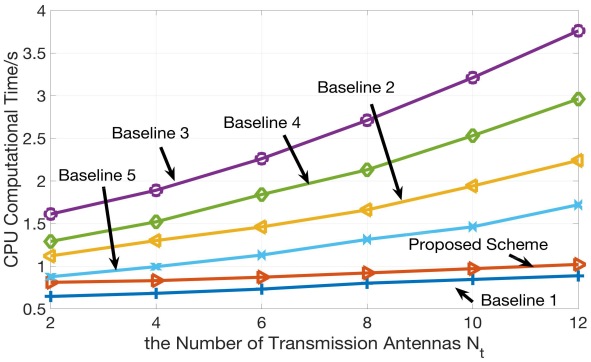


Fig. 4: CPU computational time versus the number of transmission antennas for 10^4 iterations. The number of system states $S = 4$ and the number of the receiving antennas $N_r = 2$. The system dynamics is given by the unstable matrix $\mathbf{A}_2$.

optimal performance in terms of computational complexity since the system dynamics and the optimal control solution are known. However, as shown in Figure 3-5, the CPU computational time of the proposed scheme is similar to that of Baseline 1. As such, the proposed online identification and control algorithm for the mission-critical IoT system has low computational complexity.

B. Convergence Analysis

The MSE between the learned control solution and the optimal control solution versus iteration number under a stable and an unstable system is illustrated in Fig. 6 and Fig. 7, respectively. As shown in the figures, when the system is stable, all the schemes achieve a good MSE over time. However, when the system is unstable, the learned control solutions deviate from the optimal one in Baseline 2-5, while our proposed scheme tracks the optimal one asymptotically. Specifically, the control solutions obtained via the Q-learning-based control algorithm in Baseline 2-4 and the potential-learning-based control algorithm in Baseline 5 deviate from the optimal control solution due to the "curse of dimensionality" issue during the learning process for the Q-function and

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet of Things Journal

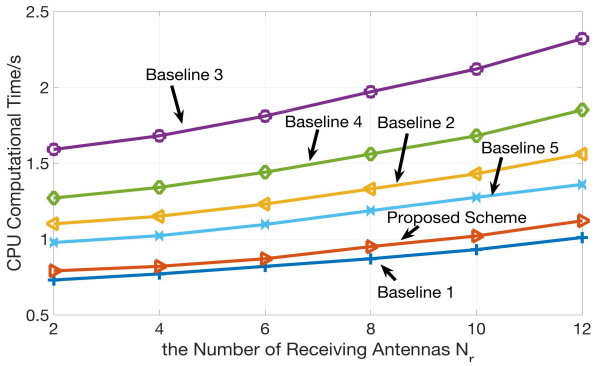


Fig. 5: CPU computational time versus the number of receiving antennas for 10^4 iterations. The number of system states $S = 4$ and the number of transmission antennas $N_t = 1$. The system dynamics are given by the unstable matrix $\mathbf{A}_2$. The control input matrix $\mathbf{B} \in \mathbb{R}^{4\times N_r}$ is randomly generated. Specifically, each element of the matrix $\mathbf{B}$ is randomly generated following a Gaussian distribution with zero mean and unit variance.

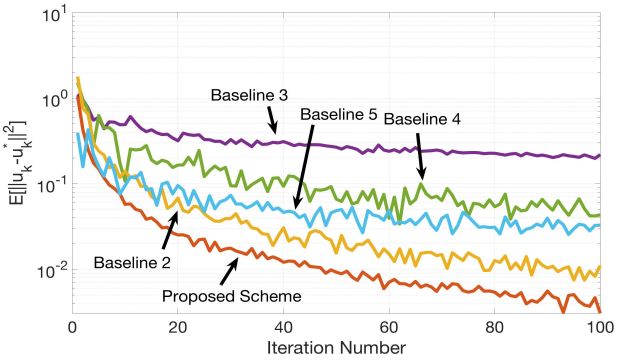


Fig. 6: MSE between the learned control solution and optimal control solution versus iteration number under a stable system. The number of system states $S = 4$, the number of transmission antennas $N_t = 1$ and the number of receiving antennas $N_r = 2$. The system dynamics are given by the stable matrix $\mathbf{A}_1$.

potential function with extended large state space. When the system is unstable, the mismatch between the learned control solution and the optimal control solution in Baseline 2-5 will be enlarged. Our proposed control algorithm, however, only learns the reduced-state value function with a small state space without the "curse of dimensionality" issue and hence it can achieve the optimal control solution.

The MSE between the learned system dynamics and the true system dynamics versus iteration number under a stable and an unstable system is illustrated in Fig. 8 and Fig. 9, respectively. As shown in figures, when the system is stable, all the schemes achieve a good MSE over time. However, when the system is unstable, the learned system dynamics of Baseline 3 and Baseline 4 cannot converge to the true system dynamics. On the other hand, the learned system dynamics via the proposed scheme will asymptotically converge to the true system dynamics. Specifically, the least-square-based identification algorithm in Baseline 3 suffers from serious numerical error when learning the system dynamics since the noisy system states collected by the IoT controller will jeopardize the performance. The SGD-based identification algorithm in Baseline 4 cannot learn the true system dynamics since the stability assumption for internal dynamics $\|\mathbf{A}\|$ is not satisfied and the boundness of the increment in the SGD update is not guaranteed. Our proposed identification scheme, however, can track the true system dynamics even in the presence of an unstable dynamic plant and wireless network due to the normalization operator in the SGD update.

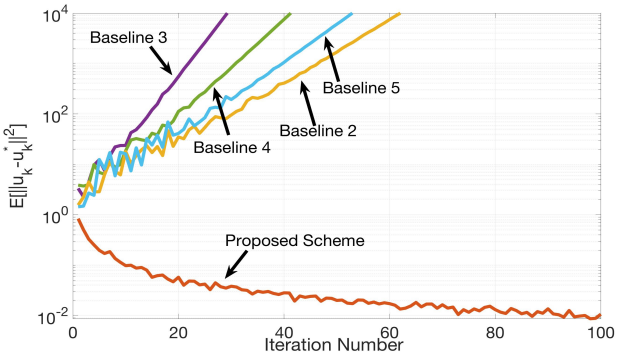


Fig. 7: MSE between the learned control solution and optimal control solution versus iteration number under an unstable system. The number of system states $S = 4$, the number of transmission antennas $N_t = 1$ and the number of receiving antennas $N_r = 2$. The system dynamics are given by the unstable matrix $\mathbf{A}_2$.

C. Stability Analysis

The sample path of the averaged $l2$-norm of the system state is illustrated in Fig. 10. As shown in the figure, the $l2$-norm of the system state of Baseline 2-5 increases over time. This is because Baseline 2-5 cannot learn the optimal control solution due to the "curse of dimensionality" issue and hence the system is unstable. Our proposed scheme, however, can stabilize the system and achieves a similar performance to that in Baseline 1. This is because the proposed control scheme tracks the optimal control solution and stabilizes the system.

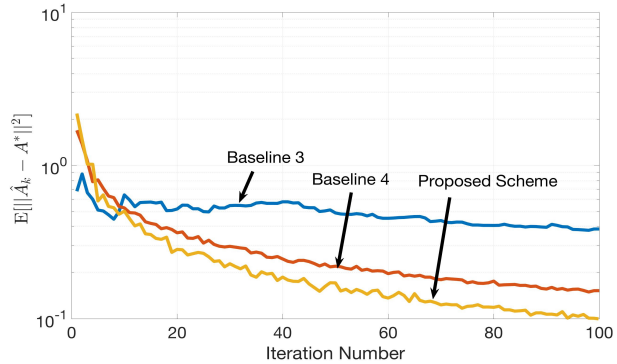


Fig. 8: MSE between the learned system dynamics and the true system dynamics versus iteration number under a stable system. The number of system states $S = 4$, the number of transmission antennas $N_t = 1$ and the number of receiving antennas $N_r = 2$. The system dynamics are given by the stable matrix $\mathbf{A}_1$.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

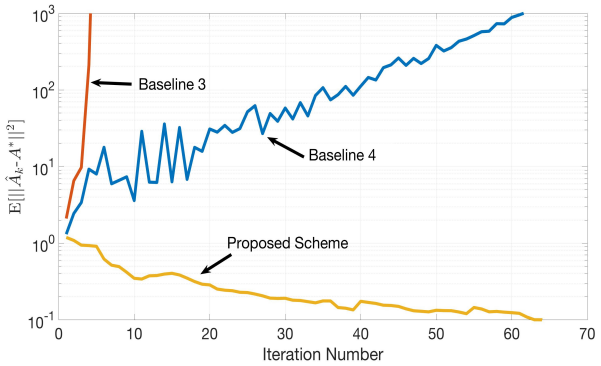


Fig. 9: MSE between the learned system dynamics and the true system dynamics versus iteration number under an unstable system. The number of system states $S = 4$, the number of transmission antennas $N_t = 1$ and the number of receiving antennas $N_r = 2$. The system dynamics are given by the unstable matrix $\mathbf{A}_2$.

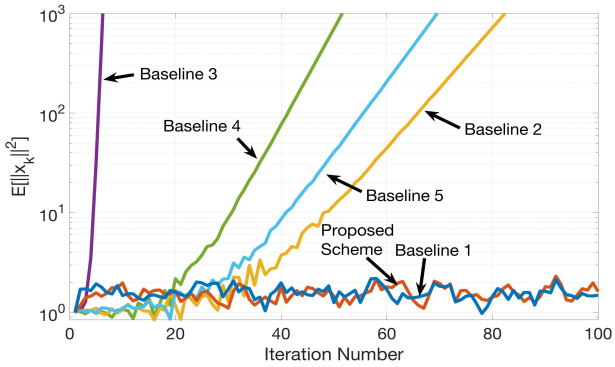


Fig. 10: Average $l2$-norm of the system state versus iteration number. The number of system states $S = 4$, the number of transmission antennas $N_t = 1$ and the number of receiving antennas $N_r = 2$. The system dynamics are given by the unstable matrix $\mathbf{A}_2$.

VI. Conclusion And Future Works

In this paper, we considered the online identification and optimal control problem for a mission-critical IoT system in the presence of wireless MIMO fading channels between the IoT controller and the actuator. Specifically, we first theoretically analyzed the existing condition for the optimal control solution of the mission-critical IoT system in the presence of the MIMO channels between the IoT controller and the actuator via the PSD decomposition technique. After that, we proposed a novel stochastic-approximation-based algorithm to learn the optimal control solution for the IoT controller in an online manner with known system dynamics. We further extended the optimal control framework to deal with unknown system dynamics, and proposed a novel normalized-SGD based identification and control algorithm that can learn the system dynamics and the optimal control solution simultaneously in an online manner. Numerical simulations demonstrate the superiority of our proposed algorithms compared to the existing approaches.

Our proposed scheme is of practical use. Notice that in the practical scenarios, many mission-critical IoT systems such as vehicle platooning IoT systems and water tank IoT systems, can be modelled as linear dynamic systems over the wireless network. Our proposed online identification and optimal control approach can be applied to identify and control such systems purely based on the real-time plant system state information and the channel state information without prior knowledge of the system dynamics of the systems. Moreover, our work can be extended to solve new practical problems. (1) Notice that the proposed online learning scheme for the optimal control solution is obtained by the state-reduction technique due to the i.i.d. properties of the fading channels across timeslots. However, when the wireless fading realizations are correlated among timeslots, the proposed approach cannot be brute-forcely applied and it is desirable to develop new methods based on the current work to learn the optimal control solution in an online manner; (2) Notice that the IoT system considered in our work follows linear dynamics. When a nonlinear system is considered, brute-force applications of our proposed system identification method will lead to model mismatch. Since the control policy applied in our work is based on the identified system dynamics, the system will be unstable under the learned control policy. As a result, it is desirable to develop new methods based on the current work to identify the nonlinear system dynamics and learn the optimal control solution in an online manner.

Appendix

A. Proof of Theorem 1

Problem 2 can be solved via MDP techniques. Specifically, Problem 2 has a solution if there exists a pair of $\left(\theta, \widetilde{V}\left(\mathbf{x}_k\right)\right)$ such that the following optimality equation is satisfied:

$$\widetilde{\theta} + \widetilde{V}\left(\mathbf{x}_k\right) = \mathbb{E}\Bigg[\min_{\mathbf{u}\left(\mathbf{x}_k, \mathbf{H}_k\right)} \Bigg[r\left(\mathbf{x}_k, \delta_k\mathbf{H}_k, \mathbf{u}\left(\mathbf{x}_k, \delta_k\mathbf{H}_k\right)\right) + \sum_{\mathbf{x}_{k+1}} \Pr\left[\mathbf{x}_{k+1} \middle| \mathbf{x}_k, \delta_k\mathbf{H}_k, \mathbf{u}\left(\mathbf{x}_k, \delta_k\mathbf{H}_k\right)\right] \widetilde{V}\left(\mathbf{x}_{k+1}\right)\Bigg], \forall \mathbf{x}_k. \tag{34}$$

To solve the optimality equation (34), we first assume that the reduced-state value function $\widetilde{V}\left(\mathbf{x}_k\right)$ has a quadratic form of $\mathbf{x}_k$ given by $\widetilde{V}\left(\mathbf{x}_k\right) = \mathbf{x}_k^T\mathbf{P}\mathbf{x}_k$ with $\mathbf{P} \in \mathbb{S}_+^S$ being a constant positive definite matrix. Then, Equation (34) can be represented as

$$\widetilde{\theta} + \mathbf{x}_k^T\mathbf{P}\mathbf{x}_k = \mathbb{E}\Bigg[\min_{\mathbf{u}_k} \Big[\mathbf{x}_k^T\mathbf{Q}\mathbf{x}_k + \mathbf{u}_k^T\left(\mathbf{R} + \mathbf{H}_k^T\mathbf{M}\mathbf{H}_k\right)\mathbf{u}_k + \mathrm{Tr}\left(\mathbf{M}\right) + \left(\mathbf{A}\mathbf{x}_k + \delta_k\mathbf{B}\mathbf{H}_k\mathbf{u}_k\right)^T\mathbf{P}\left(\mathbf{A}\mathbf{x}_k + \delta_k\mathbf{B}\mathbf{H}_k\mathbf{u}_k\right) + \mathrm{Tr}\left(\mathbf{P}\mathbf{W}\right) + \mathrm{Tr}\left(\mathbf{B}^T\mathbf{P}\mathbf{B}\right)\Big], \tag{35}$$

and $\mathbf{u}_k^*$ that achieves the minimum value of (35) is given by

$$\mathbf{u}_k^* = -\left(\mathbf{R} + \mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{H}_k^T\mathbf{B}^T\mathbf{P}\mathbf{B}\mathbf{H}_k\right)^{-1}\delta_k\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}\mathbf{A}\mathbf{x}_k. \tag{36}$$

Substituting (36) into (35), the optimality equation (35) can be further represented as

$$\widetilde{\theta} + \mathbf{x}_k^T\mathbf{P}\mathbf{x}_k = \mathrm{Tr}\left(\mathbf{M} + \mathbf{P}\mathbf{W} + \mathbf{B}^T\mathbf{P}\mathbf{B}\right) \tag{37}$$

$$+ \mathbf{x}_k^T\mathbb{E}\left[\mathbf{S}\left(\mathbf{P}\right)\right]\mathbf{x}_k, \forall \mathbf{x}_k, \tag{38}$$

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

where $\mathbf{S}(\mathbf{P})$ is a matrix-valued function in terms of $\mathbf{P}$, and is given by

$$\mathbf{S}(\mathbf{P}) = \mathbf{Q} + \mathbf{A}^T\mathbf{P}\mathbf{A} - \delta_k\mathbf{A}^T\mathbf{P}\mathbf{B}\mathbf{H}_k(\mathbf{R} + \mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{H}_k^T\mathbf{B}^T\mathbf{P}\mathbf{B}\mathbf{H}_k)^{-1}\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}\mathbf{A}.$$

Assuming $\widetilde{V}(\mathbf{x}_k)$ exists, i.e., $\mathbf{P}$ exists, it follows that

$$\widetilde{\theta} = \mathrm{Tr}\left(\mathbf{M} + \mathbf{P}\mathbf{W} + \mathbf{B}^T\mathbf{P}\mathbf{B}\right), \tag{39}$$

$$\widetilde{V}(\mathbf{x}_k) = \mathbf{x}_k^T\mathbf{P}\mathbf{x}_k = \mathbf{x}_k^T\mathbb{E}[\mathbf{S}]\mathbf{x}_k, \ \ \forall \mathbf{x}_k. \tag{40}$$

As a result, for any given realization of channel state $\delta_k\mathbf{H}_k$, the solution to Problem 2, i.e., $\widetilde{\theta}$ in (39), $\widetilde{V}(\mathbf{x}_k)$ in (40) and $\mathbf{u}_k^*$ in (36), is uniquely determined by $\mathbf{P}$. Further note that (40) is satisfied for all $\mathbf{x}_k$, it follows that $\mathbf{P}$ satisfies the following matrix equation:

$$\mathbf{P} = \mathbb{E}[\mathbf{S}(\mathbf{P})] = \mathbf{A}^T\mathbf{P}\mathbf{A} - \mathbb{E}[\delta_k\mathbf{A}^T\mathbf{P}\mathbf{B}\mathbf{H}_k(\delta_k\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}\mathbf{B}\mathbf{H}_k + \delta_k\mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{R})^{-1}\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}\mathbf{A}] + \mathbf{Q}. \tag{41}$$

As a result, it suffices to prove that the solution $\mathbf{P}^*$ to the matrix equation (41) exists, then the solutions of $\widetilde{\theta}$, $\widetilde{V}(\mathbf{x}_k)$ and $\mathbf{u}_k^*$ to Problem 2 all exist.

We now prove the existence of $\mathbf{P}^*$ such that $\mathbf{P}^* = \mathbb{E}[\mathbf{S}(\mathbf{P})]$ under the sufficient condition (13). Note that there is a $\mathbf{P}^{(1)} = \mathbf{0}$ such that $\mathbf{P}^{(1)} = \mathbf{0} < g\left(\mathbf{P}^{(1)}\right) = \mathbf{Q}$. Moreover, for any given channel state $\delta_k\mathbf{H}_k$, $\mathbb{E}[\mathbf{S}(\mathbf{P})]$ can be represented as

$$\mathbb{E}[\mathbf{S}(\mathbf{P})] = \mathbf{Q} + \mathbf{A}^T\mathbb{E}[\mathbf{P}_k^{uc}]\mathbf{A} + \mathbb{E}[\mathbf{A}^T\mathbf{P}_k^c\mathbf{A} - \mathbf{A}^T\mathbf{P}_k^c\mathbf{B}\mathbf{H}_k(\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k^c\mathbf{B}\mathbf{H}_k + \mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{R})^{-1}\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k^c\mathbf{A}], \tag{42}$$

where

$$\mathbf{P}_k^c = \mathbf{V}_k^T\begin{bmatrix} \left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T\right)_{\gamma_k} & \left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T\right)_{\gamma_k}\mathbf{\Sigma}_k \\ \mathbf{\Sigma}_k^T\left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T\right)_{\gamma_k} & \mathbf{\Sigma}_k^T\left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T\right)_{\gamma_k}\mathbf{\Sigma}_k \end{bmatrix}\mathbf{V}_k; \tag{43}$$

$$\mathbf{P}_k^{uc} = \mathbf{V}_k^T(\mathbf{I}_S - \mathbf{\Pi}_k)\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T(\mathbf{I}_S - \mathbf{\Pi}_k)\mathbf{V}_k - \mathbf{V}_k^T\mathrm{diag}\left(\mathbf{0}_{S-\gamma_k}, \mathbf{\Sigma}_k^T\left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^H\right)_{\gamma_k}\mathbf{\Sigma}_k\right)\mathbf{V}_k, \tag{44}$$

and

$$\mathbf{\Sigma}_k = \left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T\right)_{\gamma_k}^{-1}\left(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T\right)_{(1:\gamma_k;\gamma_k+1:S)}. \tag{45}$$

Additionally, denote $\widetilde{\zeta}_k = \mathrm{Diag}((\zeta_k)_{\gamma_k}^{\frac{1}{2}}, \mathbf{I}_{S-\gamma_k})$, it follows that

$$\delta_k\mathbf{B}\mathbf{H}_k(\delta_k\mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{R})^{-1}\mathbf{H}_k^T\mathbf{B}^T = \mathbf{V}_k^T\widetilde{\zeta}_k\mathbf{\Pi}_k\widetilde{\zeta}_k\mathbf{V}_k. \tag{46}$$

Let $\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T)_{\gamma_k} = \mathbf{P}_{\gamma_k}$ and $(\zeta_K)_{\gamma_k}^{\frac{1}{2}}(\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T)_{\gamma_k}(\zeta_K)_{\gamma_k}^{\frac{1}{2}} = \widetilde{\mathbf{P}}_{\gamma_k}$. The $\mathbf{P}_k^c$ dependent terms in (17) can be represented as

$$\begin{aligned}
&\mathbf{A}^T\mathbb{E}[\mathbf{P}_k^c\mathbf{A} - \delta_k\mathbf{A}^T\mathbf{P}_k^c\mathbf{B}\mathbf{H}_k(\delta_k\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k^c\mathbf{B}\mathbf{H}_k + \delta_k\mathbf{H}_k^T\\
&\mathbf{M}\mathbf{H}_k + \mathbf{R})^{-1}\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k^c]\mathbf{A}\\
&= \mathbf{A}^T\mathbb{E}[(\mathbf{P}_k^c\Xi_k\Xi_k^T + \mathbf{I}_S)^{-1}\mathbf{P}_k^c]\mathbf{A}\\
&= \mathbf{A}^T\mathbb{E}[\mathbf{V}_k^T(\mathbf{V}_k\mathbf{P}_k^c\mathbf{V}_k^T\zeta_k + \mathbf{I}_S)^{-1}\mathbf{V}_k\mathbf{P}_k^c\mathbf{V}_k^T\mathbf{V}_k]\mathbf{A}\\
&= \mathbf{A}^T\mathbb{E}[\mathbf{V}_k^T\widetilde{\zeta}_k^{-1}(\widetilde{\zeta}_k\mathbf{V}_k\mathbf{P}_k^c\mathbf{V}_k^T\widetilde{\zeta}_k\mathbf{\Pi}_k + \mathbf{I}_S)^{-1}\widetilde{\zeta}_k\\
&\quad \mathbf{V}_k\mathbf{P}_k^c\mathbf{P}_k^c\mathbf{V}_k^T\widetilde{\zeta}_k\widetilde{\zeta}_k^{-1}\mathbf{V}_k]\mathbf{A}\\
&= \mathbf{A}^T\mathbb{E}[\mathbf{V}_k^T\widetilde{\zeta}_k^{-1}]\begin{bmatrix} (\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1} & (\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1}\mathbf{P}_{\gamma_k}\Sigma_k \\ \Sigma_k^T\mathbf{P}_{\gamma_k}(\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1} & \eta_k \end{bmatrix}\\
&\quad \widetilde{\zeta}_k^{-1}\mathbf{V}_k]\mathbf{A},
\end{aligned} \tag{47}$$

where

$$\Xi_k = \delta_k\mathbf{B}\mathbf{H}_k(\delta_k\mathbf{H}_k^T\mathbf{M}\mathbf{H}_k + \mathbf{R})^{-\frac{1}{2}}, \tag{48}$$

$$\eta_k = \Sigma_k^T\mathbf{P}_{\gamma_k}\Sigma_k - \Sigma_k^T\mathbf{P}_{\gamma_k}(\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1}\mathbf{P}_{\gamma_k}\Sigma_k. \tag{49}$$

Substituting (47) into (42), it follows that

$$\begin{aligned}
\mathbb{E}[\mathbf{S}(\mathbf{P})] = \mathbf{Q} + \mathbf{A}^T\mathbb{E}[\mathbf{V}_k^T(\mathbf{I} - \mathbf{\Pi}_k)\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T(\mathbf{I} - \mathbf{\Pi}_k)\mathbf{V}_k]\mathbf{A} + \mathbf{A}^T\\
\mathbb{E}[\mathbf{V}_k^T\begin{bmatrix} (\zeta_k)_{\gamma_k}^{-\frac{1}{2}}(\mathbf{I} + \mathbf{P}_{\gamma_k}^{-1})^{-1}(\zeta_k)_{\gamma_k}^{-\frac{1}{2}} & (\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1}\mathbf{P}_{\gamma_k}\Sigma_k \\ \Sigma_k^T\mathbf{P}_{\gamma_k}(\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1} & -\widetilde{\eta}_k \end{bmatrix}]\\
\mathbf{V}_k]\mathbf{A},
\end{aligned} \tag{50}$$

where

$$\widetilde{\eta}_k = \Sigma_k^T\mathbf{P}_{\gamma_k}(\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1}\mathbf{P}_{\gamma_k}\Sigma_k. \tag{51}$$

Note that $(\mathbf{I} + \widetilde{\mathbf{P}}_{\gamma_k}^{-1})^{-1} \leq \mathbf{I}$ and $\widetilde{\eta}_k \geq \mathbf{0}$, it follows that

$$\begin{aligned}
\mathbb{E}[\mathbf{S}(\mathbf{P})] &\leq \mathbf{Q} + \mathbf{A}^T\mathbb{E}[\mathbf{P}_k^{uc}]\mathbf{A} + \|\mathbf{A}\|^2\mathbb{E}[\mathrm{Tr}((\zeta_k)_{\gamma_k}^{-1})]\mathbf{I}\\
&\leq \mathbf{A}^T\mathbb{E}[\mathbf{V}_k^T(\mathbf{I} - \mathbf{\Pi}_k)\mathbf{V}_k\mathbf{P}\mathbf{V}_k^T(\mathbf{I} - \mathbf{\Pi}_k)\mathbf{V}_k]\mathbf{A}\\
&\quad + \mathbf{Q} + \|\mathbf{A}\|^2\mathbb{E}[\mathrm{Tr}((\zeta_k)_{\gamma_k}^{-1})]\mathbf{I}.
\end{aligned} \tag{52}$$

We now construct two matrix sequences

$$\left\{\mathbf{P}_k^{(1)} : \mathbf{P}_{k+1}^{(1)} = g(\mathbf{P}_k^{(1)}), \mathbf{P}_0^{(1)} = \mathbf{P}^{(1)}, k \geq 0\right\}, \tag{53}$$

and

$$\left\{\mathbf{P}_k^{(2)} : \mathbf{P}_{k+1}^{(2)} = g(\mathbf{P}_k^{(2)}), \mathbf{P}_0^{(2)} = \mathbf{P}^{(2)}, k \geq 0\right\}. \tag{54}$$

Due to the monotonicity of $\mathbb{E}[\mathbf{S}(\mathbf{P})]$ w.r.t. $\mathbf{P}$, it follows that $\mathbf{P}_{k+1}^{(1)} \geq \mathbf{P}_k^{(1)}, \forall k \geq 0$, and $\mathbf{P}_{k+1}^{(2)} \leq \mathbf{P}_k^{(2)}, \forall k \geq 0$. Therefore, we have

$$\mathbf{P}_k^{(1)} \leq \mathbf{P}_{k+1}^{(1)} \leq \mathbf{P}_{k+1}^{(2)} \leq \mathbf{P}_k^{(2)} \leq \mathbf{P}^{(2)}. \tag{55}$$

Therefore, the monotonically increasing sequence $\left\{\mathbf{P}_k^{(1)}, k \geq 0\right\}$ is bounded from above, i.e., $\mathbf{P}_k^{(1)} \leq \mathbf{P}^{(2)}, \forall k \geq 0$, it follows that the sequence $\left\{\mathbf{P}_k^{(1)}, k \geq 0\right\}$ is convergent, i.e., there is a $\left(\mathbf{P}^{(1)}\right)^*$ such that

$$\lim_{k\to\infty}\mathbf{P}_k^{(1)} = \left(\mathbf{P}^{(1)}\right)^* = \mathbb{E}\left[\mathbf{S}\left(\left(\mathbf{P}^{(1)}\right)^*\right)\right]. \tag{56}$$

Therefore, the existence of $\mathbf{P}^*$ that satisfies (41) under the sufficient condition (13) in Theorem 1 is proved.

We further prove the uniqueness of $\mathbf{P}^*$ that satisfies (41) under the sufficient condition (13) in Theorem 1. Suppose there exist $(\mathbf{P}^*)^{(1)}$ and $(\mathbf{P}^*)^{(2)}$, which satisfies (42) and $(\mathbf{P}^*)^{(1)} \neq (\mathbf{P}^*)^{(2)}$. There is a positive constant $\gamma \in (0,1)$ such that $(\mathbf{P}^*)^{(1)} \geq \gamma(\mathbf{P}^*)^{(2)}$ and $(\mathbf{P}^*)^{(1)} \ngeq \gamma'(\mathbf{P}^*)^{(2)}$, where $\gamma' > \gamma$, since

$$\begin{aligned}
\mathbb{E}[\mathbf{S}(\mathbf{P}^*)^{(2)}] &= \mathbb{E}[\mathbf{A}(\zeta_k\zeta_k^T + (\gamma(\mathbf{P}^*)^{(2)})^{-1})^{-1}\mathbf{A}^T] + \mathbf{Q}\\
&\geq (1 + \Delta)\gamma\mathbb{E}[\mathbf{S}(\mathbf{P}^*)^{(2)}],
\end{aligned} \tag{57}$$

where $\Delta = \frac{(1-\gamma)\sigma_{min}(\mathbf{Q})}{\|\mathbb{E}[\mathbf{S}(\mathbf{P}^*)^{(2)}]\|} > 0$ and $\sigma_{min}(\mathbf{Q}) = \frac{1}{\|\mathbf{Q}^{-1}\|}$.

On the other hand, we have

$$(\mathbf{P}^*)^{(1)} \geq (1 + \Delta)\gamma\mathbb{E}[\mathbf{S}(\mathbf{P}^*)^{(2)}] = (1 + \Delta)\gamma(\mathbf{P}^*)^{(2)}. \tag{58}$$

As a result, $(1 + \Delta)\gamma > \gamma$. Since $\Delta > 0$, it contradicts our assumption, and hence $(\mathbf{P}^*)^{(1)} = (\mathbf{P}^*)^{(2)}$ and the solution to (42) is unique. This completes the proof for Theorem 1.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

B. Proof of Theorem 2

MDP theory indicates that the Bellman optimality equation (14) is satisfied by a pair of $(\theta, V(\mathbf{S}_k))$, then the following inequality holds:

$$r(\mathbf{S}_k, \mathbf{u}_k) + \sum_{\mathbf{S}_{k+1}} \Pr(\mathbf{S}_{k+1}|\mathbf{S}_k, \mathbf{u}_k) V(\mathbf{S}_{k+1}) \geq \theta + V(\mathbf{S}_k), \tag{59}$$

where the equality holds if and only if $\mathbf{u}_k = \mathbf{u}_k^*$ is the minimizer for the L.H.S. of (59).

Taking full expectation on both sides of (59), it follows that

$$\mathbb{E}[r(\mathbf{S}_k, \mathbf{u}_k)] + \mathbb{E}[V(\mathbf{S}_{k+1})] \geq \theta + \mathbb{E}[V(\mathbf{S}_k)]. \tag{60}$$

Summing up the above Inequality (60) on both sides from $k = 0$ to $k = K$ and dividing both sides by K, it follows that

$$\theta \leq \frac{1}{K} \sum_{k=0}^{K} \mathbb{E}[r(\mathbf{S}_k, \mathbf{u}_k)] + \frac{1}{K}(\mathbb{E}[V(\mathbf{S}_{k+1} - V(\mathbf{S}_0))]). \tag{61}$$

Moreover, (35) indicates that

$$V(\mathbf{S}_k) = \limsup_{T \to \infty} \sum_{t=0}^{T} \mathbb{E}_{\mathbf{u}_{k+t}^*}[r(\mathbf{S}_{k+t}, \mathbf{u}_{k+t}^*) - \theta]. \tag{62}$$

Equation (61) and (62) indicate that

$$(K+T)\theta = \limsup_{(K+T)\to\infty} \sum_{k=0}^{K+T} \mathbb{E}[r(\mathbf{S}_k, \mathbf{u}_k^*)], \tag{63}$$

and

$$\theta = \limsup_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K} \mathbb{E}[r(\mathbf{S}_k, \mathbf{u}_k^*)]. \tag{64}$$

Hence, if there exists a pair $(\theta, V(\mathbf{S}_k))$ that satisfies (59), then $V(\mathbf{S}_k)$ is the optimal value function over the extended state space $\mathbf{S}_k$ given by (62) and θ is the optimal average cost for Problem 1 given by (64). $\mathbf{u}_k^*$ is the optimizer to Problem 1. This concludes the proof for Theorem 2.

C. Proof of Theorem 3

Exploiting the i.i.d. property of the MIMO fading channels $\mathbf{H}_k$ and the controller random access δ_k, the optimality equation for Problem 1 in Theorem 2 can be represented as

$$\begin{aligned}
&\theta + V(\mathbf{x}_k, \delta_k \mathbf{H}_k) = \min_{\mathbf{u}_k}[r(\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) + \sum_{\mathbf{x}_{k+1}, \delta_{k+1}\mathbf{H}_{k+1}} \\
&\Pr(\mathbf{x}_{k+1}, \delta_{k+1}\mathbf{H}_{k+1}|\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) V(\mathbf{x}_{k+1}, \delta_{k+1}\mathbf{H}_{k+1})] \\
&= \min_{\mathbf{u}_k}[r(\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) + \sum_{\mathbf{x}_{k+1}} \Pr(\mathbf{x}_{k+1}|\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) \\
&(\sum_{\mathbf{H}_{k+1}} \Pr(\mathbf{H}_{k+1}) \sum_{\delta_{k+1}} \Pr(\delta_{k+1}) V(\mathbf{x}_{k+1}, \delta_{k+1}\mathbf{H}_{k+1}))] \\
&= \min_{\mathbf{u}_k}[r(\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) + \sum_{\mathbf{x}_{k+1}} \Pr(\mathbf{x}_{k+1}|\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) \widetilde{V}(\mathbf{x}_{k+1})].
\end{aligned} \tag{65}$$

Taking the expectation of both sides of (65) over $\delta_k \mathbf{H}_k$, it follows that

$$\begin{aligned}
\theta + \widetilde{V}(\mathbf{x}_k) = \mathbb{E}[\min_{\mathbf{u}_k}[r(\mathbf{x}_k, \delta_k \mathbf{H}_k, \mathbf{u}_k) + \sum_{\mathbf{x}_{k+1}} \Pr(\mathbf{x}_{k+1}|\mathbf{x}_k, \\
\delta_k \mathbf{H}_k, \mathbf{u}_k) \widetilde{V}(\mathbf{x}_{k+1})]].
\end{aligned} \tag{66}$$

As a result, if there exists a pair of $(\widetilde{\theta}, \widetilde{V}(\mathbf{x}_j))$ that solves (66), then $\widetilde{\theta} = \theta$ is the optimal average cost for Problem 1 given by Theorem 2. $\widetilde{V}(\mathbf{x}_k) = \mathbb{E}[V(\mathbf{S}_k)|\mathbf{x}_k]$ is the reduced-state value function, and the optimal control policy for Problem 1 is given by $\mathbf{u}_k^*$, which achieves the minimum value of the L.H.S. of the (66) and (59). This concludes the proof for Theorem 3.

D. Proof of Theorem 4

We define a Lyapunov function as follows [22]:

$$V_k = \left\| \mathbf{A} - \widehat{\mathbf{A}}_k \right\|_F^2. \tag{67}$$

The associated Lyapunov drift is given by

$$\Lambda\left(\widehat{\mathbf{A}}_k\right) = \mathbb{E}\left\{ V_{k+1} - V_k | \, \widehat{\mathbf{A}}_k \right\}. \tag{68}$$

Substitute (28) and (67) into (68), the Lyapunov drift is upper bounded as follows:

$$\begin{aligned}
&\Lambda\left(\widehat{\mathbf{A}}_k\right) \leq \\
&\mathbb{E}\Bigg\{ \mathrm{Tr}\left(\alpha_k^2 \mathbf{x}_k \mathbf{x}_k^T \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right)^T \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right) \right) \\
&+ \mathrm{Tr}\left(-2\alpha_k \mathbf{x}_k \mathbf{x}_k^T \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right)^T \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right) \right) + \alpha_k^2 \mathbf{x}_k \mathbf{x}_k^T \Big| \widehat{\mathbf{A}}_k \Bigg\}.
\end{aligned} \tag{69}$$

Note that if the bounded conditional variance condition in Theorem 4 is satisfied, the Lyapunov drift (69) can be simplified as

$$\Lambda\left(\widehat{\mathbf{A}}_k\right) \leq \eta\left(\alpha_k^2 - \alpha_k\right) \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right)^T \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right) + \alpha_k^2 \mu \mathbf{I}. \tag{70}$$

As a result, if the step-size condition in Theorem 4 is also satisfied, it follows that $\limsup_{k\to\infty} \mathbb{E}\left[\left\| \mathbf{A} - \widehat{\mathbf{A}}_k \right\|_F^2 \right] = 0$, which results in the almost sure convergence of $\widehat{\mathbf{A}}_k$ to $\mathbf{A}$. Therefore, Theorem 4 is proved.

E. Proof of Theorem 5

We shall analyze the convergence of $\widehat{\mathbf{A}}_k$ in Step 1 of Algorithm 3 using the theory of Lyapunov drift. Specifically, we define a Lyapunov function as follows [22]:

$$V_k = \left\| \mathbf{A} - \widehat{\mathbf{A}}_k \right\|_F^2. \tag{71}$$

The associated Lyapunov drift is given by

$$\Lambda\left(\widehat{\mathbf{A}}_k\right) = \mathbb{E}\left\{ V_{k+1} - V_k | \, \widehat{\mathbf{A}}_k \right\}. \tag{72}$$

Substitute (28) and (71) into (72), the Lyapunov drift is upper bounded as follows:

$$\begin{aligned}
\Lambda\left(\widehat{\mathbf{A}}_k\right) \leq \mathbb{E}\Bigg\{ &\mathrm{Tr}((\mathbf{x}_k \mathbf{x}_k^T \mathbf{1}_{\left\{\|\mathbf{x}_k\|^2 < 1\right\}} + \frac{\mathbf{x}_k \mathbf{x}_k^T}{\|\mathbf{x}_k\|^2} \mathbf{1}_{\left\{\|\mathbf{x}_k\|^2 \geq 1\right\}}) \\
&\left(\mathbf{A} - \widehat{\mathbf{A}}_k \right)^T \left(\mathbf{A} - \widehat{\mathbf{A}}_k \right)) \left(\alpha_k^{-2} - 2\alpha_k^{-1} \right) \\
&+ \alpha_k^2 \frac{\mathbf{x}_k \mathbf{x}_k^T}{\|\mathbf{x}_k\|^4} \Big| \widehat{\mathbf{A}}_k \Bigg\}.
\end{aligned} \tag{73}$$

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

Let α_k satisfy the Condition (26). It follows that $\left(k^{-2}-2k^{-1}\right) < 0$. Note that the last term in (73) is bounded due to the normalization $\alpha_k^2 \frac{\mathbf{x}_k\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2} \leq \frac{1}{k^2}\mathbf{I}_S$. Moreover, the normalized condition variance is upper bounded as $\mathbb{E}\left[\frac{\mathbf{x}_k\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2}\middle|\widehat{\mathbf{A}}_k\right] < \mu\mathbf{I}_S$, where $\mu > 0$ is a constant. Further note that $\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right] > \mathbf{W} > \mathbf{0}$, it follows that $\mathrm{Rank}\left(\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^T\middle|\widehat{\mathbf{A}}_k\right]\right) = \mathrm{Rank}\left(\mathbb{E}\left[\frac{\mathbf{x}_k\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2}\middle|\widehat{\mathbf{A}}_k\right]\right) = S$. Further note that $\mathbb{E}\left[\frac{\mathbf{x}_k\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2}\middle|\widehat{\mathbf{A}}_k\right] = \mathbf{0}$ if and only if $\mathbf{x}_k = \mathbf{0}$, it follows that there exists a positive constant η such that $\mathbb{E}\left[\frac{\mathbf{x}_k\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2}\middle|\widehat{\mathbf{A}}_k\right] > \eta\mathbf{I}_S$. Therefore, we have

$$\Lambda\left(\widehat{\mathbf{A}}_k\right) \leq \eta\left(k^{-2}-2k^{-1}\right)\left(\mathbf{A}-\widehat{\mathbf{A}}_k\right)^T\left(\mathbf{A}-\widehat{\mathbf{A}}_k\right)+k^{-2}. \tag{74}$$

It follows that

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{A}-\widehat{\mathbf{A}}_{k+1}\right\|_F^2\right] &< \left(1+\eta\left(k^{-2}-2k^{-1}\right)\right)\\ &\cdot\mathbb{E}\left[\left\|\mathbf{A}-\widehat{\mathbf{A}}_k\right\|_F^2\right]+k^{-2}, \forall k>0.\end{aligned} \tag{75}$$

According to the standard Lyapunov theory, it follows that $\limsup_{k\to\infty}\mathbb{E}\left[\left\|\mathbf{A}-\widehat{\mathbf{A}}_k\right\|_F^2\right] = 0$, which results in $\Pr\left(\lim_{k\to\infty}\widehat{\mathbf{A}}_k = \mathbf{A}\right) = 1$. Therefore, Theorem 5 is proved.

F. Proof of Theorem 6

We use the ordinary differential equation (ODE) method to analyze the convergence of $\mathbf{P}_k$. Specifically, under the condition for stepsize sequence $\{\alpha_k, k\geq 0\}$ in Theorem 5, the update rule in (30) can be approximated by the following ODE:

$$\dot{\mathbf{P}}(t) = f(\mathbf{P}(t)), \mathbf{P}(0) = \mathbf{P}_0, t\in\mathbb{R}, \tag{76}$$

and the following lemma characterizes the relationship between the convergence of the the update rule in (30) and the ODE trajectory (76).

Lemma 2: (Relationship between the Convergence Behavior of $\mathbf{P}_k$ and the ODE Trajectory (76)) If the limiting ODE (76) has a unique equilibrium point $\mathbf{P}^*$ that is globally asymptotically stable, then $\mathbf{P}_k$ via (30) will converge to $\mathbf{P}^*$ almost surely.

Proof: According to Theorem 2.1 of [56], the proof can be concluded by verifying the following three conditions:

- (a) *(Lipschitz Continuity)* $f(\mathbf{P})$ satisfies $\|f(\mathbf{P}^{(1)}) - f(\mathbf{P}^{(2)})\| \leq (1+\|\mathbf{A}\|^2)\|\mathbf{P}^{(1)}-\mathbf{P}^{(2)}\|, \forall \mathbf{P}^{(1)}, \mathbf{P}^{(2)} \in \mathbb{S}_+^S$;
- (b) *(Martingale Difference Noise)* Let $\widehat{f}(\mathbf{P}_k) = \mathbf{Q} + \mathbf{A}^T\mathbf{P}_k\mathbf{A} - \delta_k\mathbf{A}^T\mathbf{P}_k\mathbf{B}\mathbf{H}_k(\mathbf{H}_k^T(\mathbf{B}^T\mathbf{P}_k\mathbf{B}+\mathbf{M})\mathbf{H}_k + \mathbf{R})^{-1}\mathbf{H}_k^T\mathbf{B}^T\mathbf{P}_k\mathbf{A}-\mathbf{P}_k$ and $\mathbf{N}_k = \widehat{f}(\mathbf{P}_k)-f(\mathbf{P}_k)$. Then, the sequence $\{\mathbf{N}_k, k\geq 0\}$ is a martingale difference sequence w.r.t. the filtration $\left\{\mathcal{F}_k \triangleq \sigma(\mathbf{P}_0, \delta_0\mathbf{H}_0, ...\delta_k\mathbf{H}_k)\right\}$ satisfying $\mathbb{E}[\mathbf{N}_{k+1}|\mathcal{F}_k] = \mathbf{0}_{S\times S}$;
- (c) *(Square Integrability)* The sequence $\{\mathbf{N}_k, k\geq 0\}$ is square-integrable with $\mathbb{E}[\|\mathbf{N}_{k+1}\|^2|\mathcal{F}_k] \leq 2\|\mathbf{A}\|^2(1+\|\mathbf{P}_k\|^2), \forall k>0$.

For Condition (a), we first note that

$$\begin{aligned}f(\mathbf{P}^{(1)}) - f(\mathbf{P}^{(2)}) \leq \mathbb{E}[\mathbf{A}^T(\mathbf{I}-\mathbf{K}_k^{(2)}\Xi)(\mathbf{P}^{(1)}-\mathbf{P}^{(2)})\\ (\mathbf{I}-\mathbf{K}_k^{(2)}\Xi)^T\mathbf{A}].\end{aligned} \tag{77}$$

Since $\|\mathbf{A}^T(\mathbf{I}-\mathbf{K}_k^{(2)}\Xi)\| \leq \|\mathbf{A}\|$, it follows that

$$\begin{aligned}\|f(\mathbf{P}^{(1)}) - f(\mathbf{P}^{(2)})\| \leq \|\mathbf{P}^{(1)}-\mathbf{P}^{(2)}\| + \mathbb{E}[\|\mathbf{A}^T(\mathbf{I}-\mathbf{K}_k^{(2)}\Xi)\|^2\\ \|\mathbf{P}^{(1)}-\mathbf{P}^{(2)}\|] \leq (1+\|\mathbf{A}\|^2)\|\mathbf{P}^{(1)}-\mathbf{P}^{(2)}\|,\end{aligned} \tag{78}$$

and Condition (a) is verified.

For Condition (b), we note that for any given realization of $\mathbf{P}_k$, $\mathbf{N}_{k+1} = \widehat{f}(\mathbf{P}_{k+1}) - f(\mathbf{P}_{k+1})$ is a function of $\{\delta_{k+1}\mathbf{H}_{k+1}\}$. It follows that $\mathbb{E}[\mathbf{N}_{k+1}|\mathcal{F}_k] = \mathbf{0}_{S\times S}$ and Condition (b) is verified.

For Condition (c), it can be verified by

$$\begin{aligned}\mathbb{E}[\|\mathbf{N}_{k+1}\|^2\mathbf{P}_k] \leq \mathbb{E}[\|\widehat{f}(\mathbf{P}_{k+1}) - \mathbf{Q}\|^2 + \|f(\mathbf{P}_{k+1})-\\ \mathbf{Q}\|^2|\mathbf{P}_k] \leq 2\|\mathbf{A}\mathbf{P}_k\mathbf{A}^T\|^2 \leq 2\|\mathbf{A}\|^2\|\mathbf{P}_k\|^2.\end{aligned} \tag{79}$$

■

The task now turns to analyze the existence of the unique equilibrium point $\mathbf{P}^*$ for the limiting ODE (76) that is globally asymptotically stable. To achieve this goal, we introduce the following virtual fixed-point process $\left\{\widetilde{\mathbf{P}}_k, k\geq 0\right\}$:

$$\widetilde{\mathbf{P}}_{k+1} = \widetilde{\mathbf{P}}_k + \xi f\left(\widetilde{\mathbf{P}}_k\right), \widetilde{\mathbf{P}}_0 = \mathbf{P}_0,\ \forall k\geq 0, \tag{80}$$

where $\xi\in(0,1)$ is a constant. We have the following lemma to characterize the relationship between the state trajectory of the virtual fixed-point process (80) and the state trajectory of the solution to the limiting ODE (76).

Lemma 3: (Relationship between the ODE Trajectory (76) and the Virtual Fixed-point Process (80)) Let $t_k = k\xi, \forall k\geq 0$. We define a continuous piece-wise linear function $\bar{\mathbf{P}}(t), \forall t\geq 0$, by $\bar{\mathbf{P}}(t_k) = \widetilde{\mathbf{P}}_k$ with linear interpolation on the interval $[t_k, t_{k+1}]$ as

$$\bar{\mathbf{P}}(t) = \widetilde{\mathbf{P}}_k + \xi^{-1}(t-t_k)(\widetilde{\mathbf{P}}_{k+1} - \widetilde{\mathbf{P}}_k). \tag{81}$$

Let $\mathbf{P}^l(t), t\geq l$, denote the trajectory of ODE (76) with initial condition $\mathbf{P}^l(t)|_{t=l} = \bar{\mathbf{P}}(l), \forall l\in\mathbb{R}^+$. Then, for any $l>0$ and $L>0$, it follows that

$$\sup_{t\in[0,L]}\left\|\bar{\mathbf{P}}(l+t) - \mathbf{P}^l(l+t)\right\| = \mathcal{O}(\xi). \tag{82}$$

Proof: Let $L = N\xi$ for some $N>0$. For $t>0$, let $[t] = \max\{k\xi : n>0, k\xi<t\}$. For $n\geq 0$ and $1\leq l\leq L$, we have

$$\bar{\mathbf{P}}(t_{k+l}) = \bar{\mathbf{P}}(t_k) + \int_{t_k}^{t_{k+l}} f(\bar{\mathbf{P}}([t]))dt, \tag{83}$$

$$\begin{aligned}\mathbf{P}^{t_k}(t_{k+l}) = \bar{\mathbf{P}}(t_k) + \int_{t_k}^{t_{k+l}} f(\mathbf{P}^{t_k}([t]))dt+\\ \int_{t_k}^{t_{k+l}}(f(\mathbf{P}^{t_k}(t)) - f(\mathbf{P}^{t_k}([t])))dt.\end{aligned} \tag{84}$$

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

It follows that

$$\sup_{0\le j\le l}\|\bar{\mathbf{P}}(t_{k+j})-\mathbf{P}^{t_k}(t_{k+j})\|\le c_1\xi(1+\bar{\mathbf{P}}(t_k))+\xi L\|\mathbf{A}\|^2\sum_{m=0}^{l-1}\sup_{j\le m}\|\bar{\mathbf{P}}(t_{k+j})-\mathbf{P}^{t_k}(t_{k+j})\|, \tag{85}$$

where c_1 is a constant. By the Gronwall inequality, it follows that

$$\sup_{k\le j\le k+N}\|\bar{\mathbf{P}}(t_j)-\mathbf{P}^{t_k}(t_j)\|^2\le c_2\xi, \tag{86}$$

where c_2 is a constant. Since both $\sup_{t_j\le t\le t_{j+1}}\|\bar{\mathbf{P}}(t)-\bar{\mathbf{P}}(t_j)\|^2$ and $\sup_{t_j\le t\le t_{j+1}}\|\mathbf{P}^{t_k}(t)-\mathbf{P}^{t_k}(t_j)\|^2$ are $\mathcal{O}(\xi)$, it follows that

$$\sup_{t\in[0,L]}\|\bar{\mathbf{P}}(l+t)-\mathbf{P}^{l}(l+t)\|\le c_3\xi, \tag{87}$$

where c_3 is a constant. This concludes the proof. ■

As a result, the gap between the state trajectory of the virtual fixed-point process (80) and that of the limiting ODE (76) is $\mathcal{O}(\xi)$, which can be made arbitrarily small by letting $\xi\to 0$. Therefore, the convergence of the state trajectory of the virtual fixed-point process (80) under arbitrary $\xi\in(0,1)$ implies the convergence of the state trajectory of the limiting ODE (76), which in turn leads to the convergence of $\mathbf{P}_k$ updated by (30). Hence, the task turns to prove the convergence of the state trajectory of the virtual fixed-point process (80) under arbitrary $\xi\in(0,1)$.

We now analyze the convergence behaviors of the virtual fixed-point process $\left\{\widetilde{\mathbf{P}}_k, k\ge 0\right\}$ in (80) under arbitrary $\xi\in(0,1)$. Denote $\widetilde{g}(\mathbf{P})=\mathbf{P}+\xi f(\mathbf{P})$. We know that if the sufficient condition (13) in Theorem 1 is satisfied, the solution $\mathbf{P}^*$ to the fixed-point equation $\mathbf{P}^*=\widetilde{g}(\mathbf{P}^*)$ exists and is unique. Using the same techniques as in Appendix A, there is a $\widetilde{\mathbf{P}}^{(1)}=\mathbf{0}$ such that $\widetilde{\mathbf{P}}^{(1)}<\widetilde{g}\left(\widetilde{\mathbf{P}}^{(1)}\right)$, and a sufficiently large $\widetilde{\mathbf{P}}^{(2)}$ such that $\widetilde{\mathbf{P}}^{(2)}>\widetilde{g}\left(\widetilde{\mathbf{P}}^{(2)}\right)$. We now construct the following two matrix sequences:

$$\left\{\widetilde{\mathbf{P}}_k^{(1)}:\widetilde{\mathbf{P}}_{k+1}^{(1)}=\widetilde{g}(\widetilde{\mathbf{P}}_k^{(1)}),\mathbf{P}_0^{(1)}=\mathbf{0},k\ge 0\right\}, \tag{88}$$

$$\left\{\widetilde{\mathbf{P}}_k^{(2)}:\widetilde{\mathbf{P}}_{k+1}^{(2)}=\widetilde{g}(\widetilde{\mathbf{P}}_k^{(2)}),\mathbf{P}_0^{(2)}=\widetilde{\mathbf{P}}^{(2)},k\ge 0\right\}. \tag{89}$$

Let the initial condition of the fixed-point process be $\mathbf{0}\le\widetilde{\mathbf{P}}_0\le\widetilde{\mathbf{P}}^{(2)}$, it follows that $\widetilde{\mathbf{P}}_k^{(1)}\le\widetilde{\mathbf{P}}_k\le\widetilde{\mathbf{P}}_k^{(2)}$. Let $k\to\infty$ and note that $\mathbf{P}^*$ exists and is unique, it follows that

$$\mathbf{P}^*=\lim_{k\to\infty}\widetilde{\mathbf{P}}_k^{(1)}\le\lim_{k\to\infty}\widetilde{\mathbf{P}}_k\le\lim_{k\to\infty}\widetilde{\mathbf{P}}_k^{(2)}=\mathbf{P}^*. \tag{90}$$

Since $\widetilde{\mathbf{P}}^{(2)}$ can be arbitrarily large, it follows that for any bounded initial value $\widetilde{\mathbf{P}}_0$, the virtual fixed-point process $\left\{\widetilde{\mathbf{P}}_k, k\ge 0\right\}$ in (80) converges to $\mathbf{P}^*$. Therefore, the limiting ODE (76) has a unique equilibrium point $\mathbf{P}^*$ that is globally asymptotically stable. It follows that the $\mathbf{P}_k$ obtained by the proposed Algorithm 3 converges to $\mathbf{P}^*$ almost surely. Based on the structural properties in Theorem 3, it follows that $\widetilde{V}_k(\mathbf{x}_k)$ and $\mathbf{u}_k(\mathbf{x}_k)$ converges to the optimal value function $\widetilde{V}(\mathbf{x}_k)$ and optimal control action $\mathbf{u}^*(\mathbf{x}_k)$ w.p.1., respectively. Therefore, Theorem 6 is proved.

REFERENCES

[1] Y. Liu, J. Wang, J. Li, S. Niu, and H. Song, "Machine learning for the detection and identification of internet of things devices: A survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 298–320, 2021.

[2] M. Weiss, M. Luck, R. Girgis, C. Pal, and J. P. Cohen, "A survey of mobile computing for the visually impaired," *arXiv preprint arXiv:1811.10120*, 2018.

[3] S. Li, L. Da Xu, and S. Zhao, "5G internet of things: A survey," *J. Ind. Inf. Integr.*, vol. 10, pp. 1–9, 2018.

[4] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 2, pp. 905–929, 2020.

[5] Y. A. Qadri, A. Nauman, Y. B. Zikria, A. V. Vasilakos, and S. W. Kim, "The future of healthcare internet of things: a survey of emerging technologies," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 2, pp. 1121–1167, 2020.

[6] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 2, pp. 1191–1221, 2020.

[7] H. Guo, Y. Zhu, H. Ma, V. K. Lau, K. Huang, X. Li, H. Nong, and M. Zhou, "Over-the-air aggregation for federated learning: Waveform superposition and prototype validation," *J. Commun. Netw.*, vol. 6, no. 4, pp. 429–442, 2021.

[8] M. Humayun, N. Jhanjhi, B. Hamid, and G. Ahmed, "Emerging smart logistics and transportation using IoT and blockchain," *IEEE Internet Things Mag.*, vol. 3, no. 2, pp. 58–62, 2020.

[9] O. Hamdan, H. Shanableh, I. Zaki, A. Al-Ali, and T. Shanableh, "IoT-based interactive dual mode smart home automation," *IEEE Int. Conf. Consum. Electron. (ICCE)*, pp. 1–2, 2019.

[10] A. A. Mutlag, M. K. Abd Ghani, N. a. Arunkumar, M. A. Mohammed, and O. Mohd, "Enabling technologies for fog computing in healthcare IoT systems," *Future Gener. Comput. Syst.*, vol. 90, pp. 62–78, 2019.

[11] A. Maroli, V. S. Narwane, and B. B. Gardas, "Applications of IoT for achieving sustainability in agricultural sector: A comprehensive review," *J. Environ. Manage.*, vol. 298, p. 113488, 2021.

[12] D. Minovski, C. Åhlund, and K. Mitra, "Modeling quality of IoT experience in autonomous vehicles," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3833–3849, 2020.

[13] Z. Lv, Y. Han, A. K. Singh, G. Manogaran, and H. Lv, "Trustworthiness in industrial IoT systems based on artificial intelligence," *IEEE Trans. Ind. Inform.*, vol. 17, no. 2, pp. 1496–1504, 2020.

[14] M. Guida, F. Marulo, , and S. Abrate, "Advances in crash dynamics for aircraft safety," *Prog. Aerosp. Sci.*, vol. 98, pp. 106–123, 2018.

[15] Y. Sun, B. McMillin, X. Liu, and D. Cape, "Verifying noninterference in a cyber-physical system the advanced electric power grid," *IEEE Int. Conf. Softw. (QSIC)*, pp. 363–369, 2007.

[16] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[17] A. Wills and L. Ljung, "Wiener system identification using the maximum likelihood method," in *Block-oriented nonlinear system identification*. Springer, 2010, pp. 89–110.

[18] G. Saridis and G. Stein, "Stochastic approximation algorithms for linear discrete-time system identification," *IEEE Trans. Autom. Control*, vol. 13, no. 5, pp. 515–523, 1968.

[19] T. Nguyen and Z. Gajic, "Solving the matrix differential Riccati equation: a Lyapunov equation approach," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 191–194, 2009.

[20] N. A. Khan, A. Ara, and M. Jamil, "An efficient approach for solving the Riccati equation with fractional orders," *Comput. Math. Appl.*, vol. 61, no. 9, pp. 2683–2689, 2011.

[21] F. Huang, Y. Yang, Z. Zheng, G. Wu, and S. Mumtaz, "Recognizing influential nodes in social networks with controllability and observability," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6197–6204, 2020.

[22] Z. He, J. Yin, Y. Wang, G. Gui, B. Adebisi, T. Ohtsuki, H. Gacanin, and H. Sari, "Edge device identification based on federated learning and network traffic feature engineering," *IEEE Trans. Cogn. Commun. Netw.*, early access, 2021.

[23] S. I. Popoola, R. Ande, B. Adebisi, G. Gui, M. Hammoudeh, and O. Jogunola, "Federated deep learning for zero-day botnet attack detection in IoT edge devices," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3930–3944, 2022.

[24] M. Ikenoue, S. Kanae, and K. Wada, "On the recursive algorithm of bias compensated weighted least squares method," *IEEE Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, pp. 522–527, 2019.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3175965, IEEE Internet Things Journal

[25] K. Chen, Z. Li, D. Filev, Y. Wang, K. Wu, and J. Wang, "Online nonlinear dynamic system identification with evolving spatial–temporal filters: Case study on turbocharged engine modeling," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 3, pp. 1364–1371, 2020.

[26] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted L-1 balls," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 936–952, 2010.

[27] L. Ma and X. Liu, "Recursive maximum likelihood method for the identification of Hammerstein ARMAX system," *Appl. Math. Model.*, vol. 40, no. 13-14, pp. 6523–6535, 2016.

[28] R. Snyder, "Recursive estimation of dynamic linear models," *J. Res. Stat. Soc. Ser. B, Stat. Methodol.*, pp. 272–276, 1985.

[29] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Rehabil. Robots Int. Conf.*, 2011, pp. 1–7.

[30] P. Baldi, "Gradient descent learning algorithm overview: A general dynamical systems perspective," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 182–195, 1995.

[31] B. Luo, Y. Yang, H.-N. Wu, and T. Huang, "Balancing value iteration and policy iteration for discrete-time control," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 50, no. 11, pp. 3948–3958, 2019.

[32] T. Bian and Z.-P. Jiang, "Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design," *Automatica*, vol. 71, pp. 348–360, 2016.

[33] X. Li, Z. Peng, and L. Liang, "Off-policy Q-learning for infinite horizon LQR problem with unknown dynamics," *IEEE Int. Symp. Ind. Electron. (ISIE)*, pp. 258–263, 2018.

[34] W. Kang and L. C. Wilcox, "Mitigating the curse of dimensionality: sparse grid characteristics method for optimal feedback control and HJB equations," *Computational Optimization and Applications*, vol. 68, no. 2, pp. 289–315, 2017.

[35] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *nature*, vol. 473, no. 7346, pp. 167–173, 2011.

[36] C. L. Robinson and P. Kumar, "Optimizing controller location in networked control systems with packet drops," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 661–671, 2008.

[37] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the latency, rate, and reliability tradeoff in wireless networked control systems for iiot," *IEEE Internet Things J.*, vol. 8, no. 2, pp. 723–733, 2020.

[38] Y. Wu, Q. Yang, H. Li, and K. S. Kwak, "Optimal control-aware transmission for mission-critical M2M communications under bandwidth cost constraints," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5924–5937, 2020.

[39] X. Yu, G. Jin, and J. Li, "Target tracking algorithm for system with Gaussian/non-Gaussian multiplicative noise," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 90–100, 2019.

[40] F. Crevecoeur, R. Sepulchre, J.-L. Thonnard, and P. Lefèvre, "Improving the state estimation for optimal control of stochastic processes subject to multiplicative noise," *Automatica*, vol. 47, no. 3, pp. 591–596, 2011.

[41] Y. Xingkai and J. Li, "Adaptive Kalman filtering for recursive both additive noise and multiplicative noise," *IEEE Trans. Aerosp. Electron. Syst.*, early access, 2021.

[42] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[43] S. Cai and V. K. Lau, "Zero MAC latency sensor networking for cyber-physical systems," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3814–3823, 2018.

[44] Y.-S. Wang, N. Matni, and J. C. Doyle, "Localized LQR optimal control," in *Proc. 53rd IEEE Conf. Decis. Control*, 2014, pp. 1661–1668.

[45] Q. Li, X. Feng, H. Wang, and L. Sun, "Understanding the usage of industrial control system devices on the internet," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2178–2189, 2018.

[46] J. Cui, X. Chen, J. Zhang, Q. Zhang, and H. Zhong, "Toward achieving fine-grained access control of data in connected and autonomous vehicles," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 7925–7937, 2020.

[47] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, 2009.

[48] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 41, no. 1, pp. 14–25, 2010.

[49] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[50] P. Bank and M. Voß, "Linear quadratic stochastic control problems with stochastic terminal constraint," *SIAM J. Control Optim.*, vol. 56, no. 2, pp. 672–699, 2018.

[51] E. Barron and H. Ishii, "The Bellman equation for minimizing the maximum cost," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 13, no. 9, pp. 1067–1090, 1989.

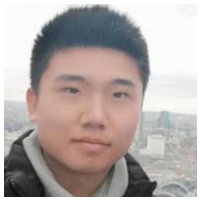
[52] M. El Chamie, Y. Yu, B. Açıkmeşe, and M. Ono, "Controlled Markov processes with safety state constraints," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1003–1018, 2018.

[53] Q. Zhang, M. Lin, L. T. Yang, Z. Chen, S. U. Khan, and P. Li, "A double deep Q-learning model for energy-efficient edge scheduling," *IEEE Trans. Serv. Comput.*, vol. 12, no. 5, pp. 739–749, 2018.

[54] A. S. Leong, S. Dey, and J. Anand, "Optimal LQG control over continuous fading channels," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 6580–6585, 2011.

[55] Y. Li and Y. Liang, "Learning overparameterized neural networks via stochastic gradient descent on structured data," *Adv. Neural Inf. Process. Syst.*, pp. 8157–8166, 2018.

[56] H. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*. Springer Science & Business Media, 2003, vol. 35.



Minjie Tang (Graduate Student Member, IEEE) received the B.Eng. degree in information and communication engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently working toward the Ph.D. degree at the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong. His research interests include wireless communication, industrial Internet of Things (IIoT), learning-driven control, wireless sensing and networked control systems.

Songfu Cai (Member, IEEE) received the Ph.D. degree in electronic and computer engineering (ECE) from The Hong Kong University of Science and Technology (HKUST) in 2019. He is currently a Post-Doctoral Research Fellow with the Department of ECE, HKUST. He received the Hong Kong Ph.D. Fellowship (HKPF) in 2013. His research interests include wireless communication, industrial Internet of Things (IIoT), learning-driven radio resource management, and networked control systems.

Vincent K. N. Lau (Fellow, IEEE) obtained B.Eng (Distinction 1st Hons) from the University of Hong Kong (1989-1992) and Ph.D. from the Cambridge University (1995-1997). He was with Bell Labs from 1997-2004 and the Department of ECE, Hong Kong University of Science and Technology (HKUST) in 2004. He is currently a Chair Professor and the Founding Director of Huawei-HKUST Joint Innovation Lab at HKUST. His current research focus includes wireless communications for 5G systems, content-centric wireless networking, wireless networking for mission-critical control, and cloud-assisted autonomous systems.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on May 23,2022 at 02:39:28 UTC from IEEE Xplore. Restrictions apply.