Q-Learning-Based Damping Control of Wide-Area Power Systems Under Cyber Uncertainties

Jiajun Duan, *Student Member, IEEE*, Hao Xu, *Member, IEEE*, and Wenxin Liu, *Senior Member, IEEE*

***Abstract*—The installation of phasor measurement units brings about system-wide synchronized real-time measurements, which makes advanced closed-loop control of wide-area power systems (WAPSs) possible. In addition to uncertainties with the physical system, network imperfections are also critical for control performance and even stability of the system. This paper targets at wide-area damping control of WAPS under both physical and cyber uncertainties. The cyber uncertainties addressed in this paper include both communication delay and package dropout. First, a linearized WAPS model is developed with considerations of random control signal arrivals. Then, a *Q*-learning-based control algorithm is designed to learn the optimal control under both physical and cyber uncertainties. Finally, to further counteract the communication uncertainties between the control center and actuators, a novel control law is designed based on the statistical analysis to make the long-term control performance approach that under ideal communications. The designed control algorithm is tested with both simplified linear model and detailed nonlinear model of WAPSs. Simulation results demonstrate the effectiveness of the designed algorithm.**

***Index Terms*—Wide-area power systems, networked control system, network imperfections, communication delay, package dropout, Q-learning algorithm.**

I. Introduction

WITH the expansion of power grid, Wide-Area Power Systems (WAPSs) are established by interconnecting the regional power systems of different areas [1]. Typical examples of WAPSs include the Western Electricity Coordinating Council (WECC) system in the U.S. [2], [3] and the South and Central China Grids [4]. The main advantages of the WAPSs include the establishment of mutual backup, the enhancement of redundancy, and increased capability and efficiency of generation allocation [3], [5]. However, the WAPSs are difficult to manage due to complexity and uncertainty.

In the past years, considerable investment has been used to develop the Wide-Area Measurement Systems (WAMS). A massive number of Phasor Measurement Units (PMUs) have been deployed in WAPSs all over the world [5], [6]. The systematic installation of synchronized PMUs makes the advanced real-time control of WAPSs possible. However, the complexities of physical and cyber systems of WAPS make the hardware investment hard to accommodate the application of advanced wide-area control sufficiently [7]. Instead of closed-loop control, open-loop monitoring remains to be the major application of the PMUs at present [5]. WAMS still has large room and potentials to be exploited.

As one of the most important problems with WAPSs, the inter-area oscillations degrade the power transfer capability and even cause system instability. Insufficient damping effort could result in system collapse with severe economic losses, such as the large-scale power outage accident of the WECC system on August 10, 1996 [3]. The traditional solution to this problem has been to install power system stabilizers (PSSs). Conventional PSSs (CPSSs) are designed based on phase compensation theory and introduced as lead-lag compensators [8]. Due to the lack of global vision and difficulty with the parameter tuning, such CPSSs cannot provide satisfactory damping performance [9]. The deployment of WAMS provides control center with the real-time global synchronized measurements that can be used to perform advanced damping control. Timely information delivery is critical for such advanced WAMS based control solutions [10]. Since the network imperfections such as communication delays and package dropouts are unavoidable in the real world, it is necessary to formulate and solve the wide-area control problem as a networked control problem [11]. However, there are not many networked control based solutions for damping control of WAPS [12]. It is mainly because of the combined cyber- and physical-complexities of the WAPS, such as network imperfections, model uncertainties, and wide range operating conditions of WAPS [10]–[14].

In [11], the time-varying delay is incorporated into a linear Networked Control System (NCS) model for WAPS control design. A Linear Matrix Inequality (LMI) method based controller is designed with wide-area measurements. Due to the conservativeness of the LMI method, a feasible solution might be difficult to find even if it exists [15]. In [10], the system outputs are estimated based on a linear model which uses the Smith prediction method to counteract the impacts of communication delays. The maximum delay that can be compensated by the designed controller is evaluated,

Manuscript received August 15, 2016; revised December 6, 2016 and March 2, 2017; accepted May 9, 2017. Date of publication June 5, 2017; date of current version October 19, 2018. Paper no. TSG-01078-2016. *(Corresponding author: Wenxin Liu.)*

J. Duan and W. Liu are with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: jid213@lehigh.edu; wel814@lehigh.edu).

H. Xu is with the Department of Electrical and Biomedical Engineering, University of Nevada, Reno, NV 89557 USA (e-mail: haoxu@unr.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2017.2711599

but the control performance of this method is sensitive to the estimation accuracy. In [13], a predictor-based $\mathcal{H}_\infty$ control is designed based on a linear time-delayed WAMS model. A unified Smith prediction method is used to improve estimation accuracy. But the method is only tested under constant communication delay.

Basically, above wide-area control algorithms require exact physical system dynamics to calculate the control signals. However, this requirement is hard to be satisfied in WAPSs due to the complexity of physical system together with the wide range of operating conditions [17]. The desired solution should have the adaptivity to learn the effective control of unknown system dynamics and uncertainties. Since reinforcement learning (RL) does not require system model and can learn the optimal control based on system's responses, it is an ideal solution for such control problems [16], [17]. It has been demonstrated that sufficient input-output data could effectively capture the practical system model even it is uncertain and time-varying. So far, there are few applications of RL-based WAPS controls. In [16], RL is used to design PSSs. Since the control law used by real-time PSSs is learned through offline simulations first, its performance will degrade if the practical operating conditions deviate from the simulations. Besides, the communication delay is counteracted by using the existing local PSSs, but the interaction among different control modes might become complicated. Considering that such deviations are unavoidable, an adaptive RL-based PSS design is proposed in [17]. But the network imperfections are neglected in their work. To the best knowledge of the authors, there are none of the existing wide-area damping control solutions that considers both cyber- and physical- system uncertainties simultaneously.

In this paper, an adaptive RL-based wide-area damping control solution is proposed for WAPS. Both cyber- and physical- system uncertainties have been properly modeled and addressed during the control design. The considered cyber-system uncertainties include random communication delays and package dropouts [21]. The control difficulty due to physical system uncertainties is circumvented together with communication delays through online learning using inputs and outputs of the system. For control design, a simplified WAPS model with network imperfections is developed. To further counteract the impacts of network imperfections, the control law is designed based on statistical analysis using the expected successful package delivery rates. Finally, RL is introduced to overcome the physical system uncertainties and to optimize control performance under communication delays. The control solution is tested with both simplified and detailed WAPS models. Simulation results demonstrate the effectiveness of the proposed solution.

The remainder of this paper is organized as follows. Section II provides the background on WAPSs, network imperfections, and RL. Section III presents control model of WAPS and Q-learning and statistical analysis based control design under cyber and physical uncertainties. Section IV presents simulation results with simplified linear and detailed nonlinear 11-bus as well as 30-bus WAPS model, respectively. Finally, Section V concludes the paper.

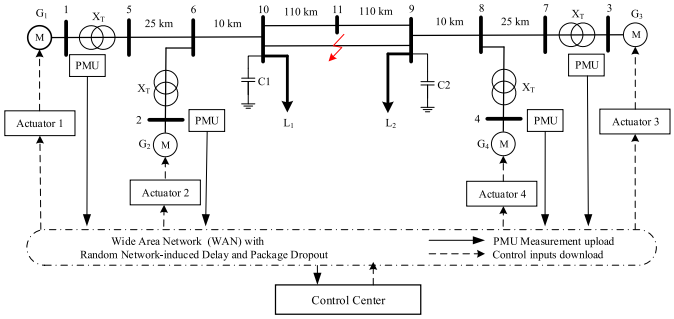


Fig. 1. Networked control of wide-area power systems.

II. Background

A. Wide-Area Power System

The interconnection of multiple subsystems in a WAPS brings about several benefits. Firstly, increase the reliability of the overall system. Whenever one of the subsystems is facing certain unrecovered contingency, it has better chance to survive or decrease the lost under supports of other inter-connected subsystem(s) [4]. Secondly, improve the energy efficiency of the overall system. Since the WAPSs enable the low-cost power supply transmitting from one subsystem to other inter-connected ones, the overall system operating cost can be optimized and reduced [5]. Thirdly, enhance the flexibility of the power grid. Because each subsystem retains the ability to work separately and autonomously, the inter-connected WAPS can be isolated into several subsystems under the extreme condition to avoid spreading of wide-area blackouts [6].

However, the effective closed-loop control of WAPSs is a long-term challenge due to the wide coverage area and the complex interactions among subsystems. Traditional power system stabilizers are designed using local information and tend to make myopic decisions without considering the impact from interconnected regions. The installation of PMUs and the deployment of WAMS bring about real-time global vision to the control center for wide-area monitoring and control. Accordingly, the damping control problem of WAPS can be formulated as a networked control problem.

The concept of the proposed networked control solution for WAPS is illustrated in the Fig. 1. Multiple PMUs are installed across the WAPS for gathering of synchronized measurements. The grid-wide PMU measurements are transmitted to the control center through the Wide-Area Network (WAN). At the control center, the control adjustments for inter-area damping is calculated based on the RL-based algorithm. The control signals are then transmitted through the WAN to the actuators of synchronous generators for control implementation. During both communication paths, random communication delays and package dropouts may occur.

B. Network Imperfections

The block diagram of WAPS under networked control in presence of network imperfections is illustrated in Fig. 2. Such a closed-loop system is referred to be a Networked Control System (NCS). In such NCS, network imperfections might occur when PMUs send the wide-area measurements

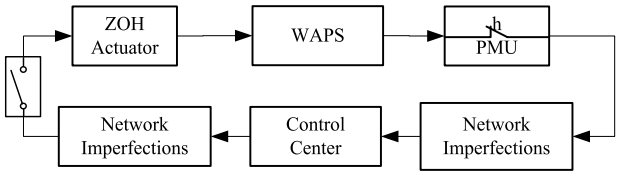


Fig. 2. Block diagram of NCS under network imperfections.

to the control center, and when control center sends the control signals back to actuators for control implementation [11]. Without considering the network imperfections, the control system may not even maintain stability under severe situations [12].

Due to the complexities of cyber uncertainties, it is impossible to consider all scenarios. This paper considers two types of most common network imperfections, random communication delay and package dropout [12]. For a discrete-time WAPS with a sampling period of h, assuming that PMUs outputs y_k at time step k take a time of $\tau_{pc,k}$ to be transmitted to the central controller. In addition, it takes a time of $\tau_{ca,k}$to be transmitted from the central controller to the actuators. Thus, the total communication delay during signal transmission is τ_k=$\tau_{pc,k}$+$\tau_{ca,k}$. According to the concept of Zero-Order Hold (ZOH) controller, the system under control keeps using previous control signal until the new control signal is successfully delivered to actuators. If τ_k<h, the delay is called short-term delay. Otherwise, the delay is called long-term delay [21]. Meanwhile, package dropout can be treated as an extreme scenario of communication delay. Two cases of delays can be classified as package dropouts. The first case is that the delay is larger than a certain threshold. The second case is that the newer control signal arrives before the older one. This paper only considers reasonable delays and package dropouts that can be effectively counteracted through control design. In another word, reasonable successful package delivery rate is required. Mathematically, the requirement can be described as: at least one among l consecutive packages is successfully delivered. It should be noted that short-term delays are normal and unavoidable, while long-term delays and package dropouts create major control difficulties.

The random communication delays and package dropouts in a discrete-time system are illustrated in Fig. 3, where the time for control calculation is neglected. In the figure, scenarios #1 and #4 are short-term delays, scenario #2 is a long term delay, and scenarios #3 and #5 are package dropouts.

C. Reinforcement Learning

Reinforcement learning (RL) includes a class of learning methods, which can approach optimal control iteratively through online learning without requiring system model. An actor involving in RL method usually interacts with system or environment and adjusts its behavior or policies to solve the optimization problems. A critic is responsible for executing the policy evaluation by assessing the results of applying current control signals into the system. Based on the assessment, the policy improvement is performed by the actor to yield an improved performance value compared to the previous one [18].

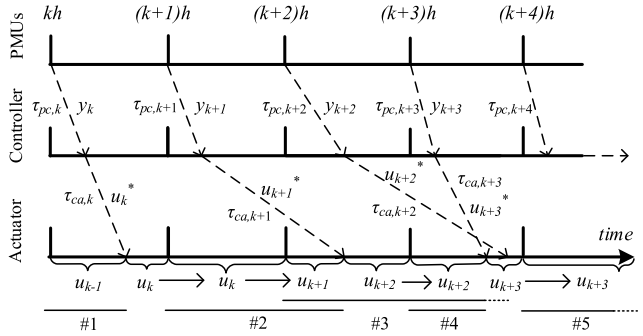


Fig. 3. Illustration of various communication delays and package dropouts.

As one of popular RL methods, Q-learning method is applied in this paper to solve the wide-area damping control problem of WAPSs. The optimal action-value function of the Q-learning method is defined in term of Bellman equation as [18],

$$Q_k^*(x,u) = E_\pi\left\{\left[r_k + \gamma V_{k+1}^*(x')\right] | x_k = x, u_k = u\right\} \quad (1)$$

where r_k is the stage cost, and $V_{k+1}^*(x') = \min_u Q_{k+1}^*(x',u')$ is the next-step value function. $0 \le \gamma \le 1$ is the discount rate and shows the importance of the future reward in the decision making process. The action-value function shows the expected return $E_\pi(\cdot)$ in state x and taking an action u under an arbitrary policy $\pi(x,u)$, and performing optimal policy $\pi^*(x,u)$ thereafter, until the optimal value of Q-function $Q_k^*(x,u)$ is reached [18]. In terms of the Q-function, the Bellman optimality equation has a simple form for a given state x, and selects the smallest value as

$$\begin{cases} V_k^*(x) = \min_u Q_k^*(x,u) \\ u_k^* = \operatorname{argmin}_u Q_k^*(x,u) \end{cases} \quad (2)$$

In the next section, the implementation of proposed Q-learning method on wide-area damping control of WAPSs is introduced in detail.

III. The Proposed Wide-Area Control Algorithm Under Cyber and Physical Uncertainties

A. Modeling of WAPSs Under Network Imperfections

According to [7] and [10], the WAPS model can be represented based on a linear small-signal model along the equilibrium point. This WPAS representation integrates the dynamics of synchronous generators (SGs), static loads, and transmission lines. By applying Kron reduction method, all of the buses get eliminated except for the generator internal buses. Thus, the original differential-algebraic-equation model can be transformed into a completed dynamic model as

$$\Delta\dot{x}(t) = A\Delta x(t) + B\Delta u(t) + D(t) \quad (3)$$

In (3), the deviations are defined as $\Delta x(t) = x(t) - x_0$ and $\Delta u(t) = u(t) - u_0$ where $x(t) = [\delta^T\ \omega^T\ E^T]^T$ is a column vector of the states of the WAPS with δ being a column vector of the rotor angle, ω being a column vector of the angular speed,

and E being a column vector of the internal generator voltage, respectively. $u(t) = [P_m{}^T \ E_f{}^T]^T$ is a column vector of the control signals to the WAPS with P_m being a column vector of the mechanical power outputs of the turbines and E_f being a column vector of the field voltage outputs of the exciters. $x_0 = [\delta_0^T \ \omega_0^T \ E_0^T]^T$ and $u_0 = [P_{m0}^T \ E_{f0}^T]^T$ are equilibrium operating points around where the linearization is performed. The matrices of $A \subset \mathbb{R}^{n\times n}$ and $B \subset \mathbb{R}^{n\times m}$ are dynamic matrixes based on the initial (equilibrium) states and parameters of the WAPS. $D(t) \subset \mathbb{R}^{n\times 1}$ denotes the time-varying disturbances of the system due to load fluctuation or other factors, and is assumed to be bounded as $\|D(t)\| \le D_M$ [19].

Since an SG's internal states ($\Delta\delta$ and ΔE) are not directly measurable, PMUs are usually deployed on generator terminal buses to measure the bus voltage V, voltage phase angle θ, and frequency f. According to [7], the following relationship exists between the PMUs measurements ($\Delta\theta$, Δf and ΔV) and SG's states ($\Delta\delta$, $\Delta\omega$, and ΔE).

$$\begin{aligned}\Delta y(t) &= \begin{bmatrix}\Delta V(t)\\ \Delta\theta(t)\\ \Delta f(t)\end{bmatrix} = \begin{bmatrix}F_1 & F_2 & 0\\ F_3 & F_4 & 0\\ 0 & 0 & \omega_s G\end{bmatrix}\begin{bmatrix}\Delta E(t)\\ \Delta\delta(t)\\ \Delta\omega(t)\end{bmatrix}\\ &= C\Delta x(t)\end{aligned} \tag{4}$$

where F_is and G are constant blocks calculated based on parameters and operating conditions of SGs. The detailed expressions of F_is and G are omitted in this paper due to space limit. Interested readers can refer to [7] and [20]. In the next subsection, the square matrix $C \subset \mathbb{R}^{n\times n}$ is learnt together with other system parameters through Q-learning. Because C is constant and invertible, the states $\Delta x(t)$ can be estimated according to $\Delta x(t) = C^{-1}\Delta y(t)$. Thus, the input-state model of (3) can be transformed into an input-output model as

$$\Delta\dot{y}(t) = A_y\Delta y(t) + B_y\Delta u(t) + D_y(t) \tag{5}$$

where $A_y = CAC^{-1} \subset \mathbb{R}^{n\times n}$, $B_y = CB \subset \mathbb{R}^{n\times m}$ and $D_y(t) = CD(t) \subset \mathbb{R}^{n\times 1}$.

The system is continuous, but control inputs are updated in discrete-time. In order to analyze the WAPS with network imperfections, the system of (5) is sampled with a period of h. Without loss of generality, consider the k^{th} interval of [hk, $h(k+1)$]. Initially, the previously applied control signal u_{k-1} is carried over. u_{k-1} is kept being applied until the possible successful delivery of the updated control signal u_k, which may or may not happen. By defining the overall delay for control signal delivery as τ_k, there will be control update if $\tau_k < h$. For example, control signal is u_{k-1} during [hk, $hk+\tau_k$] and the control signal is updated to u_k during [$hk+\tau_k$, $h(k+1)$]. For long term delays or package dropouts ($\tau_k > h$), there will be no update on control signals during the k^{th} interval. Therefore, the discrete-time control model under network imperfections can be formulated as

$$\Delta y_{k+1} = \Psi\Delta y_k + \Gamma_{0,\tau k}\Delta u_k + \Gamma_{1,\tau k}\Delta u_{k-1} + D_k \tag{6}$$

where $\Psi = e^{A_y h} \subset \mathbb{R}^{n\times n}$, $\Gamma_{0,\tau_k} = \int_0^{h-\tau_k}(e^{A_y s}B_y)ds \subset \mathbb{R}^{n\times m}$, $\Gamma_{1,\tau_k} = \int_{h-\tau_k}^{h}(e^{A_y s}B_y)ds \subset \mathbb{R}^{n\times m}$, $D_k = \int_0^h [e^{A_y s}D_y(t)]ds \subset \mathbb{R}^{n\times 1}$, $\Delta u_k = [\Delta P_{m,k}^T \ \Delta E_{f,k}^T]^T$ is the control adjustment at step k, and Δu_{k-1} is the previous value of Δu_k.

Since random network imperfections are unavoidable, it is impossible to counteract their impacts perfectly [21]. Previous studies on NCS suggest that such random network imperfections can be effectively handled through statistical analysis [12]. Assuming that τ_k obeys the normal distribution at time step k^{th}, i.e., $\tau_k(\mathbb{E}_\tau, \sigma_\tau)$, in which $\mathbb{E}_\tau$ is the expectation and σ_τ is the standard deviation of τ_k. The random control model of (6) can be transformed into a deterministic control model with Γ_{0,τ_k} being replaced by $\Gamma_{0,\mathbb{E}_\tau} = \int_0^{h-\mathbb{E}_\tau}(e^{A_y s}B_y)ds$ and Γ_{1,τ_k} being replaced by $\Gamma_{1,\mathbb{E}_\tau} = \int_{h-\mathbb{E}_\tau}^{h}(e^{A_y s}B_y)ds$.

In order to simplify the following Q-learning based control design, the control model is transformed into an auxiliary augment form as shown in (7)

$$\Delta z_{k+1} = A_d\Delta z_k + B_d\Delta u_k \tag{7}$$

where

$$\Delta z_k = \begin{bmatrix}\Delta y_k\\ \Delta u_{k-1}\\ 1\end{bmatrix}, \quad A_d = \begin{bmatrix}\Psi^{n\times n} & \Gamma_{1,\mathbb{E}_\tau}^{n\times m} & D_M^{n\times 1}\\ \mathbf{0}^{m\times n} & \mathbf{0}^{m\times m} & \mathbf{0}^{m\times 1}\\ \mathbf{0}^{1\times n} & \mathbf{0}^{1\times m} & 1^{1\times 1}\end{bmatrix},$$

$$B_d = \begin{bmatrix}\Gamma_{0,\mathbb{E}_\tau}^{n\times m}\\ I^{m\times m}\\ \mathbf{0}^{1\times m}\end{bmatrix},$$

$\mathbf{0}$ being a zero matrix, and I being an identity matrix.

According to the optimal control theory [18], the optimal control inputs for system (7) can be designed as

$$\Delta u_k^* = -\left(B_d^T P B_d + R\right)^{-1} B_d^T P A_d \Delta z_k \tag{8}$$

where $P \ge 0 \subset \mathbb{R}^{n\times n}$ is the Riccati equation solution [18] and $R \subset \mathbb{R}^{m\times m}$ is the positive-definite matrix.

Since the NCS suffers from network imperfections, the effective control signal (Δu_k) in (6) and (7) might be different from the current control signal calculated by the central controller (Δu_k^d). According to the assumption made in Section II-B, there is at least one successful control signal delivered in every l steps. The assumption can be formulated according to (9).

$$\begin{aligned}\Delta u_k &= p_k\Delta u_k^d + \left[(1-p_k)p_{k-1}\right]\Delta u_{k-1}^d + \cdots\\ &\quad + \left[(1-p_k)(1-p_{k-1})\cdots(1-p_{k-l+2})p_{k-l+1}\right]\Delta u_{k-l+1}^d\\ &= \sum_{i=1}^{l}\left[\prod_{j=1}^{i-1}\left(1-p_{k-j+1}\right)\right]p_{k-i+1}\Delta u_{k-i+1}^d\end{aligned} \tag{9}$$

In (9), p_k is an indicator, where $p_k = 1$ stands for successful information delivery, and $p_k = 0$ stands for unsuccessful information delivery. Δu_k^d is the control signal calculated in control center at the k_{th} time step considering communication delay.

Similar to the above problem formulation, statistical analysis is introduced again in control design to counteract the effect of network imperfections. If the occurrence probability of on-time package delivery $P(p_k)|_{\mathbb{E}_p,\sigma_p}$ obeys the normal

distribution in which $\mathbb{E}_p$ is the expectation of $P(p_k)$ and σ_p is the standard deviation, the designed control inputs u_k^d can be modified as [25]

$$\begin{aligned}\Delta u_k^d &= \frac{1}{\mathbb{E}_p}\Delta u_k^* - \left[\left(1-\mathbb{E}_p\right)\mathbb{E}_p\right]\Delta u_{k-1}^* - \cdots \\ &\quad - \left[\left(1-\mathbb{E}_p\right)^{l-1}\mathbb{E}_p\right]\Delta u_{k-l+1}^* \\ &= \frac{1}{\mathbb{E}_p}\left[\Delta u_k^* - \sum_{i=2}^{l}\left(1-\mathbb{E}_p\right)^{i-1}\mathbb{E}_p\Delta u_{k-i+1}^*\right] \end{aligned} \tag{10}$$

In (10), Δu_k^* and its previous values are calculated according to the Q-learning algorithm which is introduced in the next subsection. The model-free algorithm does not require physical system model, neither the statistical data on network imperfections. However, there is no guarantee that the calculated control signal Δu_k^* is effectively delivered. To further counteract the cyber uncertainties between the control center and the actuators, the instantaneous control signals Δu_k^*s calculated through Q-learning are weighted together in (10). The introduction of historic control signals can effectively counteract the previous impact of network imperfections on control signal delivery. In this way, the long term expectation on control performance can be expected to approach that under ideal conditions.

B. Q-Function Setup for Control Design

The control design in (8) requires known system dynamics to solve the Riccati equation as well as the optimal control inputs. To relax the requirements, Q-learning method is applied to solve the optimal control problem. An infinite-horizon value function [18] can be defined with Riccati equation solution P as quadratic in the state,

$$V_{k+1}^*(\Delta z_{k+1}) = \Delta z_{k+1}^T P \Delta z_{k+1} \tag{11}$$

The Bellman equation can be formulated accordingly as

$$Q^*\left(\Delta z_k, \Delta u_k^*\right) = \Delta z_k^T \Upsilon \Delta z_k + \Delta u_k^{*T} R \Delta u_k^* + V_{k+1}^*(\Delta z_{k+1}) \tag{12}$$

where $\Upsilon \subset \mathbb{R}^{n\times n}$ and $R \subset \mathbb{R}^{m\times m}$ are positive-definite cost matrixes.

By substituting (7) and (11) into the Bellman equation of (12), one can get

$$\begin{aligned} Q^*\left(\Delta z_k, \Delta u_k^*\right) &= \Delta z_k^T \Upsilon \Delta z_k + \Delta u_k^{*T} R \Delta u_k^* \\ &\quad + \left(A_d \Delta z_k + B_d \Delta u_k^*\right)^T P \left(A_d \Delta z_k + B_d \Delta u_k^*\right) \\ &= \begin{bmatrix} \Delta z_k \\ \Delta u_k^* \end{bmatrix}^T \begin{bmatrix} A_d^T P A_d + \Upsilon & B_d^T P A_d \\ A_d^T P B_d & B_d^T P A_d + R \end{bmatrix} \begin{bmatrix} \Delta z_k \\ \Delta u_k^* \end{bmatrix} \end{aligned} \tag{13}$$

Introduce a kernel matrix $H \in \mathbb{R}^{q\times q}$ as

$$H = \begin{bmatrix} A_d^T P A_d + \Upsilon & A_d^T P B_d \\ B_d^T P A_d & B_d^T P B_d + R \end{bmatrix} = \begin{bmatrix} H_{zz} & H_{zu} \\ H_{uz} & H_{uu} \end{bmatrix} \tag{14}$$

Then the Bellman equation of (13) can be rewritten as

$$Q^*\left(\Delta z_k, \Delta u_k^*\right) = \begin{bmatrix} \Delta z_k \\ \Delta u_k^* \end{bmatrix}^T H \begin{bmatrix} \Delta z_k \\ \Delta u_k^* \end{bmatrix} \tag{15}$$

Recall to (8), it yields the policy

$$\Delta u_k^* = -H_{uu}^{-1} H_{uz} \Delta z_k \tag{16}$$

Therefore, the optimal control gain can be obtained more conveniently by solving the kernel matrix H in (15) instead of using system dynamics A_d and B_d in (13). The brief derivation of numerical computing method to realize the online updating of H is given in the next subsection.

C. Model-Free Online Tuning of Control

To learn the kernel matrix H and approximate the optimal control signals, the Recursive Least Square (RLS) method is used in this paper [22]. First, one set of parametric structures $\hat{Q}(\bar{z}, h_i)$ and $\hat{u}_i(z)$ are selected to approximate the actual Q^* and u^* at the i^{th} iteration, respectively. It should be noted that i is the iteration step which is different from the time step k.

$$\hat{Q}_i(\Delta\bar{z}, h_i) = \Delta Z^T H_i \Delta Z = h_i^T \Delta\bar{z} \tag{17}$$

$$\Delta\hat{u}_i(\Delta z) = -H_{uu,i}^{-1} H_{uz,i} \Delta z \tag{18}$$

where $\Delta Z = \begin{bmatrix} \Delta z^T & \Delta u^T \end{bmatrix}^T \in \mathbb{R}^q$, $\Delta\bar{z} = (\Delta Z_1^2, \ldots, \Delta Z_1 \Delta Z_q, \Delta Z_2^2, \Delta Z_2 \Delta Z_3, \ldots, Z_{q-1} Z_q, Z_q^2) \in \mathbb{R}^{q(q+1)/2}$ is the Kronecker product quadratic polynomial basis vector; $h = v(H)$ with $v(\cdot)$ being a vector function which is constructed by stacking the columns of the squared matrix into a one-column vector. Therefore, $v(\cdot)$ transforms a $q \times q$ matrix H into a $q(q+1)/2 \times 1$ vector h, and H can be solved as $H = v^{-1}(h)$ [23].

Then, to find h_{i+1}, the desired target function is written as

$$\begin{aligned} d(\bar{z}_k, h_i) &= \Delta z_k^T \gamma \Delta z_k + \Delta\hat{u}_i^T(\Delta z_k) R \Delta\hat{u}_i(\Delta z_k) \\ &\quad + Q_i\left(\Delta z_{k+1}, \Delta\hat{u}_{i+1}(\Delta z_{k+1})\right) \end{aligned} \tag{19}$$

Due to the approximate error, the Bellman equation in (15) does not hold anymore. Then it can be defined as $h_{i+1}^T \Delta\bar{z}_k = d(\bar{z}_k, h_i) + e_k$, where e_k is the Bellman equation error. Therefore, h_{i+1} is found over a compact set Ω to minimize the Bellman equation error e_k in a least-square sense as

$$h_{i+1} = \arg\min_{h_{i+1}} \left\{ \int_\Omega \left| h_{i+1}^T \Delta\bar{z}_k - d(\bar{z}_k, h_i) \right|^2 d\Delta z_k \right\} \tag{20}$$

Then it can be solved as

$$h_{i+1} = \left(\int_\Omega \Delta\bar{z}_k \Delta\bar{z}_k^T d\Delta z_k \right)^{-1} \int_\Omega \Delta\bar{z}_k d(\bar{z}_k, h_i) d\Delta z_k \tag{21}$$

The kernel matrix H stops updating when Bellman equation error is less than certain small valve ε, e.g., $\|h_{i+1} - h_i\| < \varepsilon$. References [22] and [23]. If the Bellman equation error e_k becomes zero, the system optimality is achieved. It should be mentioned that the quantity of the historic information (i.e., compact set Ω) required to solve (21) is dependent on the system dimension, i.e., the number of PMUs used in the NCS. The overall procedure of the proposed Q-learning based algorithm for wide-area control is summarized in Table I. One advantage of the algorithm is that it makes incremental

TABLE I
IMPLEMENTATION OF THE Q-LEARNING BASED CONTROL ALGORITHM

1) Initialize the system state set Δz, control action Δu, value function $Q(\Delta z, \Delta u)$, and kernel matrix H,
2) **for** $i = 0$ to a given number iterations,
   if $\|h_{i+1} - h_i\| > \varepsilon$, **do**
   2.1) Observe the state Δz of the WAPS
   2.2) Calculate the least squares to solve the kernel matrix H_{i+1} (21)
   2.3) Perform policy update to approximate the optimal control signal $\Delta\hat{u}_i(\Delta z)$ (18)
   2.4) Implement the control signal and observe the state $\Delta z'$ (7)
   2.5) Update the Q-function (17)
3) **end**

improvements and the convergence of learning can be theoretically proved based on the mild assumptions as introduced in the paper. This means that it can avoid the abrupt/risky control adjustments while performance improvement can be guaranteed even before the optimal solution is found.

IV. SIMULATION STUDIES

It is a common sense that the practical WAPS model is complicated and highly nonlinear. Due to the complexity of the physical system and cyber uncertainties, it is impossible to model such a large system precisely. To solve the wide-area damping control problem effectively, a model-free control algorithm is proposed based on the Q-learning technique. In order to evaluate its real-world performance, the linear model based control algorithm is tested with variable types of simplified linearized and detailed nonlinear WAPS models, i.e., IEEE 11-bus and 30-bus models [24].

To select the next control signal Δu_k replacing the currently effective control signal Δu_{k-1}, each of the most recent l successive control signals (Δu_k^d, Δu_{k-1}^d, ..., Δu_{k-l+1}^d) is assigned with a random probability $P|_{(\mathbb{E}_p,\sigma_p)} \subset [0, 1]$ with expectation being $\mathbb{E}_p$ and deviation being σ_p. If the random probability P is larger than $\mathbb{E}_p$, the corresponding control signal will be successfully delivered (p=1), otherwise, the control signal will not be delivered in time due to network imperfections (p=0). The search of Δu_k starts from Δu_k^d to Δu_{k-l+1}^d. During the searching process, the first control signal with nonzero p will be selected as Δu_k. If none of the l control signals is delivered and the situation has already happened for l-1 times, one of the l signals will be randomly selected. This is because of the assumption that there is at least one successful control signal delivery in every l steps.

To decide the instant of time when the new control signal Δu_kwill be deployed, similar statistical analysis is employed. A random number within the range of [0, h] is generated to decide τ_k. The random number has an expectation of $\mathbb{E}_\tau$ and deviation of σ_τ. Once the τ_k is determined, the Δu_k will be deployed at the instant of $hk+\tau_k$ at the k^{th} step.

During simulation, h is set to 0.1s, l is set to 3, $\mathbb{E}_p$ is set to 0.9, σ_p is set to 0.05, $\mathbb{E}_\tau$ is set to 0.75h, and σ_τ is set to 0.2h. It should be noted that $\mathbb{E}_p$ is intentionally set to a small value (frequent network imperfection) to challenge the proposed control algorithm.

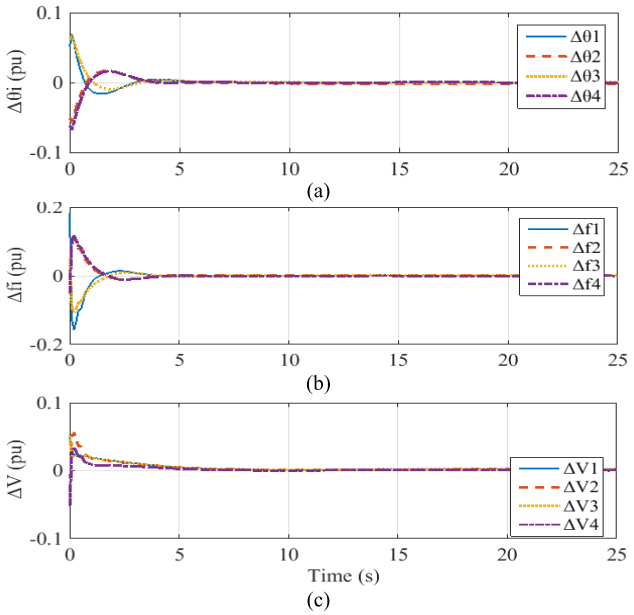


Fig. 4. Simulation results of the proposed Q-learning controller with linearized 11-bus model: (a). Phase angle deviation ($\Delta\theta$); (b). Frequency deviations (Δf); (c). Terminal bus voltage deviations (ΔV).

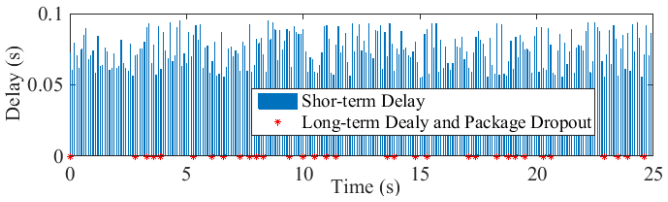


Fig. 5. Distribution of communication delay and package dropout.

A. Simulation With Linearized 11-Bus Model

The initial large deviations are set up to 15% of the equilibrium operating points. In addition, up to 5% disturbances (D_k) are simulated to investigate the effectiveness of the designed algorithm under small and continuous system disturbances. The damping performance of the proposed algorithm under aforementioned conditions are shown in Fig. 4. Among the three plots, Fig. 4(a) shows the responses of bus voltage phase angle deviation ($\Delta\theta$), Fig. 4(b) shows the responses of frequency deviation (Δf), and Fig. 4(c) shows the responses of terminal bus voltage deviation (ΔV), all in per-unit values. The distribution of communication delay and package dropout during the simulation is shown in the Fig. 5. It can be seen that the wide-area oscillations get damped effectively by the proposed control algorithm under both cyber and physical uncertainties. Besides, the robustness of the proposed algorithm against the network imperfections is also demonstrated.

The control signal responses are presented in Fig. 6. Because the control signals of the four subsystems are quite similar for small signal disturbance, only control inputs (ΔE_{f1} and ΔP_{m1}) of subsystem #1 are presented. Due to the continuous system disturbances, it can be seen that the control adjustments oscillate around zero. As illustrated in the Fig. 7,

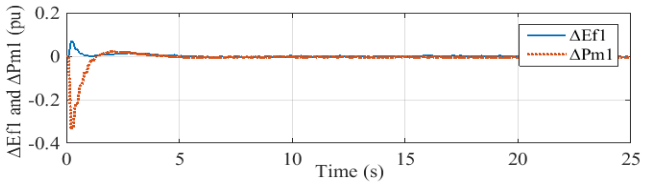


Fig. 6. Wide-area damping control signals of generator #1 (adjustments of field voltage ΔE_{f1} and mechanical power input ΔP_{m1}).

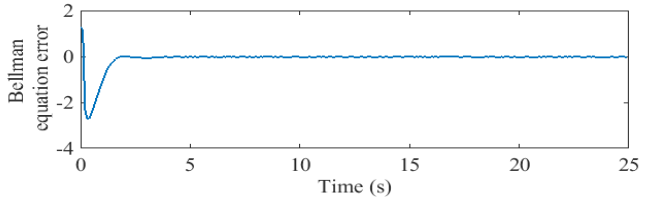


Fig. 7. Bellman equation error e_k of 11-bus WAPS.

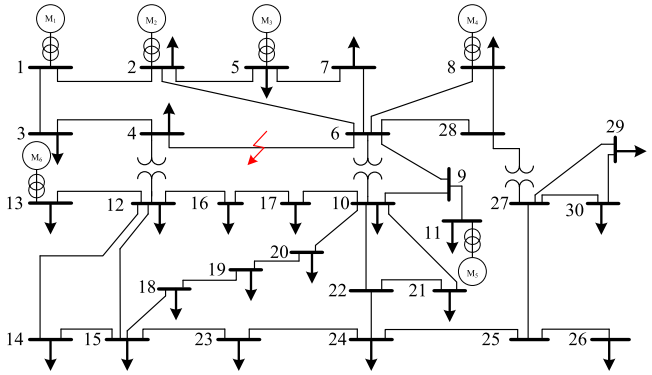


Fig. 8. IEEE 30-bus WAPS.

the Bellman equation error e_k converges to a small neighborhood of zero, which demonstrates achievement of the system optimality.

B. Simulation With Linearized 30-Bus Model

In order to further test evaluate the effectiveness of the proposed control algorithm on variable types of WAPS, a similar simulation is conducted on a linearized IEEE 30-bus system as shown in Fig. 8 [26]. The parameters of an IEEE 30-bus system can be found online in [26].

The damping performance of the proposed algorithm under aforementioned conditions are shown in Fig. 9. Among the three plots, Fig. 9(a) shows the response of bus voltage phase angle deviation ($\Delta\theta$), Fig. 9(b) shows the response of frequency deviation (Δf), and Fig. 9(c) shows the response of terminal bus voltage deviation (ΔV), all in per-unit values. The Bellman equation error e_k is shown in the Fig. 10. As can be observed, the proposed control algorithm can effectively damp the oscillation for a more complicated WAPS under network imperfections. Moreover, the proposed Q-learning is a novel online time-based learning technique. Compared with conventional offline policy or value iteration schemes, the time-based learning techniques are updating along with time

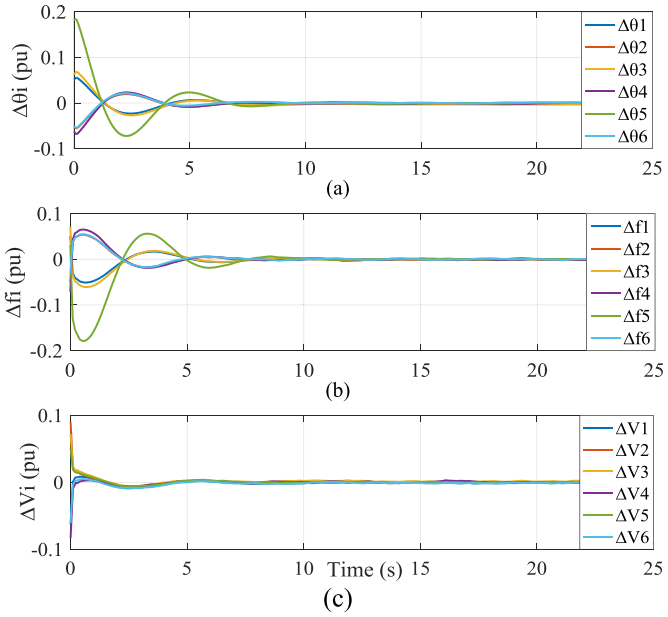


Fig. 9. Simulation results of the proposed Q-learning controller with linearized 30-bus model: (a). Phase angle deviation ($\Delta\theta$); (b). Frequency deviations (Δf); (c). Terminal bus voltage deviations (ΔV).

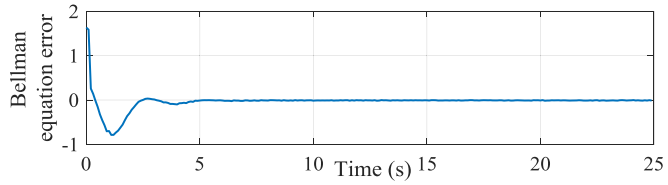


Fig. 10. Bellman equation error e_k of IEEE 30-bus WAPS.

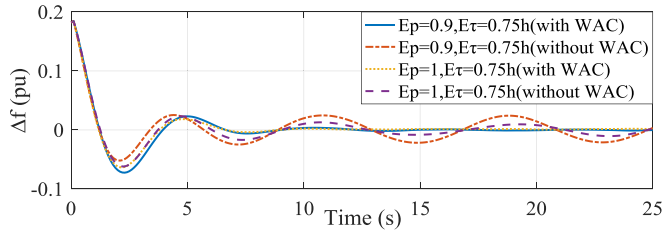


Fig. 11. Frequency responses under different network imperfections with and without design considerations.

online instead of iteration offline, which have been demonstrated as an effective method to overcome the issue from high computational complexity [28].

In addition, to better illustrate the influence from the network imperfections and to demonstrate the effectiveness of the proposed controller against this problem, the comparison results of frequency responses are shown in Fig. 11. The simulation is conducted on an IEEE 30-bus system with package dropout ($\mathbb{E}_p = 0.9$) and without package dropout ($\mathbb{E}_p = 1$) under the same communication delay ($\mathbb{E}_\tau = 0.75h$). It can be observed that the network imperfections have a tiny influence on damping performance of the proposed controller. However, the communication network quality, i.e., package dropout and communication delay, if not properly considered, can result low frequency oscillations in WAPS.

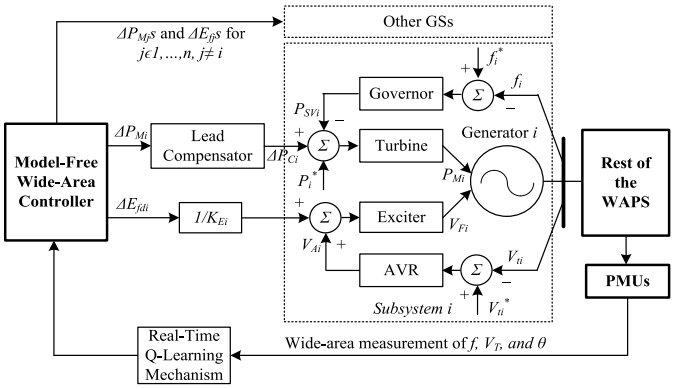


Fig. 12. Implementation of the proposed control algorithm for WAPSs.

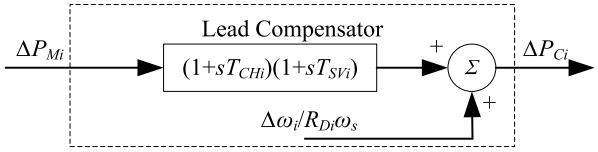


Fig. 13. Lead compensator for the turbine.

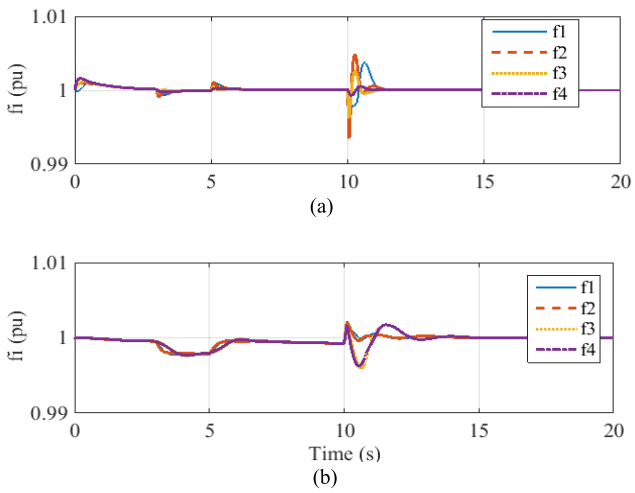


Fig. 14. Frequency responses (f) of 11-bus WAPS under system disturbances, (a). The proposed Q-learning networked controller; (b). CPSS.

C. Simulation With Nonlinear Detailed 11-Bus Model

In this section, the detailed nonlinear WAPS model is used to evaluate the performance of the proposed control algorithm under the practical application. The simulation studies are conducted on the Simulink module that comes with the SimPowerSystem™ toolbox for inter-area oscillations [27]. The specific targets are focused on testing the algorithm performance under large system disturbances and network imperfections. In addition, the control performance is also compared against well-tuned CPSSs. The simulation settings of network imperfections are same as that in linear-model-based simulations in Section V-A. The implementation of the proposed control algorithm with the detailed WAPSs model is illustrated in Fig. 12.

It is known that two control signals ΔE_{fi} and ΔP_{mi} (physical quantities) have to be realized by the physical components (exciters and turbines). Because the electrical controller (exciter) has a fast response speed and the inertial delay can be ignored, only a gain $1/K_{Ei}$ is added after the filed voltage adjustment (ΔE_{fi}) [24]. However, the inertial of mechanical controller (turbine) introduces significant delay to ΔP_{mi}. Thus, a simple lead compensator is designed based on the turbine dynamics to counteract the delay. The linearized dynamics of a non-reheat steam turbine can be represented using two-order functions as [20]

$$\Delta\dot{P}_{mi}(t) = \frac{1}{T_{CHi}}[-\Delta P_{Mi} + \Delta P_{SVi}] \tag{22}$$

$$\Delta\dot{P}_{SVi}(t) = \frac{1}{T_{SVi}}\left[-\Delta P_{SVMi} + \Delta P_{Ci} - \Delta\omega_i/R_{Di}\omega_s\right] \tag{23}$$

where R_{Di} is the speed regulation quantity; P_{Ci} is the control input of the governor, namely the power change setting; P_{SVi} is the steam valve position; T_{SVi} and T_{CHi} are the reheater main inlet volumes and steam chest time constants, respectively. Based on the dynamics in (22) and (23), a compensator can be designed to counteract the delay of the turbine response as shown in Fig. 13. The parameters of R_D, T_{SV} and T_{CH} are set to be 10, 0.001and 0.015 in the simulation, respectively.

Because the generator internal voltages cannot be easily acquired, the generator terminal voltage measurements are used instead as an input to the wide-area controller. A series of different operating conditions are simulated in the study. Initially, the system is running stably. At time 3 second, a 200 MW step load increasing is performed at bus # 10 (from 967 MW to 1167 MW). At time 5 second, the load at bus #10 returns back to its original value (967 MW). At time 10 second, a three-phase short circuit is performed at the transmission line disconnecting bus #9 and bus #10. After 6 cycles, the fault is removed by opening the two circuit breakers located at the both ends of the transmission line, as presented in Fig. 1. The total simulation time is set to 20 seconds.

In order to fully understand the performance of the proposed model-free control algorithm, it is compared with CPSSs which are simulated under the same power system scenarios. In the simulation, four well-tuned CPSSs from [24] are deployed at four generators respectively. It should be noted that the problem of network imperfections is not considered in CPSSs since they are local controllers.

The simulation results of frequency responses (f) under aforementioned circumstances are shown in Fig. 14, among which Fig. 14(a) presents the control performance using the proposed Q-learning networked controller and Fig. 14(b) is the damping result under CPSSs (without using the proposed controller), all in per units. Similarly, the terminal voltages responses (V) are shown in Fig. 15, among which Fig. 15(a) presents the control performance using the proposed model-free networked controller and Fig. 15(b) is the damping result under CPSSs (without using the proposed controller), all in per units.

As introduced before, the proposed Q-learning method uses a linear WAPS model to approximate the optimal control of

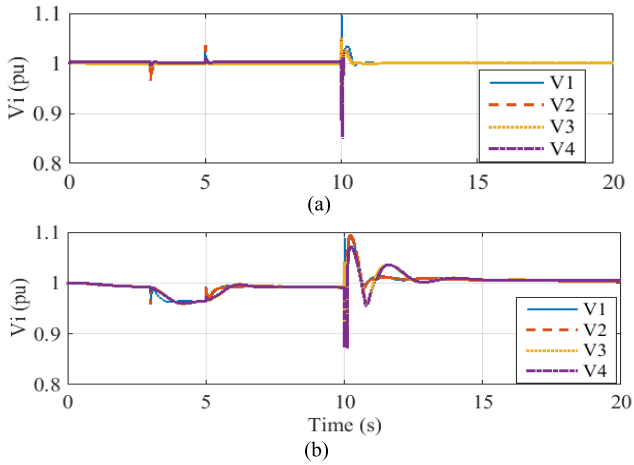


Fig. 15. Terminal voltages responses (V) under system disturbances, (a). The proposed Q-learning networked controller; (b). CPSS.

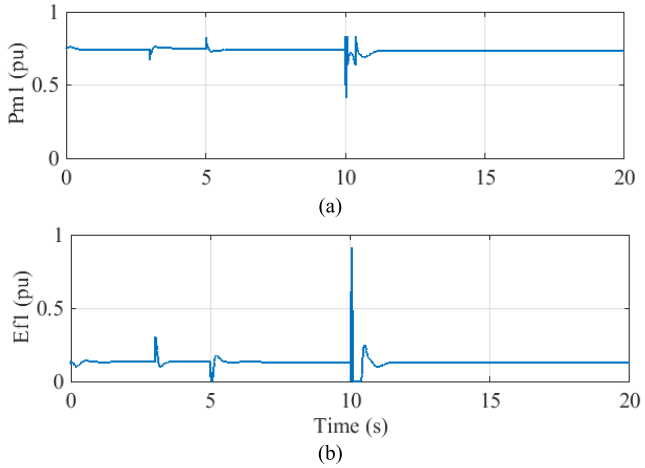


Fig. 16. Overall control inputs to generator #1, (a). Mechanical input power P_{m1}; (b). Field voltage E_{f1}.

a nonlinear system. However, it can be observed from the Fig. 14(a) and Fig. 15(a) that the possible measurement deviations from a nonlinear model do not degrade the performance of the linear-based control design. It is because that the system dynamics with disturbances are able to be learned and updated online. Even if the CPSSs shown in Fig. 14(b) and Fig. 15(b) can also stabilize system under load change and fault situation, it takes 5 seconds to damp out oscillations caused by the line fault. While the proposed controller only takes less than 3 seconds to realize the system stabilization. It should be noted that the voltage drop in Fig. 15(b), is not a steady state error. It is because CPSSs cannot recover the low order frequency oscillations in a timely manner. If given sufficient time, the deviation under CPSSs will disappear. Because frequency f and voltage V of the designed control method are adjusted at the same time through coordination, the transient performance is much better than that under conventional controllers.

The overall control inputs (P_{m1} and E_{f1}) to generator #1 are shown in Fig. 16. For protection purpose, ramp rates and bounds are applied to the overall control inputs. Although the practical control constraints further increases system uncertainties, control performance is not degraded.

D. Simulation With Nonlinear Detailed 30-Bus Model

In this subsection, similar simulations are performed on a detailed IEEE 30-bus model as presented in Fig. 8. The performance of the proposed wide-area controller is compared with the behavior of CPSSs (without using the proposed controller). In order to make the paper condensed, only a three-phase self-clean fault is conducted between bus #4 and bus #6. The total simulation time is 15s and the fault happens at the 2s during the steady state. After 3 cycles, the fault is removed by opening the two circuit breakers located at the both ends of the transmission line. The frequency responses of using the proposed wide-area controller and the CPSSs are shown in the Fig. 17(a) and Fig. 17(b), respectively. As can be observed, the proposed Q-learning controller can achieve the wide-area oscillation damping within 4s, while CPSSs have to take more than 8s to damp out the oscillations. Actually, the

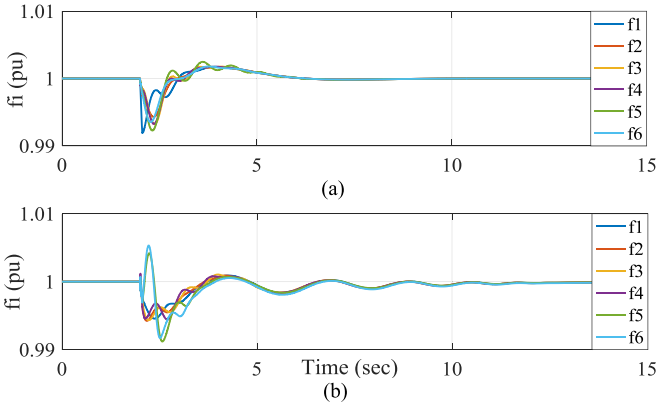


Fig. 17. Frequency responses (f) of IEEE 30-bus WAPS under system disturbances, (a). The proposed Q-learning networked controller; (b). CPSS.

tuning process of CPSSs is full of difficulties, and the low-frequency oscillation is a common phenomenon when system operating condition changes. One major reason is because the CPSSs only use the local measurements. Therefore, WAMS has created a great opportunity to better manage WAPSs.

Above simulation results demonstrate the effectiveness of the proposed wide-area damping control algorithm. Usually, RL method requires significant time to obtain the desired knowledge, and the learning from scratch is difficult as shown in our previous paper [29]. In this work, once the initial knowledge is obtained, the incremental online learning afterward will be converged much faster. So the learning speed is not a problem since the utility can afford the cost of capable infrastructure. Besides, even though the design is linear-based, the obtained performance is not compromised. Practical control applications prefer such simple yet effective algorithms.

V. Conclusion

Both cyber and physical uncertainties in WAPS are challenging the operation and control of WAPS. Currently, PMUs are mainly used for open-loop wide-area monitoring instead of closed-loop wide-area control. Targeting at wide-area damping

control of WAPS under both physical and cyber uncertainties, a learning-based networked control algorithm is presented in this paper. The problems of network imperfections and physical uncertainties are properly addressed through the model-free Q-learning based control algorithm. Besides, statistical analysis based control law is also designed to further counteract the cyber uncertainties. The proposed algorithm is evaluated through simulation with both linearized and detailed nonlinear models of WAPS. Simulation results demonstrate the effectiveness of the algorithm.

Even though Q-learning can improve the adaptivity of the linear model based control algorithm through online learning, the control performance cannot beat those sophisticated nonlinear model based control designs. Developing simple yet effective nonlinear models of WAPS for control design is a challenging problem. In addition, an advanced experimental setup should be used to well evaluate the control algorithms for cyber-physical systems. It is also desirable to emulate the actual delays and dropouts instead of merely emulating the randomness at actuator side. Future research should investigate the relationship between network imperfections and system stability and transient performance. Overall, there are still many open problems to investigate with control of cyber-physical WAPS. The research requires close collaboration among researchers of multiple disciplines.

REFERENCES

[1] C. Lu, Y. Zhao, K. Men, L. Tu, and Y. Han, "Wide-area power system stabiliser based on model-free adaptive control," *IET Control Theory Appl.*, vol. 9, no. 13, pp. 1996–2007, Aug. 2015.

[2] C. W. Taylor, D. C. Erickson, K. E. Martin, R. E. Wilson, and V. Venkatasubramanian, "WACS-wide-area stability and voltage control system: R&D and online demonstration," *Proc. IEEE*, vol. 93, no. 5, pp. 892–906, May 2005.

[3] A. Chakrabortty, "Wide-area damping control of power systems using dynamic clustering and TCSC-based redesigns," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1503–1514, Sep. 2012.

[4] J. Ma, T. Wang, Z. Wang, J. Wu, and J. S. Thorp, "Design of global power systems stabilizer to damp interarea oscillations based on wide-area collocated control technique," in *Proc. IEEE Power Energy Soc. General Meeting*, San Diego, CA, USA, 2011, pp. 1–7.

[5] A. Chakrabortty, J. H. Chow, and A. Salazar, "A measurement-based framework for dynamic equivalencing of large power systems using wide-area phasor measurements," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 68–81, Mar. 2011.

[6] F. Dörler, M. R. Jovanović, M. Chertkov, and F. Bullo, "Sparsity-promoting optimal wide-area control of power networks," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2281–2291, Sep. 2014.

[7] A. Chakrabortty and P. P. Khargonekar, "Introduction to wide-area monitoring and control," in *Proc. Amer. Control Conf. (ACC)*, Washington, DC, USA, 2013, pp. 6758–6770.

[8] N. Kishor, L. Haarla, J. Turunen, M. Larsson, and S. R. Mohanty, "Controller design with model identification approach in wide area power system," *IET Gener. Transm. Distrib.*, vol. 8, no. 8, pp. 1430–1443, Aug. 2014.

[9] J. B. Zhang, C. Y. Chung, and Y. D. Han, "A novel modal decomposition control and its application to PSS design for damping interarea oscillations in power systems," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2015–2025, Nov. 2012.

[10] T. Zabaiou, F. A. Okou, L.-A. Dessaint, and O. Akhrif, "Time-delay compensation of a wide-area measurements-based hierarchical voltage and speed regulator," *Can. J. Elect. Comput. Eng.*, vol. 33, no. 2, pp. 77–85, Sep. 2008.

[11] S. Wang, X. Meng, and T. Chen, "Wide-area control of power systems through delayed network communication," *IEEE Trans. Control Syst. Technol.*, vol. 20, no. 2, pp. 495–503, Mar. 2011.

[12] C. Lu, X. Zhang, X. Wang, and Y. Han, "Mathematical expectation modeling of wide-area controlled power systems with stochastic time delay," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1511–1519, May 2015.

[13] B. Chaudhuri, R. Majumder, and B. C. Pal, "Wide-area measurement-based stabilizing control of power system considering signal transmission delay," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1971–1979, Nov. 2004.

[14] W. Zhang, W. Liu, X. Wang, L. Liu, and F. Ferrese, "Distributed multiple agent system based online optimal reactive power control for smart grids," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2421–2431, Sep. 2014.

[15] S. Wang, W. Gao, J. Wang, and J. Lin, "Synchronized sampling technology-based compensation for network effects in WAMS communication," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 837–845, Jun. 2012.

[16] R. Hadidi and B. Jeyasurya, "Reinforcement learning based real-time wide-area stabilizing control agents to enhance power system stability," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 489–497, Mar. 2013.

[17] C. Lu, Y. Zhao, K. Men, L. Tu, and Y. Han, "Wide-area power system stabiliser based on model-free adaptive control," *IET Control Theory Appl.*, vol. 9, no. 13, pp. 1996–2007, Aug. 2015.

[18] F. L. Lewis, D. Vrabie, and V. Syrmos, *Optimal Control*, 3rd ed. New York, NY, USA: Wiley, 2012.

[19] F. L. Lewis, S. Jagannathan, and A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*. London, U.K.: Taylor & Francis, 1999.

[20] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*, 2nd ed. Piscataway, NY, USA: Wiley, 2002.

[21] W. Zhang, M. S. Branicky, and S. M. Philips, "Stability of networked control systems," *IEEE Control Syst. Mag.*, vol. 21, no. 1, pp. 84–99, Feb. 2001.

[22] P. J. G. Teunissen, *Dynamic Data Processing: Recursive Least-Squares*. Delft, The Netherlands: Delft Univ. Press, 2001.

[23] A. Tl-Asma, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to $\mathcal{H}$-infinity control," *Automatica*, vol. 43, no. 3, pp. 473–481, 2007.

[24] P. Kundur, *Power Systems Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.

[25] J. L. Myers, A. Well, and R. F. Lorch, *Research Design and Statistical Analysis*, 3rd ed. New York, NY, USA: Routledge, 2010.

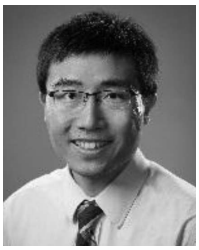
[26] *Power Systems Test Case Archive*. Accessed on Aug. 2014. [Online]. Available: https://www2.ee.washington.edu/research/pstca/pf30/pg_tca30bus.htm

[27] *Performance of Three PSS for Inter-Area Oscillations*. Accessed on Sep. 2014. [Online]. Available: http://www.mathworks.com/help/physmod/sps/examples/performance-of-three-pss-for-interarea-oscillations.html

[28] H. Xu, Q. Zhao, and S. Jagannathan, "Finite-horizon near-optimal output feedback neural network control of quantized nonlinear discrete-time systems with input constraint," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1776–1788, Aug. 2015.

[29] Y. Xu, W. Zhang, W. Liu, and F. Ferrese, "Multiagent-based reinforcement learning for optimal reactive power dispatch," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1742–1751, Nov. 2012.

Jiajun Duan (S'14) was born in Lanzhou, China, in 1990. He received the B.S. degree in power system and its automation from Sichuan University, Chengdu, China, in 2013, and the M.S. degree in electrical engineering from Lehigh University, Bethlehem, PA, USA, in 2015, where he is currently pursuing the Ph.D. degree. His research interest includes power system, power electronics, control systems, renewable energy, and smart microgrid.



Hao Xu (M'12) was born in Nanjing, China. He received the master's degree in electrical engineering from Southeast University, Nanjing, in 2009, and the Ph.D. degree from the Missouri University of Science and Technology, Rolla, MO, USA, in 2012. He was with Texas A&M University–Corpus Christi, Corpus Christi, TX, USA, as an Assistant Professor with the College of Science and Engineering. He is currently with the University of Nevada, Reno, where he is an Assistant Professor with the Department of Electrical and Biomedical Engineering. His current research interests include intelligent control design for advanced power systems, smart grid, autonomous unmanned aircraft systems, wireless passive sensor network, localization, detection, networked control system, cyber-physical system, distributed network protocol design, optimal control, and adaptive control.

Wenxin Liu (S'01–M'05–SM'14) received the B.S. degree in industrial automation and the M.S. degree in control theory and applications from Northeastern University, Shenyang, China, in 1996 and 2000, respectively, and the Ph.D. degree in electrical engineering from the Missouri University of Science and Technology (formerly, University of Missouri–Rolla), Rolla, MO, USA, in 2005. Then, he was an Assistant Scholar Scientist with the Center for Advanced Power Systems, Florida State University, Tallahassee, FL, USA, until 2009. From 2009 to 2014, he was an Assistant Professor with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM, USA. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, USA. His research interests include power systems, power electronics, and controls. He is an Editor of the IEEE TRANSACTIONS ON SMART GRID and a Guest Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.