Summer 2012

Stochastic optimal adaptive controller and communication protocol design for networked control systems

Hao Xu

Department: Electrical and Computer Engineering

STOCHASTIC OPTIMAL ADAPTIVE CONTROLLER AND COMMUNICATION PROTOCOL DESIGN FOR NETWORKED CONTROL SYSTEMS

by

HAO XU

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE & TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2012

Approved

Jagannathan Sarangapani, Advisor
S.N. Balakrishnan
Sanjay Madria
Maciej Zawodniok
Yahong Rosa Zheng

PUBLICATION DISSERTATION OPTION

This dissertation contains the following four articles:

Paper 1, H. Xu, S. Jagannathan, and F. L. Lewis, "Stochastic Optimal Control of Unknown Linear Networked Control System in the Presence of Random Delays and Packet Losses," In Press, Automatica.

Paper 2, H. Xu and S. Jagannathan, "Stochastic Optimal Design for Unknown Linear Discrete-time System Zero-sum Games under Communication Constraints," has been submitted to Automatica.

Paper 3, H. Xu, and S. Jagannathan, "Stochastic Optimal Controller Design for Uncertain Nonlinear Networked Control System via Neuro Dynamic Programming," conditional accepted in the IEEE Transaction on Neural Networks.

Paper 4, H. Xu and S. Jagannathan, "Stochastic Optimal Design for Unknown Networked Control System using Communication Network Protocols," to be submitted to Automatica.

Paper 5, Hao Xu, and S. Jagannathan, "A Cross Layer Approach to the Novel Distributed Scheduling Protocol and Event-triggered Controller Design for Cyber Physical Systems", to be submitted to IEEE Transactions on Communication.

Research in part by NSF grant ECCS#1128281 and Intelligent Systems Center

ABSTRACT

Networked Control System (NCS) is a recent topic of research wherein the feedback control loops are closed through a real-time communication network. Many design challenges surface in such systems due to network imperfections such as random delays, packet losses, quantization effects and so on. Since existing control techniques are unsuitable for such systems, in this dissertation, a suite of novel stochastic optimal adaptive design methodologies is undertaken for both linear and nonlinear NCS in presence of uncertain system dynamics and unknown network imperfections such as network-induced delays and packet losses. The design is introduced in five papers.

In Paper 1, a stochastic optimal adaptive control design is developed for unknown linear NCS with uncertain system dynamics and unknown network imperfections. A value function is adjusted forward-in-time and online, and a novel update law is proposed for tuning value function estimator parameters. Additionally, by using estimated value function, optimal adaptive control law is derived based on adaptive dynamic programming technique. Subsequently, this design methodology is extended to solve stochastic optimal strategies of linear NCS zero-sum games in Paper 2.

Since most systems are inherently nonlinear, a novel stochastic optimal adaptive control scheme is then developed in Paper 3 for nonlinear NCS with unknown network imperfections. On the other hand, in Paper 4, the network protocol behavior (e.g. TCP and UDP) are considered and optimal adaptive control design is revisited using output feedback for linear NCS. Finally, Paper 5 explores a co-design framework where both the controller and network scheduling protocol designs are addressed jointly so that proposed scheme can be implemented into next generation Cyber Physical Systems.

ACKNOWLEGEMENTS

I would like to express my deep thanks to my advisor, Dr. Jagannathan Sarangapani for his valuable guidance, patience and support over the last several years. I also would like to thank Dr. S. N. Balakrishnan, Dr. Sanjay Madria, Dr. Maciej Zawodniok, Dr. Yahong (Rosa) Zheng, and Dr. Al Salour for serving on my doctoral committee. In addition, I would like to thank the National Science Foundation (NSF) and Intelligent System Center (ISC) for providing financial support through a graduate research assistantship.

I am greatly thankful to my wife, Ran Cai, my parents, Heqing Xu and Ning Ge, as well as the rest of my family for their love and support. Their encouragement and understanding over the course of my education has truly been a blessing. I also would like to thank all my fellow colleagues who made my time in Ph.D. program fun and exciting.

Finally, I would like to thank the staff of ECE department for their continuous assistance and also would like to thank the staff of Curtis Laws Wilson Library for providing me with the necessary literature.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

Table Page

1. INTRODUCTION

In the past decade, significant advances in theoretical and applied research have occurred in computation, communication, and control areas. Control system has made great strides from analog control (first generation) to digital (second generation) [1] with the appearance of digital computer in 1940s. Similarly, wireless communication is being preferred over wired communication as it allows mobility. In the recent control applications, reinforcement learning which is used for computational intelligence has been introduced in complex control system design. Most recently, a communication network is combined with modern control system to form a networked feedback control system due to the presence of a real-time communication network. This novel networked control system [2-5] (NCS) concept is considered as a third generation control system [1]. In NCS, a communication packet carries the reference input, plant output, and control input which are exchanged by using a communication network among control system components such as sensor, controller and actuators as shown in Fig 1.1.

Compared with traditional control systems, a NCS can not only reduce system wiring with ease of system diagnosis and maintenance, but also increases the system agility which is one of most critical factor in developing practical modern system. Because of these advantages, a NCS, as shown in Fig 1.2., has been implemented in the manufacturing industry. Multiple devices sense data from controlled plants by using embedded sensors then packetize the data and transmit the sensed data to remote controllers through the wireless network. When the respective controllers receive information from the controlled plant, suitable control inputs can be designed based on that information and transmitted back to the respective devices through the network.

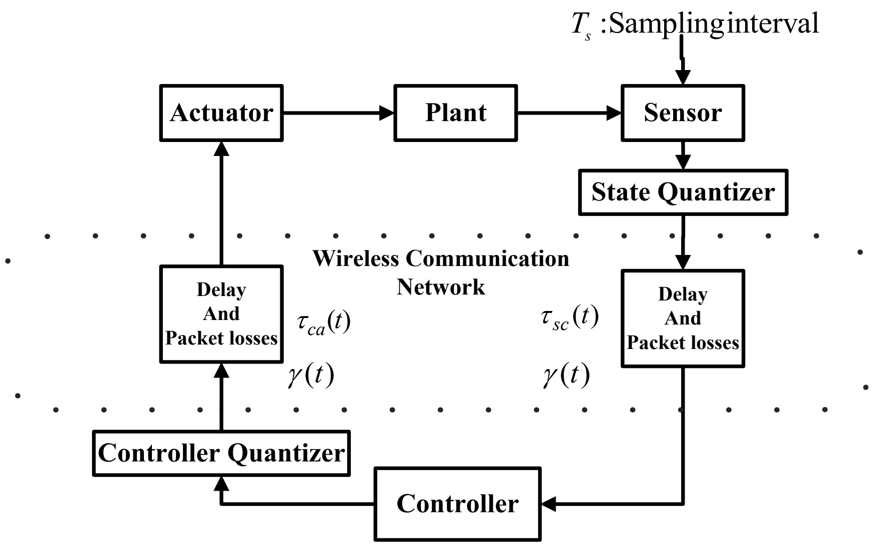


Fig 1.1. Networked control systems.

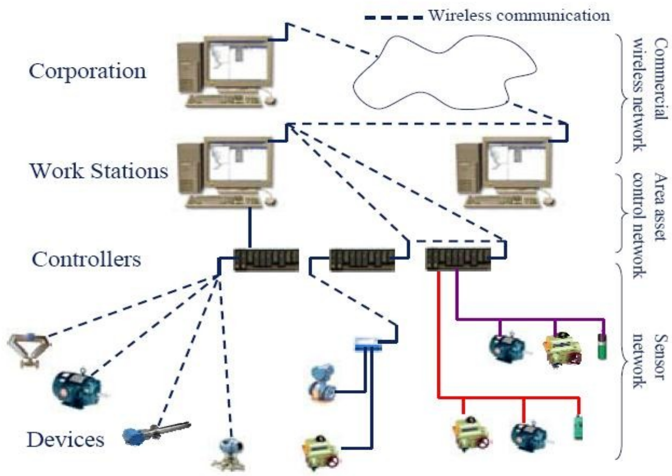


Fig 1.2. Networked control system in manufacturing [19].

Similarly, in Fig 1.3, a NCS is implemented on the smart grid which is considered as the next generation power system. The sensors can report the consumer demand to the smart grid processor, which can decide how to deal with these demands whether to request more power generation or utilize the stored energy. Compared with traditional power system, network control based smart grid can manage the power resource more efficiently.

However, due to the unreliable communication network, a NCS has many challenging issues to be solved before reaping their benefits. The first issue is the network-induced delay that occurs while exchanging data among devices connected to the shared communication media. This delay, either constant or random, can degrade the performance of control system and even destabilize the system when the delay is not explicitly considered in the design process [8]. The second issue is packet losses due to unreliable network transmission which can cause a loss in control input resulting in instability [15]. Because of limited network capacity, sensed plant data and designed control inputs need to be quantized prior to transmission which may lead to quantization errors for both measured states and control inputs [5,7]. Since these quantization errors can cause instability of a NCS [7], it is considered as the third issue.

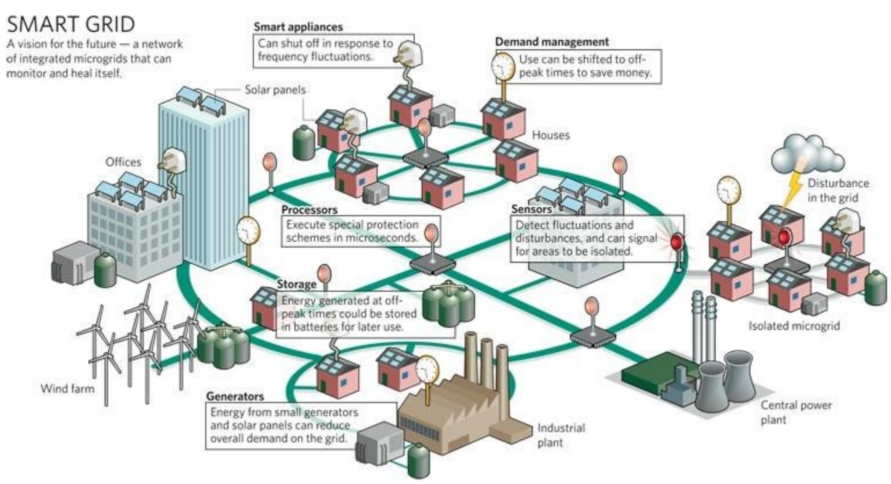


Fig 1.3. NCS in Smart Grid[1]

[1]Photo courtesy of: http: http://www.consumerenergyreport.com/wp-content/ uploads/ 2010/04/smartgrid.jpg

Next, an overview of current methodologies for the NCS design is presented, and their shortcomings are discussed. Subsequently, the organization and the contributions of this dissertation are introduced.

1.1. OVERVIEW OF NETWORKED CONTROL METHODOLGIES

As introduced in the above section, although the NCS can offer several advantages, it also brings many challenging issues (e.g. network-induced delays and packet losses) due to the presence of a communication network and its associated network protocol utilized for packet transmission. For the NCS shown in Fig. 1.1, researchers [2-16] analyzed the stability of such NCS starting 1990s. In [4,5,8,9], authors evaluated stability and performance of a NCS with constant network-induced delay and derived a stability region of linear NCS. Selecting a conventional stable controller with constant gains, related maximum allowable transfer interval (MATI) and maximum allowable delay can be calculated based on the stability region of NCS [8-9]. In addition, the effect of packet losses on the NCS has been analyzed in [14].

Similar to network-induced delay, authors derived a stability region for packet losses based on stochastic control [18]. Zhang et. al [8] conducted the stability analysis of the NCS in the presence of packet losses and delays and proposed the region of stability. Experimental studies in [8-9] have illustrated that conventional controller can still maintain a NCS stable in the mean when network-induced delays and packet losses fall within the region of stability.

On the other hand, optimal control design is also pursued in the NCS research. Nilsson [6] introduced the optimal design problem and derived optimal controller for the

NCS with a short network-induced delay (i.e. delay less than one sampling interval). In [6], Nilsson represented the NCS dynamics with augment states. Then, optimal controller has been derived using standard Riccati Equation-based optimal control theory [18].

Recently, Hu and Zhu extended linear NCS optimal controller design with network-induced delays of over several sampling intervals (i.e. delay is more than one sampling interval) [13]. Compared with previous NCS schemes, the work in [14] considered specific network protocol such as a transmission control protocol (TCP) and user datagram protocol (UDP), and derived optimal control design for NCS under TCP and UDP. However, all these methodologies [2-16] required full knowledge of system dynamics and network imperfections (i.e. network-induced delay and packet losses) which are not known beforehand in practical NCS. Therefore, methods developed in [2-16] may not be suitable to yield best performance during implementation. Also, literature on NCS focuses only on linear dynamic systems. However, practical systems are inherently nonlinear. Therefore, the control design for such nonlinear NCS is important and necessary.

In addition, network protocol design is critical for NCS design [5][9]. At present, limited effort [9][14] has been in place to understand the effect of protocols and most of them merely evaluate the behaviors of existing network protocols by separating the controller and network protocol design. However, since controller and network protocol design are related to each other closely, they cannot be separated in a truly NCS design. Thus, in this dissertation, a novel controller and network protocol designs are introduced jointly to address the drawbacks described above. Additionally, stability guarantees are provided by comparing the proposed schemes with that of the existing NCS.

1.2. ORGANIZATION OF THE DISSERTATION

In this dissertation, novel stochastic optimal adaptive control and network protocol designs for linear/nonlinear NCS are undertaken while relaxing the knowledge of the system dynamics and network imperfections. This dissertation is presented in five papers, and their relationship to one another is illustrated in Fig. 1.2. The common theme in the five papers is the optimal adaptive control of linear/nonlinear NCS.

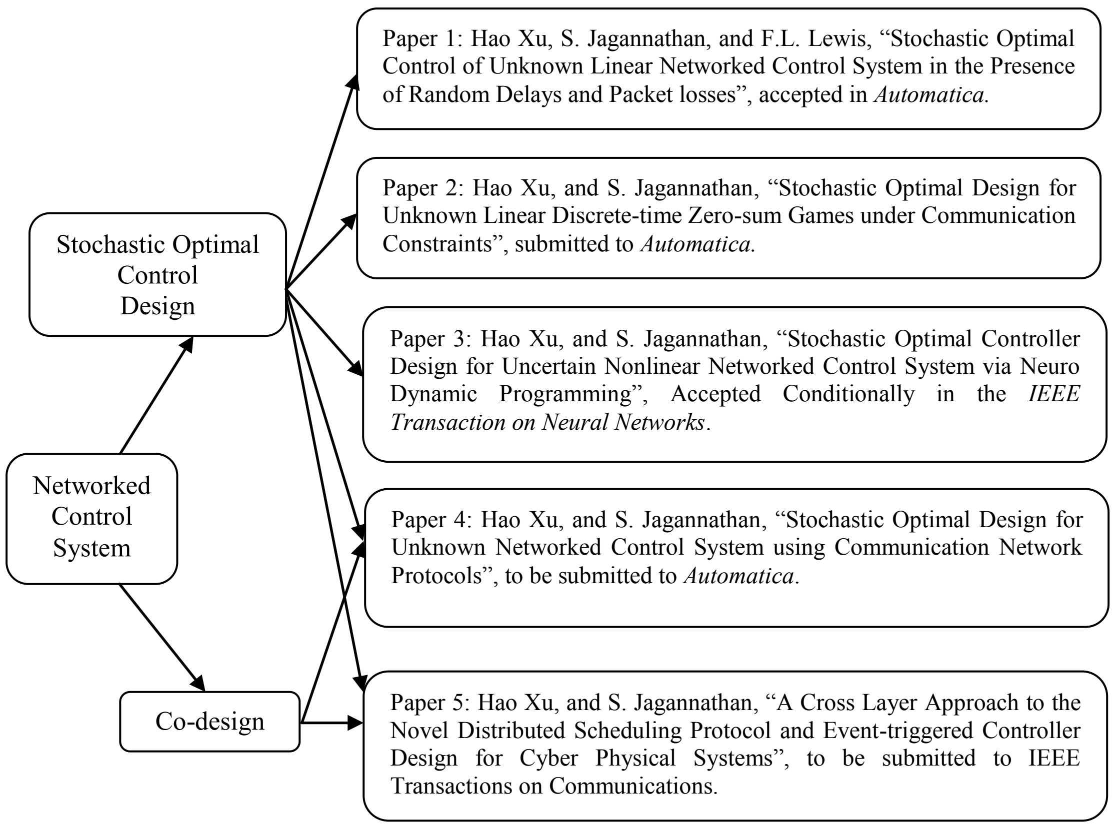


Fig. 1.4 Dissertation Outline

In the first paper, a novel stochastic optimal adaptive control of linear NCS with uncertain system dynamics in the presence of network imperfections such as random delays and packet losses is derived. The value function approximation and Q-learning ideas are used to solve the infinite horizon optimal regulation of unknown linear NCS. Then, using certainty equivalence property, a stochastic suboptimal adaptive control scheme is proposed. Lyapunov theory is applied to guarantee that all signals are asymptotically stable in the mean and estimated control signals converge to optimal or suboptimal control inputs respectively. In comparison with other works [2-16], the proposed method relaxes the requirement of system dynamics and network imperfections.

Subsequently, the control design in the first paper is extended to generate optimal strategies for linear NCS zero-sum games in the second paper. System dynamics and network imperfections are not needed for the proposed optimal strategy.

In the third paper, a novel nonlinear NCS representation incorporating the system uncertainties and network imperfections is introduced first by using input and output measurements for facilitating output feedback. Further, an online neural network (NN) identifier is introduced to estimate the control coefficient matrix. Subsequently, the critic and action NNs are employed along with NN identifier to determine the forward-in-time, time-based stochastic nonlinear optimal adaptive control of NCS without using value and policy iterations. Lyapunov techniques are used to show that that all the closed-loop signals and NN weights are uniformly ultimately bounded (*UUB*) in the mean while the approximated control input convergences close to its target value over time in the mean.

By contract, in the fourth paper, TCP and UDP are considered with NCS for evaluating the impact of network protocol reliability on controller performance. Here, a novel observer is derived to estimate the system states in the presence of unknown system dynamics and network imperfections first. Next, stochastic optimal adaptive output-feedback controller by using ADP is utilized to solve the infinite horizon optimal regulation of linear NCS under TCP and UDP respectively. Finally, Lyapunov stability analysis indicate that all signals are asymptotically stable in the mean for linear NCS under TCP, and uniformly ultimately bounded in the mean for linear NCS under UDP.

Finally, the last chapter proposed a cross layer co-design for the Cyber Physical System (CPS), which is considered as a new breed of promising emerging dynamic systems. First, by maximizing the utility function which is generated based on the information from both application and network layer, a novel distributed scheduling protocol is derived via cross layer approach. Subsequently, a novel adaptive model based optimal event-triggered control scheme is developed for real-time CPS with unknown system dynamics in the application layer. Compared with traditional scheduling algorithms, the proposed distributed scheduling scheme via cross layer approach can not only allocate the network resource efficiently but also improves the performance of the overall real-time CPS. Finally, simulation results are included to illustrate the effectiveness of proposed cross-layer co-design.

1.3. CONTRIBUTIONS OF THE DISSERTATION

This dissertation provides contributions to the field of linear and nonlinear NCS in the design of an optimal and suboptimal adaptive controller and network protocol. As a consequence, proposed designs can not only render a desired performance in terms of attaining optimality but also maintain the NCS stable in the mean in the presence of unknown system dynamics and network imperfections. Traditionally, the optimal control design for NCS [2-16] work backward-in-time and require full knowledge of system dynamics and network imperfections. The ADP-based available optimal techniques, on the other hand, relax the need for system dynamics and use value and/or policy iterations which may be difficult to be implemented on hardware since the number of iterations needed is not known. The proposed effort overcomes these deficiencies.

The contributions of Paper 1 include the stochastic optimal and suboptimal adaptive control design of linear NCS in forward-in-time in the presence of unknown system dynamics and network imperfections by using a value function estimator. A suitable representation of the linear NCS is derived first by using augmented stated and then optimal and suboptimal adaptive controller is designed by using Q-function approach. Next, these results are extended to linear NCS zero-sum games, and stochastic optimal adaptive control and disturbance inputs are now derived in Paper 2.

On the other hand, the contributions of Paper 3 include the development of nonlinear optimal adaptive controller for nonlinear NCS in presence of system uncertainties and network imperfections by using input-output measured data. Here as a first step, a suitable nonlinear NCS representation is obtained for the controller design in the input-output form. The need for control coefficient matrix is relaxed when compared to [18] by using a nonlinear neural network identifier.

In addition, a novel optimal adaptive controller incorporating the network protocol behavior such as the TCP or UDP is introduced. In all the above papers, closed-loop stability is demonstrated by using Lyapunov analysis. For the case of linear NCS, asymptotic stability is demonstrated and for the case of nonlinear NCS, uniform ultimate boundedness of the closed-loop is shown.

Finally, Paper 5 will consider a promising new class of emerging dynamic system referred to as CPS by using a co-design framework where the optimal adaptive model-based event-triggered controller and network distributed scheduling protocol designs are performed in a joint manner through the selection of suitable utility function. It is important to note that available current CPS literature [20-22] usually separates the control and network protocol designs as two separate problems. Instead, in the proposed co-design framework, the utility function is developed based on information from both network layer and application layer which in turn optimizes the performance of control system and the network efficiency.

PAPER I

STOCHASTIC OPTIMAL CONTROL OF UNKNOWN LINEAR NETWORKED CONTROL SYSTEM IN THE PRESENCE OF RANDOM DELAYS AND PACKET LOSSES

H. Xu, S. Jagannathan, and F. L. Lewis

Abstract - In this paper, the stochastic optimal control of linear networked control system (NCS) with uncertain system dynamics and in the presence of network imperfections such as random delays and packet losses is derived. The proposed stochastic optimal control method uses an adaptive estimator (AE) and ideas from Q-learning to solve the infinite horizon optimal regulation of unknown NCS with time-varying system matrices. Next, a stochastic suboptimal control scheme which uses AE and Q-learning is introduced for the regulation of unknown linear time-invariant NCS that is derived using certainty equivalence property. Update laws for online tuning the unknown parameters of the AE to obtain the Q-function are derived. Lyapunov theory is used to show that all signals are asymptotically stable (AS) in the mean and that the estimated control signals converge to optimal or suboptimal control inputs in the mean. Simulation results are included to show the effectiveness of the proposed schemes. The result is an optimal control scheme that operates forward-in-time manner for unknown linear systems in contrast with standard Riccati equation-based schemes which function backward-in-time.

Key words— Networked Control System (NCS), Q-function, Adaptive Estimator, Optimal Control.

1. Introduction

Feedback control systems with control loops closed through a real-time network are called Networked Control Systems (NCS) (Halevi and Ray, 1988; Branicky et al., 2000; Wu and Chen, 2007; Cloosterman et al., 2009). In NCS, a communication packet carries the reference input, plant output, and control input which are exchanged using a network among control system components such as sensors, controller, and actuators. The primary advantages of NCS are reduced system wiring, ease of system diagnosis and maintenance, and increased system agility. However, insertion of the communication network in the feedback loop brings many challenging issues.

The first issue is the network-induced delay that occurs while exchanging data among devices connected to the shared medium. This delay, either constant or random, can degrade the performance of control system and even destabilize the system when the delay is not explicitly considered in the design process. The second issue is packet losses due to unreliable network transmission which can cause a loss in control input resulting in instability. These issues have been identified in the literature and are being studied.

For instance, Cloosterman et al. (2009) analyzed the stability of NCS with network-induced delays. Walsh et al. (1999) and Lian et al. (2001) considered stability performance of NCS with constant delays. Azimi-Sadjadi (2003), Wu and Chen (2007), Schenato et al. (2007) analyzed the stability performance of NCS with packet losses. Eventually Zhang et al. (2001) conducted the stability analysis of NCS with delays and packet losses and proposed a stability region.

While stable controllers are encouraging, optimality is generally preferred for NCS which is very difficult to attain. Lian et al. (2002) proposed the optimal controller design by using classical optimal control theory (Lewis and Syrmos, 1995) for NCS with multiple constant delays embedded into the NCS representation. Using the stochastic optimal control theory (Stengel, 1986; Bertsekas and Shreve, 1978; Åstrom, 1970), Nilsson et al. (1998) proposed the optimal and suboptimal controller designs for linear NCS with random delays. Although these optimal and suboptimal controller designs have resulted in satisfactory performance, they all require information about the system NCS dynamics and information on delays and packet losses which are not known beforehand.

On the other hand, adaptive dynamic programming (ADP) schemes proposed by Werbos (1991), Watkins (1989), intend to solve optimal control problems forward-in-time by using value and policy iterations. There are four techniques in ADP (i.e. heuristic dynamic programming (HDP), action dependent HDP (ADHDP), dual heuristic programming (DHP) and action dependent DHP (ADDHP)), but they all require policy and value iterations.

Al-Tamimi, Lewis and Abu-Khalaf (2007) used the *Q*-learning policy iteration method to solve the optimal strategies for linear discrete-time system quadratic zero-sum games in forward-in-time without requiring the system dynamics wherein the system dynamics are defined as constant matrices. It is important to note that policy and value iteration-based schemes are difficult to implement on hardware (Dierks et al., 2009) since it is not clear how to select the number of iterations required for convergence and stability while keeping the hardware constraints. Inadequate number of policy and value iterations can result in instability (Dierks et al., 2009). Therefore, Dierks and Jagannathan (2009) used two time-based neural networks (NN) to solve the Hamilton-Jacobi-Bellman (HJB) equation forward-in-time for the optimal control of a class of general nonlinear affine discrete-time systems without using policy or value iterations. However, these papers did not consider the effects of delays and packet losses which are normally found in a NCS. The delays and packet losses cause instability (Zhang et al., 2001) if they are not considered carefully which in turn make the optimal controller design more involved and different than (Al-Tamimi et al., 2007; Zhang et al., 2009).

Thus, this paper introduces ADHDP technique for the optimal and suboptimal control of linear NCS with uncertain system dynamics and in the presence of unknown random network-induced delays and packet losses. In other words, first a linear NCS with random delays and packet losses will be represented by a time-varying linear system with unknown system matrices. The suboptimal approach in (Al-Tamimi et al., 2007; Zhang et al., 2009) is not directly applicable to the NCS due to the inclusion of network imperfections such as these delays and packet losses.

A novel approach is undertaken to the optimal regulation of linear NCS with random delays and packet losses to solve the Bellman equation (Wonham, 1968) online and forward-in-time without using policy and value iterations. Using an initial stabilizing control, an adaptive estimator (AE) (Franklin et al., 1994) is tuned online to learn the stochastic cost function without needing to solve the stochastic Riccati equation (SRE). Then, using the idea of Q-learning, the optimal controller which minimizes the stochastic cost function can be calculated based on the information provided by the AE. Thus the proposed AE-based scheme relaxes the requirements for system dynamics and information on random delay and packet losses. Next, the suboptimal controller design is derived based on NCS representation that is obtained by using certainty equivalence property. For the suboptimal control, linear NCS is modeled as a time-invariant system with unknown matrices. The suboptimal controller reduces computational complexity.

This paper is organized as follows. First, NCS background representation is given in Section 2. In Section 3, the stochastic optimal and suboptimal regulation controls of NCS are introduced. Section 4 illustrates the effectiveness of proposed schemes via numerical simulations, and Section 5 provides concluding remarks.

2. Background

The basic structure of NCS considered in this paper is shown as Figure 1 where the feedback control loop is closed over a wireless network. Since wireless network bandwidth is limited, two types of network-induced delays and one type of packet losses are included in this structure: (1) $\tau_{sc}(t)$: sensor-to-controller delay, (2) $\tau_{ca}(t)$: controller-to-actuator delay, and (3) $\gamma(t)$: indicator of packet received.

The following assumption is needed similar to other works (Liou and Ray, 1991; Hu and Zhu, 2003):

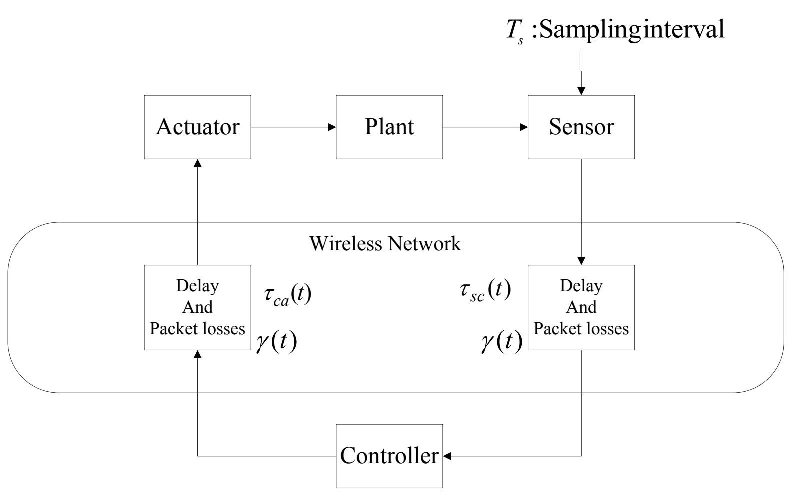


Fig 1. Networked Control System (NCS)

Assumption 1:

a). Sensor is time-driven while the controller and actuator are event-driven (Hu and Zhu, 2003).

b). The communication network considered is a wide area wireless network so that the two types of network-induced delays are independent, ergodic and unknown whereas their probability distribution functions are considered known (Liou and Ray, 1991; Hu and Zhu, 2003; Goldsmith, 2003).

c). The sum of the two delay types is bounded (Liou and Ray, 1991) while the initial state of linear system is deterministic (Hu and Zhu, 2003).

A linear time-invariant system $\dot{x}(t) = Ax(t) + Bu(t)$ is considered. However, considering the effects of network-induced delays and packet losses, the original controlled plant can be expressed as

$$\dot{x}(t) = Ax(t) + \gamma(t)Bu(t - \tau(t)) \tag{1}$$

where $\gamma(t) = \begin{cases} \mathbf{I}^{n\times n} & \text{if the control input is received at time } t \\ \mathbf{0}^{n\times n} & \text{if the control input is lost at time } t \end{cases}$, $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m$ and $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}$ representing system matrices. From Assumption 1, we can assume that the sum of network-induced delays is bounded above i.e. $\tau(t) = \tau_{sc}(t) + \tau_{ca}(t) < \bar{d}T_s$ where $\bar{d}$ represents the delay bound while T_s is the sampling interval.

During a sampling interval $[kT_s, (k+1)T_s) \ \ \forall k$, the controller input $u(t)$ to the plant is a piecewise constant. According to Assumption 1, there are at most $\bar{d}$ current and previous control input values that can be received at the actuator. If several control inputs are received at the same time, only the newest control input is allowed to act on the controlled plant during any sampling interval $[kT_s,(k+1)T_s)\ \ \forall k$, and other previous control inputs are deduced. Since controller is event driven, the plant will implement control input at these time instant $kT_s + t_i^k, i = 0,1,...,\bar{d}$ and $t_i^k < t_{i-1}^k$ where $t_i^k = \tau_i^k - iT$ as illustrated in Figure 2 (Liou and Ray, 1991).

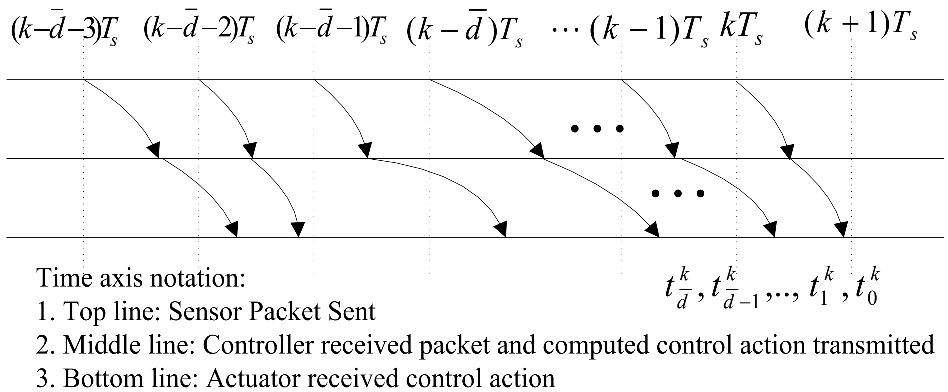


Fig 2. Timing diagram of signals in NCS.

For the event driven controller, the control input becomes u_k in response to the sensor signal x_k. Integration of (1) over a sampling interval $[kT_s,(k+1)T_s)\ \ \forall k$ yields

$$\begin{aligned} x_{k+1} &= A_s x_k + B_0^k u_k^a + B_1^k u_{k-1}^a + \cdots + B_{\bar{d}}^k u_{k-\bar{d}}^a, \\ u_{k-i}^a &= \gamma_{k-i} u_{k-i} \end{aligned} \tag{2}$$

where $x_k = x(kT)$, $A_s = e^{AT}$, u_k is the control input computed at controller, u_k^a is control input received at actuator with $B_i^k = \int_{\tau_i^k - iT_s}^{\tau_{i-1}^k-(i-1)T_s} e^{A(T_s - s)} ds B \bullet \varphi\left(T_s + \tau_{i-1}^k - \tau_i^k\right) \bullet \varphi\left(\tau_i^k - iT_s\right)\ \ \forall i = 1,2,...,\bar{d}$,

$B_0^i = \int_{\tau_0^k}^{T_s} e^{A(T_s - s)} ds B \bullet \varphi\left((k+1)T_s - \tau_0^k\right)$, $\varphi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$; and

$$\gamma_{k-i} = \begin{cases} 0, \text{ if } u_{k-i} \text{ was received during } \left[kT_s,(k+1)T_s\right) \\ 1, \text{ if } u_{k-i} \text{ was lost during } \left[kT_s,(k+1)T_s\right) \end{cases}$$

Using (2), a new augment state variable vector consisting of current state and past inputs, $z_k = \begin{bmatrix} x_k^T & u_{k-1}^T & \cdots & u_{k-\bar{d}}^T \end{bmatrix}^T \in \mathbb{R}^{n+\bar{d}m}$, is defined such that (2) can be expressed as

$$z_{k+1} = A_{zk} z_k + B_{zk} u_k \tag{3}$$

where the time-varying system matrices are given by

$$A_{zk} = \begin{bmatrix} A_s & \gamma_{k-1} B_1^k & \cdots & \gamma_{k-i} B_i^k & \cdots & \gamma_{k-\bar{d}} B_{\bar{d}}^k \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & I_m & \cdots & \cdots & 0 & 0 \\ \vdots & 0 & I_m & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & I_m & 0 \end{bmatrix}, \qquad B_{zk} = \begin{bmatrix} \gamma_k B_0^k \\ I_m \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

In this paper, we derive the optimal controller to minimize the stochastic cost function

$$J_k = \underset{\tau,\gamma}{E}\left[\sum_{m=k}^{\infty}\left(x_m^T S x_m + u_m^T R u_m\right)\right] \qquad k = 0,1,2,... \tag{4}$$

where S and R are symmetric positive semi-definite and symmetric positive definite constant matrices respectively and $\underset{\tau,\gamma}{E}(\bullet)$ is the expected operator (in this case the mean value) of $\sum_{m=k}^{\infty}\left(x_m^T S x_m + u_m^T R u_m\right)$ based on the random network-induced delays and packet losses. After redefining the augment state vector z_k, original stochastic cost function, (4) can be expressed as

$$J_k = \underset{\tau,\gamma}{E}\left[\sum_{m=k}^{\infty}\left(z_m^T S_z z_m + u_m^T R_z u_m\right)\right] \qquad k = 0,1,2,... \tag{5}$$

where $S_z = \begin{bmatrix} S & 0 & \cdots & 0 \\ 0 & \frac{R}{\bar{d}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{R}{\bar{d}} \end{bmatrix}$, and $R_z = \frac{1}{\bar{d}} R$.

Note the matrices S_z and R_z are still symmetric positive semi-definite and symmetric positive definite respectively.

Remark 1: if $\bar{d} = 1$, then the overall network-induced delay (sum of the two delays) is less than a sampling interval. Then stochastic optimal and suboptimal designs for known NCS system dynamics can be found in (Nilsson et al., 1998).

3. Optimal and Suboptimal Regulator Design

In this section, the idea of Q-learning (Watkins, 1989) and concept of AE are utilized to develop the stochastic optimal and suboptimal control designs for NCS with unknown linear time-varying dynamics in the presence of unknown random delays and packet losses. In Section 3.1, a novel stochastic optimal control will be formulated for the NCS. First, Q-function is set up for NCS with random delays and packet losses. Second, model-free online tuning of the parameters based on AE and Q-learning algorithm will be proposed. Eventually the convergence proof is given. Subsequently, stochastic suboptimal control is proposed in Section 3.2.

3.1. Optimal Control

In this section, stochastic optimal control of NCS is obtained without knowledge of system dynamics and wireless network imperfections. First, NCS dynamics (3) with random delays and packet losses is used. It is important to note when time-varying NCS dynamics A_{zk}, B_{zk} are known, the stochastic optimal control can be obtained by solving

Stochastic Ricatti Equation (SRE) in backward-in-time manner. However, in practice random delays and packet losses will affect the dynamics A_{zk}, B_{zk} which makes the dynamics uncertain. Therefore, a novel Q-function approach is introduced to overcome this drawback but this cost function is not known before hand. Consequently, a novel AE is proposed to learn this $Q(\bullet)$ function online. Eventually, even when NCS dynamics A_{zk}, B_{zk}, random delays and packet losses are unknown, stochastic optimal control still can be obtained in terms of estimated Q-function in a forward-in-time manner without using value and policy iterations in contrast with existing Q-function based ADP schemes where value and policy iterations are needed. Next the Q-function setup is described.

3.1.1. Q-function Setup

Consider the NCS in the presence of practical random delays and packet losses described by (3) as $z_{k+1} = A_{zk} z_k + B_{zk} u_k$ where $\|B_{zk}\|_F \le B_M$ (Note $\|\bullet\|_F$ denotes the Frobenius norm and B_M is the Frobenius norm bound of B_{zk}). Given the unique equilibrium point at $z = 0$ for the NCS system on a set Ω, assume that the states are considered measurable. According to these conditions, the stochastic optimal control input which minimizes the cost function J_k (5) for NCS system (3) can be derived as $u_k^* = -K_k z_k$ with K_k being the optimal gain and u_k^* being the control input. According to the optimal control theory (Lewis and Syrmos, 1995), the stochastic cost function can be represented as

$$J_k = \underset{\tau,\gamma}{E}\left(z_k^T P_k z_k\right) \tag{6}$$

where $P_k \ge 0$ is the solution to the SRE (Wonham, 1968). The optimal action dependent value function or simply Q-function denoted as $Q(\bullet)$ of NCS is defined in terms of expected value as

$$Q(z_k, u_k) = \underset{\tau,\gamma}{E}\{[r(z_k, u_k) + J_{k+1}]\} = \underset{\tau,\gamma}{E}\{[z_k^T \; u_k^T] H_k [z_k^T \; u_k^T]^T\} \tag{7}$$

where $r(z_k,u_k)=z_k^T S_z z_k + u_k^T R_z u_k$. It is important to note that the matrix H_k is time-varying as opposed to the one defined for stochastic suboptimal control in the next section. Since the stochastic optimal control, u_k^*, is dependent on state z_k which is known at time k, Q-function can be expressed as $Q(z_k,u_k)=[z_k^T \; u(z_k)^T]\underset{\tau,\gamma}{E}(H_k)[z_k^T \; u(z_k)^T]^T$. Then using Bellman equation (Lewis and Syrmos, 1995) and stochastic cost function definition, the following equation can be formulated by applying Q-function (7) as

$$\begin{aligned}
\begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \underset{\tau,\gamma}{E}(H_k)\begin{bmatrix} z_k \\ u_k \end{bmatrix} &= \underset{\tau,\gamma}{E}\{[r(z_k,u_k)+J_{k+1}]\} \\
&= z_k^T S_z z_k + u_k^T R_z u_k + \underset{\tau,\gamma}{E}\left(z_{k+1}^T P_{k+1} z_{k+1}\right) \\
&= \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} S_z & 0 \\ 0 & R_z \end{bmatrix}\begin{bmatrix} z_k \\ u_k \end{bmatrix} + \underset{\tau,\gamma}{E}\left\{\left(\begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} A_{zk}^T \\ B_{zk}^T \end{bmatrix}^T P_{k+1} \begin{bmatrix} A_{zk}^T \\ B_{zk}^T \end{bmatrix}\begin{bmatrix} z_k \\ u_k \end{bmatrix}\right)\right\} \\
&= \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} S_z + \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} A_{zk}\right) & R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} B_{zk}\right) \end{bmatrix}\begin{bmatrix} z_k \\ u_k \end{bmatrix}
\end{aligned} \tag{8}$$

Therefore $\underset{\tau,\gamma}{E}(H_k)$ can be written in terms of the system matrices and solution to the SRE as

$$\begin{aligned}
\bar{H}_k = \underset{\tau,\gamma}{E}(H_k) &= \begin{bmatrix} \bar{H}_k^{zz} & \bar{H}_k^{zu} \\ \bar{H}_k^{uz} & \bar{H}_k^{uu} \end{bmatrix} \\
&= \begin{bmatrix} S_z + \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} A_{zk}\right) & R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} B_{zk}\right) \end{bmatrix}
\end{aligned} \tag{9}$$

The optimal action dependent value function $Q(z_k,u_k)$ is equal to stochastic cost function J_k. Therefore, we have

$$J_k = Q(z_k,u_k) \tag{10}$$

Then using (9) and stochastic control theory (Stengel, 1986), the optimal time-varying gain can be expressed in terms of $\overline{H}_k$ as

$$K_k = [R_z + \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} B_{zk})]^{-1} \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} A_{zk}) = \left(\overline{H}_k^{uu}\right)^{-1} \overline{H}_k^{uz} \tag{11}$$

Remark 2: According to (11), if the solution to the SRE, P_{k+1} is known, then the time-varying system matrices A_{zk}, B_{zk} are still required to compute the controller gains in a backward-in-time manner. On the other hand, if time varying matrix $\overline{H}_k$ can be learned online at time k without the knowledge of linear time-varying system dynamics, optimal controller gain can be solved not only without NCS system matrices, but also forward-in-time.

3.1.2. Model-free Online Tuning based on Adaptive Estimator and Q-Learning

The proposed online tuning approach entails one AE which is used to learn Q-function. Since Q-function include $\overline{H}_k$ matrix, this matrix can be solved online and the control signal can be obtained using (11). We make the following assumption (Middleton and Goodwin, 1988) since the NCS is linear, the delays of NCS are bounded above, packet losses satisfy the Bernoulli distribution and the delays and packet losses change slowly (Goldsmith, 2003).

Assumption 2: The Q-function, $Q(z_k, u_k)$, can be expressed as the linear in the unknown parameters (LIP).

By using the stochastic adaptive control theory (Jagannathan, 2006) and the definition of Q-function (7), $Q(z_k, u_k)$ can be represented in vector form similar to the AE representation as

$$Q(z_k, u_k) = w_k^T \overline{H}_k w_k = \overline{h}_k^T \overline{w}_k \tag{12}$$

where $\overline{h}_k = vec(\overline{H}_k), w_k = [z_k^T \; u^T(z_k)]^T, w_k \in \mathbb{R}^{n+(\bar{d}+1)m=l}$, and $\overline{w}_k = (w_{k1}^2, \ldots, w_{k1}w_{kl}, w_{k2}^2, \ldots, w_{kl-1}w_{kl}, w_{kl}^2)$ is the Kronecker product quadratic polynomial stochastic independent basis vector and $\overline{h}_k = vec(\overline{H}_k)$ with the vector function acting on $l \times l$ matrices thus yielding a $l(l+1)/2 \times 1$ column vector (Note: the $vec(\bullet)$ function is constructed by stacking the columns of the matrix into one column vector with the off-diagonal elements which can be combined as $H_{mn} + H_{nm}$).

The time-varying matrix $\overline{H}_k$ can be considered as slowly varying (Middleton and Goodwin, 1988). Then Q-function can be expressed as unknown time-varying target parameter vector and the regression function $\overline{w}_k$. Next, the Q-function $Q(z_k,\ u_k)$ estimation will be considered.

3.1.3. Q-function Estimation for Optimal Regulator Design

The *Q*-learning was originally proposed in (Watkins, 1989; Werbos, 1992) to solve the optimal control problems for time-invariant systems by using policy or value iterations. Here, the Q-function scheme is extended to time-varying linear systems without using iterative approach. According to the definition of Q-function and relationship between Q-function and stochastic cost function (10), the relationship between $\overline{H}_k$ matrix in (9) and the stochastic cost function is given as

$$J_k = w_k^T \overline{H}_k w_k = \overline{h}_k^T \overline{w}_k \tag{13}$$

Then Q-function $Q(z_k, u_k)$ can be approximated by an AE as

$$\hat{Q}(z_k, u_k) = \hat{\overline{h}}_k^T \overline{w}_k \tag{14}$$

where $\hat{\bar{h}}_k^T$ is the estimate value of the target parameter vector $\bar{h}_k^T$ with regressor satisfying $\|\bar{w}_k\| = 0$ for $\|z_k\| = 0$.

It is observed that Bellman Equation can be rewritten as $J_{k+1} - J_k + r(z_k, u_k) = 0$. This relationship, however, is not guaranteed to hold when the estimated matrix $\hat{\bar{H}}_k$ is applied. Hence, using delayed values for convenience, the residual error, e_{hk}, associated with (14) can be expressed as $\hat{J}_k - \hat{J}_{k-1} + r(z_{k-1}, u_{k-1}) = e_{hk}$, i.e.

$$\begin{aligned} e_{hk} &= r(z_{k-1}, u_{k-1}) + \hat{\bar{h}}_k^T \bar{w}_k - \hat{\bar{h}}_k^T \bar{w}_{k-1} \\ &= r(z_{k-1}, u_{k-1}) + \hat{\bar{h}}_k^T (\bar{w}_k - \bar{w}_{k-1}) \\ &= r(z_{k-1}, u_{k-1}) + \hat{\bar{h}}_k^T \Delta W_{k-1} \end{aligned} \tag{15}$$

where $\Delta W_{k-1} = \bar{w}_k - \bar{w}_{k-1}$.

The residual dynamics in (15) are then rewritten as

$$e_{hk+1} = r(z_k, u_k) + \hat{\bar{h}}_{k+1}^T \Delta W_k \tag{16}$$

Next, we define an auxiliary residual error vector as

$$\Xi_{hk} = \Gamma_{k-1} + \hat{\bar{h}}_k^T \Omega_{k-1} \in \mathbb{R}^{1\times(1+i)} \tag{17}$$

where $\Gamma_{k-1} = [r(z_{k-1}, u_{k-1}) \quad r(z_{k-2}, u_{k-2}) \quad \cdots \quad r(z_{k-1-i}, u_{k-1-j})]$ and $\Omega_{k-1} = [\Delta W_{k-1} \quad \Delta W_{k-2} \quad \cdots \quad \Delta W_{k-1-j}\ ], 0 < j < k-1 \in \mathbb{N}$ with $\mathbb{N}$ being the set of positive natural numbers. It is important to note that (17) indicates a time history of the previous $j+1$ residual errors (15) recalculated by using the most recent $\hat{\bar{h}}_k$.

The dynamics of the auxiliary vector (17) are generated similar to (16) and revealed to be

$$\Xi_{hk+1} = \Gamma_k + \hat{\bar{h}}_{k+1}^T \Omega_k \tag{18}$$

Now define the update law of the time-varying matrix $\overline{H}_k$ as

$$\hat{\bar{h}}_{k+1} = \Omega_k \left(\Omega_k^T \Omega_k\right)^{-1} \left(\alpha_h \Xi_{hk}^T - \Gamma_k^T\right) \tag{19}$$

where $0 < \alpha_h < 1$. Substituting (19) into (18) results

$$\Xi_{hk+1} = \alpha_h \Xi_{hk} \tag{20}$$

It is observed that the stochastic cost function J_k and AE (14) will become zero only when $z_k = 0$. Hence, when the system states have converged to zero, the Q-function approximation is no longer updated. It can be seen as a persistency of excitation (PE) requirement for the inputs to the Q-function estimator wherein the system states must be persistently exiting long enough for the AE to learn the stochastic cost function. In this paper, exploration noises are added to satisfy the PE condition.

Definition 1: *(Persistence of Excitation)* A stochastic vector $\beta_k \in \mathbb{R}^p$ is said to be persistency exciting if there exist positive constants δ, α and $k_0 \geq 1$, such that

$$\sum_{k=k_0}^{k_0+\delta} E[\beta_k \beta_k^T] \geq \alpha \mathbf{I}$$

where $\mathbf{I}$ is identity matrix, $E\{\bullet\}$ is the mean value of $\{\bullet\}$.

Lemma 1: The vector ΔW_k in (15) and Ω_k satisfy the persistently exciting condition with exploration noise.

Proof: Refer to the Appendix.

Now define the parameter estimation error to be $\tilde{\bar{h}}_k = \bar{h}_k - \hat{\bar{h}}_k$. Rewrite Bellman Equation using the target AE representation (13) revealing $\bar{h}_{k+1}^T \bar{w}_k = r(z_k, u_k) + \bar{h}_{k+1}^T \bar{w}_{k+1}$ which can be expressed as

$$r(z_k, u_k) = \bar{h}_{k+1}^T \bar{w}_k - \bar{h}_{k+1}^T \bar{w}_{k+1} = -\bar{h}_{k+1}^T \Delta W_k \tag{21}$$

Substituting $r(z_k, u_k)$ into (16) and utilizing (15) with $e_{hk+1} = \alpha_h e_{hk}$ from (20) yields

$$\tilde{\bar{h}}_{k+1}^T \Delta W_k = -\alpha_h r(z_{k-1}, u_{k-1}) - \alpha_h \hat{\bar{h}}_k^T \Delta W_{k-1} \tag{22}$$

Similar to $r(z_k, u_k)$, we define $r(z_{k-1},\ u_{k-1}) = -\bar{h}_k^T \Delta W_{k-1}$, and substitute this expression into (22), to get

$$\tilde{\bar{h}}_{k+1}^T \Delta W_k = \alpha_h \tilde{\bar{h}}_k^T \Delta W_{k-1} \tag{23}$$

Next, the convergence of the stochastic cost function estimation error with adaptive estimation error dynamics $\tilde{\bar{h}}_k$ given by (23) is demonstrated for an initial admissible control policy. The linear NCS time varying system dynamics are shown to be asymptotically stable in the mean if an initial admissible control policy can be applied provided the system matrices are known. However, introducing the estimated Q-function results in estimation errors for the stochastic cost function J_k , and stability of the estimated stochastic cost function needs to be studied. Subsequently, the results of Theorem 1 will be used for proving the overall closed-loop system stability in Theorem 2 by using an initial admission control policy. In order to proceed, the following definition is needed before presenting the theorem.

Definition 2: *(Asymptotic stability)* An equilibrium point x_e is said to be asymptotically stable (AS) if there exists a set $\mathbf{S} \subset \mathbb{R}^n$ such that, for every initial condition $x_0 \in \mathbf{S}$, one has $\|x_k - x_e\| \to 0$ as $k \to \infty$. In other words, the state x_k converges to x_e.

Theorem 1: *(Asymptotic Stability of the Cost AE Errors)*. Given the initial conditions for the AE parameter vectors $\hat{\bar{h}}_0$ be bounded in the set $\mathbf{S}$, let $u_0(z_k)$ be an initial admissible control policy for the linear NCS (3). Let the AE parameter update law be given by (19). Then, there exists a positive constant α_h satisfying $0 < \alpha_h < 1$ such that the adaptive parameter estimator errors converge to zero asymptotically.

Proof: Consider the positive definite Lyapunov candidate

$$V_J(\tilde{\bar{h}}_k) = (\tilde{\bar{h}}_k^T \Delta W_{k-1})^2 \tag{24}$$

The first difference is given by $\Delta V_J(\tilde{\bar{h}}_k) = (\tilde{\bar{h}}_{k+1} \Delta W_k)^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2$, and using (23) yields

$$\Delta V_J(\tilde{\bar{h}}_k) = (\alpha_h \tilde{\bar{h}}_k \Delta W_{k-1})^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2 \le -(1-\alpha_h^2)(\tilde{\bar{h}}_k \Delta W_{k-1})^2 \tag{25}$$

Since $V_J(\tilde{\bar{h}}_k)$ is positive definite and $\Delta V_J(\tilde{\bar{h}}_k)$ is negative definite (due to PE condition; Lemma 1) provided α_h is selected as above. Therefore, the parameter errors converge to zero asymptotically. This implies that $\hat{J}_k \to J_k$ and $\tilde{\bar{h}}_k \to 0$ when $k \to \infty$.

Next, we show that the estimated control input based on this estimated matrix will indeed converge to the optimal control input.

3.1.4. Estimation of the Optimal Feedback Control Signal

There are two ways to estimate the optimal control signal for regulating the NCS. One is based on time-varying matrix $\overline{H}_k$, and the other one is based on standard optimal control theory by minimizing the stochastic cost function. The difference being that the latter method requires the system dynamics and it solves the optimal controller backward. However, it is shown here that ultimately both are equivalent and therefore are used in the proofs.

Method I: As mentioned before, the time-varying matrix $\overline{H}_k$ can be estimated by using an AE. According to *Q*-learning and equation (11), the estimated optimal control input for NCS can be expressed by using the adaptive estimation of $\overline{H}_k$ as

$$\hat{u}_{1k} = -\hat{K}_k z_k = -(\hat{\overline{H}}_k^{uu})^{-1} \hat{\overline{H}}_k^{uz} z_k \tag{26}$$

Method II: Alternatively, the estimated optimal control signal which minimizes the estimated stochastic cost function (13) with the adaptive estimator (AE) $\hat{\overline{H}}_k$ as

$$\hat{u}_{2k} = -\frac{1}{2} R_z^{-1} \underset{\tau,\gamma}{E}\left(B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \right) \tag{27}$$

where

$$\hat{J}_{k+1} = \underset{\tau,\gamma}{E}\left(w_{k+1}^T \hat{H}_{k+1} w_{k+1}\right) = \underset{\tau,\gamma}{E}\left(z_{k+1}^T \hat{P}_{k+1} z_{k+1}\right) \tag{28}$$

Next, it will be shown that the optimal control input obtained by method I and II are equivalent.

***Lemma 2*:** The optimal control obtained using the estimated value of $Q(z_k, u_k)$ is same as the optimal control calculated by minimizing the stochastic cost function J_k, i.e. $\hat{u}_{1k} = \hat{u}_{2k}$.

Proof: Refer to the Appendix.

Since the equality proven in this lemma is in both ways and noting the drawback of second method, we use the first method to solve the optimal controller design for NCS.

However, we will use the Lemma 1 to complete the convergence proof since they are equivalent. Next, the stability of the cost estimation, control estimation, and adaptive estimation error dynamics are considered.

3.1.5. Closed-loop System Stability

In this section, it will be shown that time-varying $\overline{H}_k$ matrix and related stochastic cost function estimation errors dynamics are asymptotically stable in the mean. Further, the estimated control input for NCS (26) will approach the optimal control signal asymptotically. Before introducing the theorem on system stability, we present the block diagram in Figure 3 for the proposed stochastic optimal regulator of linear NCS with unknown system dynamics.

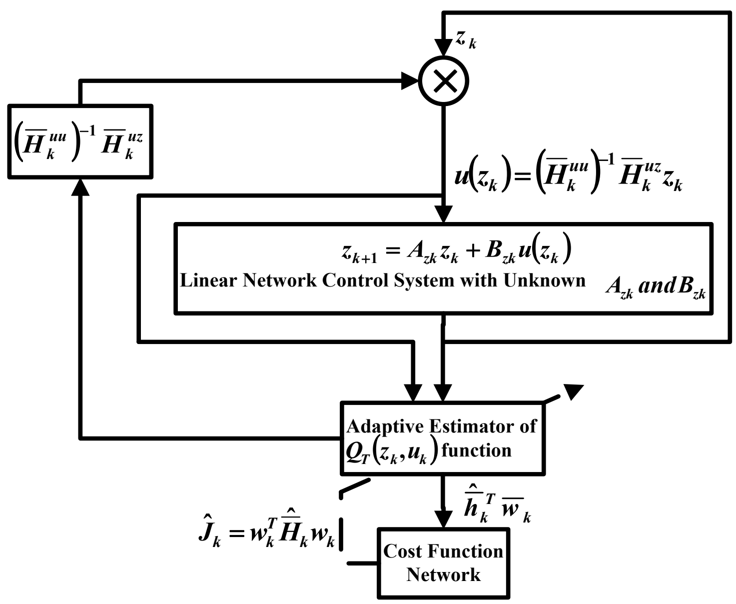


Fig 3. Stochastic optimal regulator block diagram

Next, the initial system states are considered to reside in the same set as that of the initial stabilizing control input u_{0k}. Further sufficient condition for the AE tuning gain

α_h is derived to ensure the all future states will converge to zero. Then it can be shown that the actual control input approaches the optimal control asymptotically.

Before the convergence proof, the following result is needed to establish bounds on the optimal closed loop dynamics when the optimal control is applied to the NCS (3) with random delays and packet losses.

Lemma 3: There exists an admissible control policy be applied to the NCS (3) such that the system dynamics $A_{zk}z_k + B_{zk}u_k$ are bounded above with the bounds satisfying

$$\|A_{zk}z_k + B_{zk}u_k\|^2 \le k^*\|z_k\|^2 \tag{29}$$

where $0 < k^* < 1/2$ is a constant.

Proof: Consider the Lyapunov function candidate

$$V_D(z_k) = z_k^T z_k \tag{30}$$

whose first difference of $V_D(z_k)$ is given by $\Delta V_D(z_k) = z_{k+1}^T z_{k+1} - z_k^T z_k$. Note that since u_k is an admissible control policy, it follows from the definition of admissible control that the NCS dynamics (3) with optimal control applied are asymptotically stable in the mean, and the sequence $\|z_k\|,\ \ k = 1,2,...,\infty$ monotonically decreases until it reaches zero. This result directly implies that $z_{k+1}^T z_{k+1} - z_k^T z_k < 0$ or $\Delta V_D(z_k) < 0$. Using the fact $\Delta V_D(z_k) < 0$, it is clear that $z_{k+1}^T z_{k+1} < z_k^T z_k$. Substituting the system dynamics $z_{k+1} = A_{zk}z_k + B_{zk}u_k$ yields

$$\begin{aligned}\Delta V_D(z_k) &= (A_{zk}z_k + B_{zk}u_k)^T (A_{zk}z_k + B_{zk}u_k) - z_k^T z_k < 0 \\ &= \|A_{zk}z_k + B_{zk}u_k\|^2 - \|z_k\|^2 < 0\end{aligned} \tag{31}$$

Next, we must identify a bound on $\left\| A_{zk} z_k + B_{zk} u_k \right\|$ which guarantees the sufficient condition $\Delta V_D(z_k) < 0$ for stability is still met. Selecting the bound shown in optimal control policy, (38) reveals $\Delta V_D(z_k) < -(1 - k^*)\, z_k^T z_k < 0$ as required.

Theorem 2 *(Convergence of the Optimal Control Signal):* Given the initial conditions for the system state z_0, cost function and AE parameter vectors $\hat{\bar{h}}_0$ be bounded in the set $\mathbf{S}$, let u_{0k} be any initial admissible control policy for the NCS (3) with random delays and packet losses satisfying the bounds given by (29) for $0 < k^* < 1/2$. Let the AE parameter be tuned and estimation control policy be provided by (12) and (26) respectively. Then, there exist positive constants α_h given by Theorem 1 such that the system states z_k and stochastic cost function parameter estimator errors $\tilde{\bar{h}}_k$ are all asymptotically stable in the mean. In other words, as $k \to \infty, z_k \to 0$, $\tilde{\bar{h}}_k \to 0, \hat{J}_k \to J_k$ and $\hat{u}_{2k} \to u_k^*, \hat{u}_{1k} \to u_k^*$.

***Proof*:** Refer to the Appendix.

Remark 3: It is important to note that when the delay bound is increased in NCS, the dimension of augmented state z_k increases and computational complexity also goes up due to the presence of the communication network within the control loop. While the recent embedded processors can handle the computational complexity to some extent, the delay bound due to the network phenomenon can be reduced by a suitable design of networking protocols, which is relegated as part of future effort. In the next subsection, suboptimal control scheme is presented in order to reduce the computational complexity of the controller.

3.2. Suboptimal Control

In the previous section, a new stochastic optimal control policy for NCS is introduced. It is important to note that if the linear system is time-varying, no steady-state solution to the Riccati equation (Lewis and Syrmos, 1996) can be found in general. Therefore, in this section the stochastic suboptimal control is introduced based on certainty equivalence property (Maybeck 1982; Hespanha et. al. 2007). Although, the performance of stochastic suboptimal control is not same as that of stochastic optimal control, computational complexity is reduced significantly due to constant feedback control gains without too much loss in performance from the optimality.

By using certainty equivalence property and random process (Papoulis, 1991), the NCS dynamics (3) can be approximated as a deterministic system as

$$z_{k+1} = A_z z_k + B_z u_k \tag{32}$$

where $A_z = \begin{bmatrix} A_s & \mu_1 & \cdots & \mu_{\bar{d}-1} & \mu_{\bar{d}} \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & I_m & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_m & 0 \end{bmatrix}$, $B_z = \begin{bmatrix} \mu_0 \\ I_m \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

and

$$\mu_i = \lim_{n\to\infty} \frac{1}{k} \sum_{j=1}^{n} \gamma_{j-i} B_i^j = \lim_{n\to\infty} \frac{1}{k} \sum_{j=1}^{k} (\gamma_{j-i} \int_{\tau_i^j - iT_s}^{\tau_{i-1}^j - (i-1)T_s} e^{A(T_s - s)} ds B \bullet \varphi\left(T_s + \tau_{i-1}^j - \tau_i^j\right) \bullet \varphi\left(\tau_i^j - iT_s\right))$$

$$\forall i = 1,2\ldots,\bar{d};\ \mu_0 = \lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \gamma_{j-i} B_i^j = \lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \gamma_{j-i} \int_{\tau_0^j}^{T_s} e^{A(T_s - s)} ds B \bullet \varphi\left((k+1)T_s - \tau_0^j\right)$$

It is important to note that the system matrices A_z and B_z in (32) are deterministic. Suboptimal control can be obtained by solving algebraic Riccati Equation (ARE) for the known certainty equivalence deterministic NCS representation. However, due to unknown network imperfections, the system matrices are still unknown. In this section, suboptimal controller is obtained based on AE proposed in Section 3.1 and *Q*-learning without requiring A_z and B_z .

3.2.1. Q-function Setup for Suboptimal Control of NCS

Based on linear optimal control theory (Lewis and Syrmos, 1995), stochastic cost function of (32) can be expressed as $J_k = z_k^T P z_k$ where $P \geq 0$ satisfies the ARE. Then, applying the *Q*-learning method (Watkins, 1989; Al-Tamimi, Lewis et al., 2007), stochastic suboptimal control can be formulated without knowing NCS system dynamics (32) but with policy iteration. In this part, we extend the *Q*-learning to certainty equivalence deterministic NCS representation given by (32) without using the iterative approach.

The optimal action dependent value function $Q(\bullet)$ of certainty equivalence deterministic NCS during sampling interval $[kT_s,(k+1)T_s)$ can be defined as

$$
\begin{aligned}
Q(z_k,u_k) &= r(z_k,u_k) + J_{k+1} = z_k^T S_z z_k + u_k^T R_z u_k + z_{k+1}^T P z_{k+1} \\
&= \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} S_z & 0 \\ 0 & R_z \end{bmatrix} \begin{bmatrix} z_k \\ u_k \end{bmatrix} + \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} A_z^T \\ B_z^T \end{bmatrix}^T P \begin{bmatrix} A_z^T \\ B_z^T \end{bmatrix} \begin{bmatrix} z_k \\ u_k \end{bmatrix} \\
&= \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} S_z + A_z^T P A_z & A_z^T P B_z \\ B_z^T P A_z & R_z + B_z^T P B_z \end{bmatrix} \begin{bmatrix} z_k \\ u_k \end{bmatrix} \\
&= \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} H_{zz} & H_{zu} \\ H_{uz} & H_{uu} \end{bmatrix} \begin{bmatrix} z_k \\ u_k \end{bmatrix} = \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T H \begin{bmatrix} z_k \\ u_k \end{bmatrix}
\end{aligned}
\tag{33}
$$

where H is the constant matrix associated with P which is a solution of the ARE. The relationship between P and the H can be written as

$$H = \begin{bmatrix} H_{zz} & H_{zu} \\ H_{uz} & H_{uu} \end{bmatrix} = \begin{bmatrix} S_z + A_z^T P A_z & A_z^T P B_z \\ B_z^T P A_z & R_z + B_z^T P B_z \end{bmatrix}$$

Therefore, based on optimal control theory (Lewis and Syrmos, 1995), optimal controller gain can be expressed in terms of H as

$$K = \left(R_z + B_z^T P B_z\right)^{-1} B_z^T P A_z = \left(H_{uu}\right)^{-1} H_{uz} \tag{34}$$

Thus, if the matrix H is known, the NCS system dynamics (32) are not needed to calculate the suboptimal controller gain. According to ADP-based value iteration approach (Al-Tamimi et al., 2007), the $Q(\bullet)$ function and H matrix can be learned at every sampling interval iteratively provided the number of iterations are large. In addition, in many cases a simulator or a model is needed to obtain the states in order to perform iterations which are not impossible in the case of NCS. Therefore, in this paper, the AE is tuned to learn $Q(\bullet)$ and H matrix online and subsequently utilized to calculate the suboptimal gain (34). This control input is then applied to the certainty equivalence NCS (32). Next, the suboptimal controller design is introduced.

3.2.2. Adaptive Estimation of Q-function

First, based on adaptive estimation used in stochastic optimal control (Section 3.1), we define the Q-function for the NCS (32) as

$$Q\left(z_k, u_k\right) = w_k^T H w_k = h^T \overline{w}_k \tag{35}$$

while the adaptive estimation of Q-function (35) can be expressed as

$$\hat{Q}\left(z_k, u_k\right) = w_k^T \hat{H}_k w_k = \hat{h}_k^T \overline{w}_k \tag{36}$$

where the augment state $w_k = \left[z_k^T \; u_k^T\right]^T, w_k \in \mathbb{R}^{(n+\bar{d}m)=l}$, $\overline{w}_k = (w_{k1}^2, \ldots, w_{k1}w_{kl}, w_{k2}^2, \ldots, w_{kl-1}w_{kl}, w_{kl}^2)$ which is nothing but the Kronecker product quadratic polynomial stochastic indepent basis vector, and $h = vec(H)$.

Using the adaptive estimation algorithm that is proposed as part of stochastic optimal control in Section 3.1, the update law of estimated H matrix can be written as

$$\hat{h}_{k+1} = \Omega_k \left(\Omega_k^T \Omega_k\right)^{-1} \left(\alpha_h \Xi_k^T - \Gamma_k^T\right) \tag{37}$$

where $0 < \alpha_h < 1$, $\Delta\mathbf{W}_k$ and Γ_k are defined in equation (19), Ξ_k is an auxiliary residual error vector with residual error e_k is defined as

$$e_{k+1} = r(z_k, u_k) + \hat{h}_{k+1}^T \Delta W_k \tag{38}$$

Based on Theorem 1, the parameter errors $\tilde{h}_k = h - \hat{h}_k$ converge to zero asymptotically. This implies that $\hat{J}_k \to J_k, \hat{Q}(z_k, u_k) \to Q(z_k, u_k)$ and $\tilde{h}_k \to 0$ when $k \to \infty$.

Next, with the estimated H matrix and equation (34), stochastic suboptimal control can be obtained as

$$\hat{u}_k = -(\hat{H}_k^{uu})^{-1} \hat{H}_k^{uz} z_k \tag{39}$$

3.2.3. Closed-loop System Stability for Suboptimal Control

In this section, it will be shown that matrix H and related stochastic suboptimal cost function estimation errors dynamics are asymptotically stable in the mean. Further, the estimated control input for NCS (32) will approach the suboptimal control signal asymptotically in the mean.

Theorem 3 *(Convergence of the Suboptimal Control Signal):* Given the initial system conditions for the system state z_0, cost function and AE parameter vectors $\hat{h}_0$ be bounded in the set $\mathbf{S}$, let u_{0k} be any initial admissible control policy for the NCS (32) with bounds given by $\|A_z z_k + B_z u_k\|^2 \le k^* \|z_k\|^2$ and $0 < k^* < 1/2$. Let the AE parameter be tuned and estimation control policy be provided by (37) and (39) respectively. Then, there exist positive constants α_s such that the system states z_k and stochastic suboptimal cost function parameter estimator errors $\tilde{\tilde{h}}_k$ are all asymptotically stable in the mean. In other words, as $k \to \infty, z_k \to 0, \tilde{h}_k \to 0, \hat{J}_k \to J_k$ and $\hat{u}_k \to u_k^*$.

Proof: Refer to the Appendix.

The performance of this suboptimal control design (37) and (39) with adaptive estimation algorithm will be shown to be close in comparison (Nilsson et al., 1998) to a tradition suboptimal control with known NCS system dynamics even though no knowledge of NCS system dynamics are required here. Although controller derived based on certainty equivalence deterministic NCS representation (32) is suboptimal, it is still of a great practical interest (Hespanha, et. al. 2007).

4. Simulation Results

In this section, stochastic suboptimal and optimal control of NCS is evaluated. At the same time, the standard suboptimal and optimal control of NCS with known dynamics and network imperfections is also simulated for comparison.

Example: The continuous-time version of a batch reactor system dynamics are given by (Carnevale et al. 2007)

$$\dot{x} = \begin{bmatrix} 1.38 & -0.2077 & 6.715 & -5.676 \\ -0.5814 & -4.29 & 0 & 0.675 \\ 1.067 & 4.273 & -6.654 & 5.893 \\ 0.048 & 4.273 & 1.343 & -2.104 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 5.679 & 0 \\ 1.136 & -3.146 \\ 1.136 & 0 \end{bmatrix} u \qquad (40)$$

where $x \in \mathbb{R}^{4\times1}$ and $u \in \mathbb{R}^{2\times1}$. It is important to note that this example has developed over the years as a benchmark example for NCS, see e.g., (Carnevale et al. 2007; Walsh et al. 1999).

The parameters of this NCS are selected as

1. The sampling time: $T_s = 100ms$;

2. The bound of delays is two, i.e. $\bar{d} = 2$;

3. The mean random delay values are $E(\tau_{sc}) = 80ms,\ \ E(\tau) = 150ms$;

4. Packet losses follow Bernoulli distribution with $p = 0.3$.

The distribution of random delays between the sensor and actuator are shown in Figure 4 and the packet losses are shown in Figure 5. In order to incorporate the random delays $\tau(t)$ and packet losses $\gamma(t)$ to the batch reactor (40), the original time-invariant system (40) was discretized and represented as a time-varying linear NCS given by (3) in Section 2. For instance, when $k = 20$, $t = 2\sec$, $\tau_{20} = 170ms, \tau_{19} = 168ms$, $\tau_{18} = 121ms$, $\gamma_{20} = 1, \gamma_{19} = 1$, and $\gamma_{18} = 1$, the NCS dynamics can be calculated based on (3) as

$$z_{k+1} = \begin{bmatrix} 1.48 & 0.04 & 1.47 & -1.23 & 0.01 & -0.14 & 0.02 & -0.23 \\ -0.10 & 0.45 & -0.13 & 0.20 & 0.07 & 0.01 & 0.06 & 0.02 \\ 0.07 & 0.54 & -0.18 & 1.09 & 0.1 & -0.06 & 0.11 & 0.01 \\ -0.18 & 0.54 & -1.33 & 2.21 & 0.09 & 0.14 & 0.11 & 0.21 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} z_k$$

$$+ \begin{bmatrix} 0.0035 & 0.1170 & 0.1097 & 0.1096 & 1 & 0 & 0 & 0 \\ -0.0697 & 0.0021 & -0.1886 & 0.0671 & 0 & 1 & 0 & 0 \end{bmatrix}^T u_k$$

when $k=50, t=5\sec$, $\tau_{50}=150ms,\quad \tau_{49}=152ms,\quad \tau_{48}=145ms$, $\gamma_{50}=1, \gamma_{49}=1$

and $\gamma_{48}=0$, the NCS dynamics become

$$z_{k+1}=\begin{bmatrix} 1.48 & 0.04 & 1.47 & -1.23 & 0.01 & -0.13 & 0 & 0 \\ -0.10 & 0.45 & -0.13 & 0.20 & 0.12 & 0.01 & 0 & 0 \\ 0.07 & 0.54 & -0.18 & 1.09 & 0.13 & -0.17 & 0 & 0 \\ -0.18 & 0.54 & -1.33 & 2.21 & 0.13 & 0.12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} z_k$$

$$+\begin{bmatrix} 0.0002 & 0.0317 & 0.0234 & 0.0234 & 1 & 0 & 0 & 0 \\ -0.0041 & 0 & -0.0577 & 0.0040 & 0 & 1 & 0 & 0 \end{bmatrix}^T u_k$$

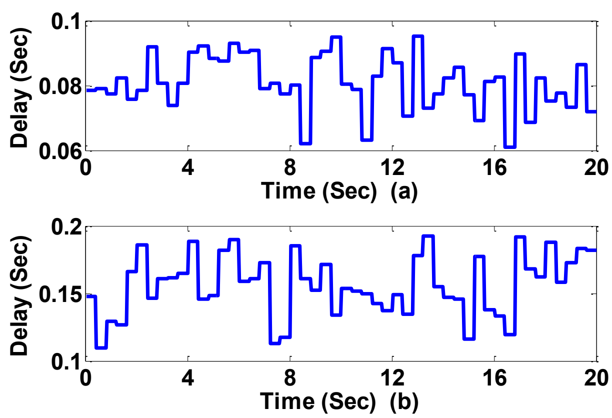


Fig 4. The distribution of random delays in NCS: (a) delay between sensor and controller τ_{sc}; (b) delay in total NCS τ .

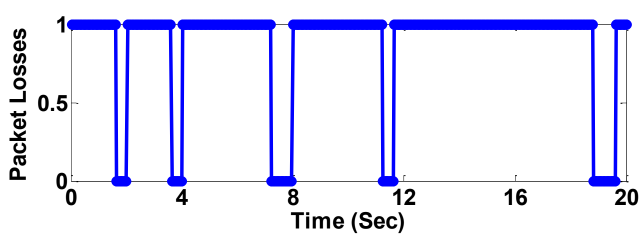


Fig 5. The distribution of packet losses ("1" means packet received, "0" means packet lost).

On the other hand, certainty equivalence deterministic NCS representation (32) can be calculated based on information of Batch reactor dynamics (61) and network imperfections (i.e. networked-induce delay and packet losses) as

$$z_{k+1} = \begin{bmatrix} 1.48 & 0.04 & 1.47 & -1.23 & 0.02 & -0.19 & 0.01 & -0.1 \\ -0.10 & 0.45 & -0.13 & 0.20 & 0.46 & 0.01 & 0.15 & 0 \\ 0.07 & 0.54 & -0.18 & 1.09 & 0.36 & -0.22 & 0.16 & -0.07 \\ -0.18 & 0.54 & -1.33 & 2.21 & 0.35 & -0.03 & 0.16 & -0.02 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} z_k$$

$$+\begin{bmatrix} 0.0056 & 0.2361 & 0.1338 & 0.1337 & 1 & 0 & 0 & 0 \\ -0.0636 & 0.0014 & -0.1170 & -0.0110 & 0 & 1 & 0 & 0 \end{bmatrix}^T u_k$$

where $z_k = [x_k \ u_{k-1} \ u_{k-2}]^T \in \mathbb{R}^{8\times 1}$.

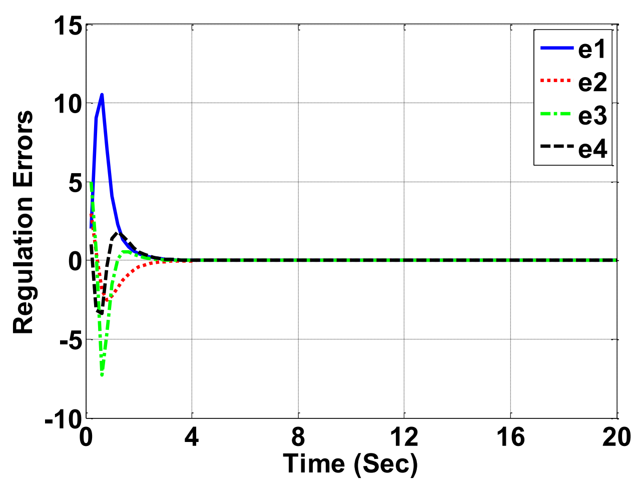


Fig 6. Performance of conventional stochastic optimal controller.

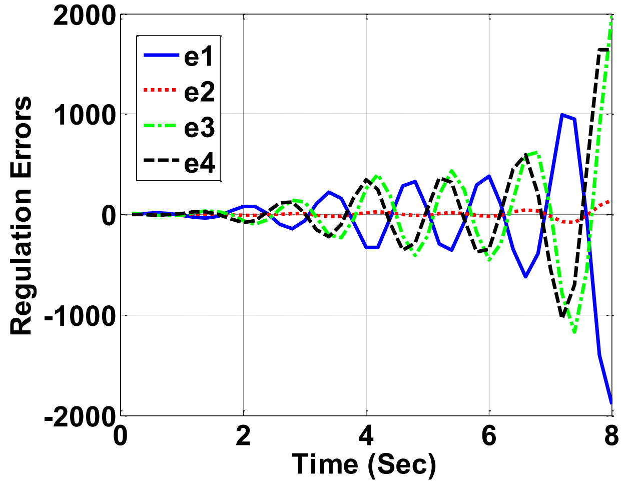
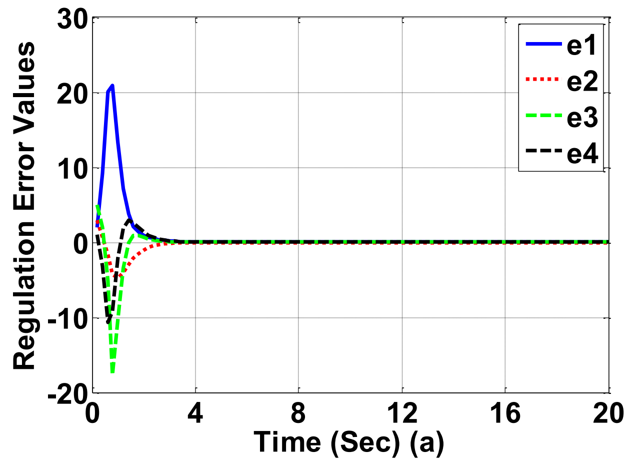


Fig7. State regulation errors of ADP value iteration (VI) controller when random delays and packet losses are present.

First, Figure 6 indicates that the stochastic optimal control of NCS with known dynamics and information of network imperfections (e.g. random delays and packet losses) obtained by solving the SRE backward-in-time. The controllers can make the state regulation errors converge to zero while ensuring the NCS stable in the mean when the delays and packet losses are accurately known.



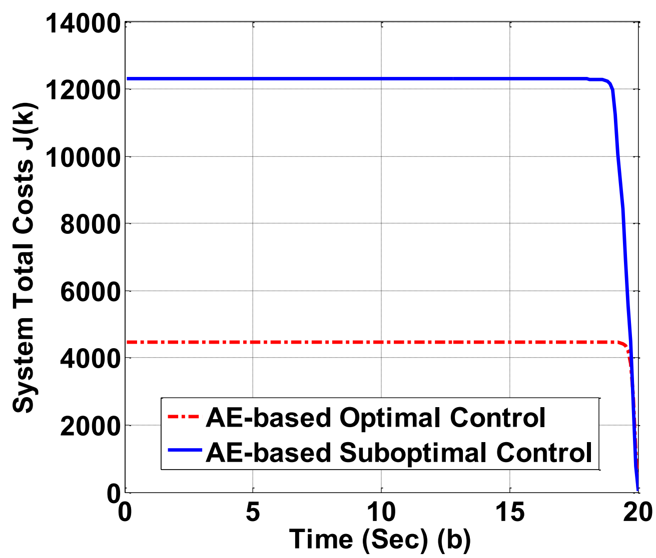


Fig 8. Performance of suboptimal and optimal controller for NCS with unknown dynamics: (a) State regulation errors with AE-based optimal control; (b) Comparison of system costs with AE-based optimal and suboptimal controllers.

Next, the ADP value iteration (VI) control (Al-Tamimi et.al., 2007) input $u_k = -\begin{bmatrix} 0.801 & 0.868 & -0.242 & 1.377 \\ -1.150 & 0.035 & -1.961 & 1.855 \end{bmatrix} x_k$ is designed by using policy iteration scheme. This ADP VI controller though does not require system dynamics cannot maintain batch reactor system stable in the mean in the presence of random delays and packet losses as shown in Figure 7.

Finally, the proposed adaptive stochastic optimal and stochastic suboptimal controller designs are implemented for the NCS with unknown system dynamics in presence of random delays and packet losses. The augment state z_k is generated as $z_k = [x_k \ u_{k-1} \ u_{k-2}]^T \in \mathbb{R}^{8\times 1}$ or $w = [z \ u] \in \mathbb{R}^{10\times 1}$. The initial stabilizing policy for the algorithm was selected as $u_0(z_k) = -\begin{bmatrix} 0.88 & 0.77 & -0.11 & 1.07 & 0.25 & 0.01 & 0.14 & 0.02 \\ -1.65 & -0.08 & -2.93 & 2.61 & -0.02 & 0.68 & -0.03 & 0.51 \end{bmatrix} z_k$ while the regression function for Q-function was generated as $\{w_1^2, w_1w_2, w_1w_3, \ldots, w_2^2, \ldots, w_9^2, \ldots, w_{10}^2\}$ as per (24).

The design parameter for Q-function $Q(z_k, u_k)$ was selected as $\alpha_h = 10^{-6}$ while initial parameters for the AE were set to zero at the beginning of the simulation. The initial parameters of the action control network were chosen to reflect the initial stabilizing control. The simulation was run for 200 times steps, and for the first 50 times steps, exploration noise with mean zero and variance 0.006 was added to the system at odd time steps and exploration noise with mean zero and variance 0.003 was added to the system at even time steps in order to ensure the persistency of excitation (PE) condition holds (Lemma 1).

In Figure 8 and 9, the performance of proposed AE-based optimal controller is evaluated. As shown in Figure 8(a), the proposed AE-based optimal controller can also make the NCS state regulation errors converge to zero even when the NCS dynamics are unknown which implies that the proposed controller can make the NCS closed-loop system stable in the mean. The cost-to-go function of proposed optimal and suboptimal controllers is compared in Figure 8(b) where the proposed AE-based optimal controller can minimize the cost-to-go ($J_k = \underset{\tau,\gamma}{E}[\sum_{m=k}^{\infty}(z_m^T Q_z z_m + u_m^T R_z u_m)]$) function more than proposed suboptimal controller based on certainty equivalence deterministic NCS model.

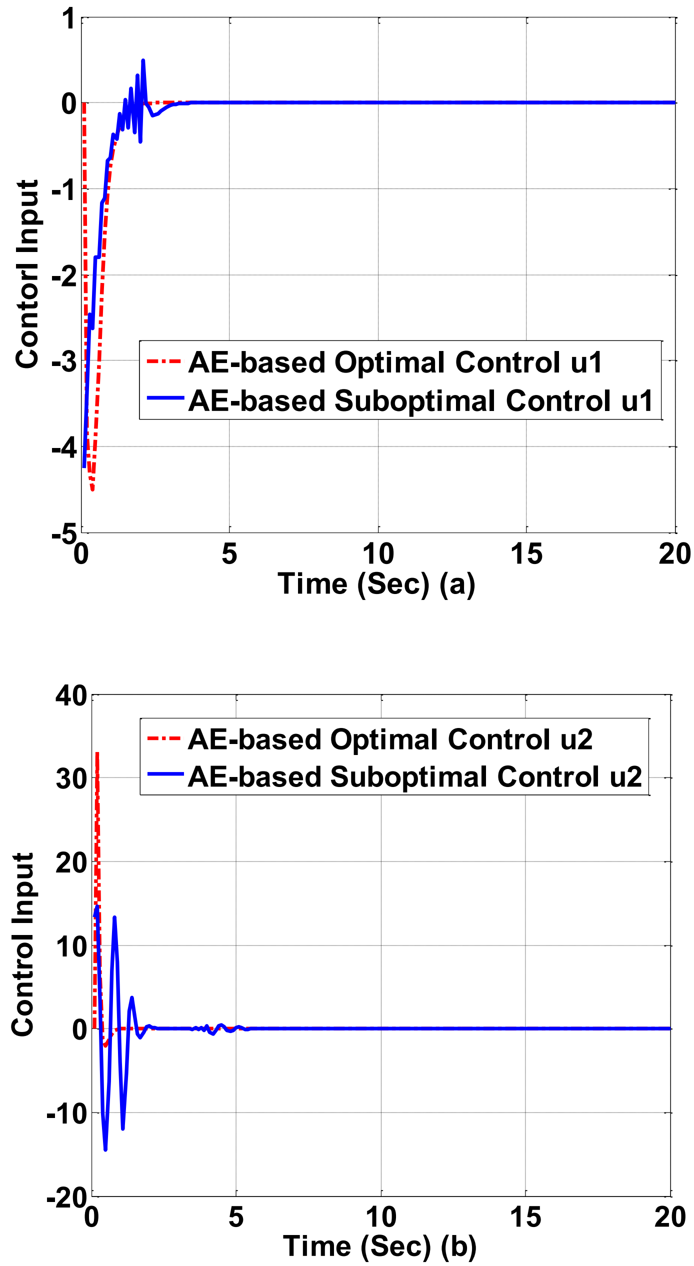


Fig 9. Comparison of control inputs with stochastic optimal and suboptimal controllers $u = (u_1\, u_2)^T \in \mathbb{R}^{2\times 1}$.

This clearly shows that the proposed AE-based optimal controller is more effective than suboptimal control based on certainty equivalence deterministic NCS representation. Now in Figure 9(a) (b), the control inputs of proposed AE-based optimal and suboptimal controllers are compared. The proposed AE-based optimal controller can force the NCS states converge to zero quicker than suboptimal control based on certainty equivalence deterministic NCS model. The proposed suboptimal controller has a smaller overshoot initially when compared to stochastic optimal controller.

According to the above results (Figures 6 through 9), the performance of proposed AE-based stochastic optimal and suboptimal controllers nearly has the same performance of stochastic optimal/suboptimal control with known system dynamics and wireless imperfections. The slightly higher overshoot observed at the beginning for the proposed optimal/suboptimal controller is due to an initial online learning phase needed to tune the optimal/suboptimal controller. After a short time, proposed AE-based stochastic optimal and suboptimal controllers will have similar performance even when NCS system dynamics and wireless imperfections are unknown.

5. Conclusions

In this work, we proposed an online adaptive dynamic programming technique based on AE to solve the stochastic optimal and suboptimal regulation control of NCS with uncertain dynamics in presence of unknown random delays and packet losses.

The availability of past state values ensure that NCS system dynamics were not needed when an AE generates an estimated *Q*-function and a novel stochastic optimal control law based on the estimation of $Q(z_k, u_k)$. An initial admissible control policy ensures that the system is stable in the mean while the AE learns *Q*-function $Q(z_k, u_k)$ and the matrix $\overline{H}_k$, stochastic cost function and optimal control signal. All AE parameters were tuned online using proposed update law and Lyapunov theory demonstrated the asymptotic stability of overall closed-loop system.

References

[1] Al-Tamimi, A., Lewis, F.L., and Abu-Khalaf, M. (2007). Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control. Automatica, vol. 43, pp. 473-481.

[2] Antsaklis, P. and Baillieul, J. (2004). Special issue on networked control systems. IEEE Trans. Automat. Contr., vol. 49, pp. 1421-1423.

[3] Åstrom, K.J. (1970). Introduction to stochastic control theory. Academic Press, New York.

[4] Azimi-Sadjadi, B. (2003). Stability of networked control systems in the presence of packet losses. in Proc. Conf. on Decision and Contr., pp. 676-681.

[5] Barto, A.G., Sutton, R.S., and Anderson, C.W. (1983). Neuron like elements that can solve difficult learning control problems. IEEE Trans. on Syst. Man and Cyber., vol.13, pp. 835-846.

[6] Bertsekas, D.P., and Shreve, S.E. (1978). Stochastic optimal control: the discrete time case, Academic Press, New York.

[7] Branicky, M.S., Phillips, S.M., and Zhang, W. (2000). Stability of networked control systems: explicit analysis of delay. in Proc. Amer. Control Conf., pp. 2352-2357.

[8] Busoniu, L., Babuska, R., and Schutter, B.D. (2010). Reinforcement learning and dynamic programming using function approximators. CRC Press, New York.

[9] Carnevale, D., Teel, A.R., and Nesic, D. (2007). A Lyapnov proof of improved maximum allowable transfer interval for networked control systems. IEEE Trans. Autom. Contr., vol. 52, pp. 892-897.

[10] Cloosterman, M.B.G, van de Wouw, N., Heemels, W.P.M.H., and Nijmeijer, H. (2009). Stability of networked control systems with uncertain time-varying delays. IEEE Tran. Automat. Contr., vol. 54, pp. 1575-1580.

[11] Dierks, T., and Jagannathan, S. (2009). Optimal control of affine nonlinear discrete-time systems with unknown internal dynamics. in Proc. Conf. on Decision and Contr., pp. 6750-6755.

[12] Dierks, T., Thumati, T. B., and Jagannathan, S., (2009). Adaptive dynamics programming-based optimal control of unknown affine nonlinear discrete-time systems. in Proc. Intern. Joint Conf. Neur. Netwo., vol.711-716.

[13] Franklin, G.F., Powell, J.D., and Emani-Naeini, A. (1994). Feedback control of dynamic systems, 3rd Edition, Addison-Wesley, Reading, Massachusetts.

[14] Goldsmith, A. (2003). Wireless Communications. Cambridge University Press, Cambridge, U.K.

[15] Green, M., and Moore, J.B. (1986). Persistency of excitation in linear systems. Syst. and Contr. Lett., vol. 7, pp. 351-360.

[16] Guo, L. (1990). Estimating time-varying parameters by the Kalman filter based algorithm: stability and convergence. IEEE Trans. Automat. Contr., vol.35, pp. 141-147.

[17] Halevi, Y., and Ray, A. (1988). Integrated communication and control systems: part I—analysis. J. Dynamic Syst., Measure. Contr., vol. 110, pp. 367-373.

[18] Heemels, W.P.M.H, Teel, A.R., van de Wouw, N., and Nesic, D. (2010). Networked control systems with communication constraints: tradeoff between transmission intervals, delays and performance. IEEE Trans. Automat. Contr., vol. 55, pp. 1781-1796

[19] Hesphanha, J.P., Naghshtabrizi, P., and Xu, Y. (2007). A survey of recent results in networked control system. Proc. IEEE, vol. 95, pp. 138-162.

[20] Hu, S.S., and Zhu, Q.X. (2003). Stochastic optimal control and analysis of stability of networked control systems with long delay. Automatica, vol. 39, pp. 1877-1884.

[21] Jagannathan, S. (2006). Neural network control of nonlinear discrete-time systems, CRC Press.

[22] Lewis, F.L., and Syrmos, V.L. (1995). Optimal control, 2nd Edition, Wiley, New York.

[23] Lian, F., Moyne, J. and Tilbury, D. (2003). Modeling and optimal controller design of networked control systems with multiple delays. Int. J. Contr., vol. 76, pp. 591-606.

[24] Liou, L.W., and Ray, A. (1991). A stochastic regulator for integrated communication and control systems: part I—formulation of control law. ASME Jour. of Dynamic Syst., Measureme. and Contr., vol. 4, pp. 604-611.

[25] Liu, X.H., and Goldsmith, A. (2004). Wireless medium access control in networked control systems. in Proc. Ameri. Control Conf., pp. 688-694.

[26] Maybeck, P.S. (1982), Stochastic models, estimation and control. Academic press, New York.

[27] Middleton, R.H., and Goodwin, G.C. (1988). Adaptive control of time-varying linear systems. IEEE Trans. Automat. Contr., vol. 33, pp. 150-155.

[28] Nilsson, J., Bernhardsson, B., and Wittenmark, B. (1998). Stochastic analysis and control of real-time systems with random time delays. Automatica, vol. 34, pp. 57-64.

[29] Papoulis, A. (1991). Probability, random variables, and stochastic processes, 3rd Edition, New York: McGraw-Hill.

[30] Sastry, S. and Bodson, M. (1989). Adaptive control: stability, convergence, and robustness. Englewood Cliffs, NJ: Prentice Hall.

[31] Schenato, L., Sinopoli, B., Franceschetti, M., Poolla, K., and Sastry, S. (2007). Foundations of control and estimation over lossy networks. Proc. IEEE, vol. 95, pp. 163-187.

[32] Stengel, R.F. (1986). Stochastic optimal control, theory and application, John Wiley and Sons, New York.

[33] Walsh, G.C., Ye, H., and Bushnell, L. (1999). Stability analysis of networked control systems. in Proc. Amer. Control Conf., pp. 2876-2880.

[34] Watkins, C., (1989). Learning from delayed rewards, Ph.D. Thesis, Cambridge,England: Cambridge University.

[35] Werbos, P.J. (1991). A menu of designs for reinforcement learning over time. MA: MIT Press.

[36] Werbos, P.J. (1992). Approximate dynamic programming for real-time control and neural modeling. Handbook of Intelligent Control, New York, Van Nostrand Reinhold.

[37] Wonham, W.M. (1968). On a matrix Riccati equation of stochastic control. SIAM J. Contr., vol.6, pp. 681-697.

[38] Wu, J. and Chen, T. (2007). Design of networked control systems with packet dropouts. IEEE Trans. Automat. Control, vol. 52, pp. 1314-1319.

[39] Zhang, W., Branicky, M.S., and Phillips, S. (2001). Stability of networked control systems. IEEE Contr. Syst. Mag., vol. 21, pp. 84-99.

[40] Zhang, H., Luo, Y., and Liu, D. (2009), Neural network based near optimal control for a class of discrete-time affine nonlinear system with control constraints. IEEE Trans. Neur. Netwo., vol. 20, pp. 1490-1503.

Appendix

Proof for Lemma 1. The exploration noise vector which obeys normal distribution with variances changing over time is added (i.e. $\mathbf{n}^{pe} \sim \mathcal{N}(0, \sigma_k^2 \mathbf{I})$).

For given positive constant δ and $k_0 \geq 1$, exploration noise vectors can be added to system, in turn vector w_k is as

$$w_k = [z_k^T \ u^T(z_k)]^T + (\mathbf{n}_k^{pe})^T \quad k = k_0, ..., k_0 + \delta, ... \tag{A.1}$$

where $\mathbf{n}_k^{pe} = [n_k^{pe}(1) \ n_k^{pe}(2) \ ... \ n_k^{pe}(l)] \in \mathbb{R}^{n+(\bar{d}+1)m=l}$ is a row vector and $n_k^{pe}(i)$ represent ith scalar element in the row vector.Without loss of generality, the Kronecker product quadratic polynomial stochastic independent basis vector $\overline{w}_k$ can be expressed $\overline{w}_k = [w_k^2(1), w_k(1) \times w_k(2), ..., w_k(l-1) \times w_k(l), \ w_k^2(l)]^T$. Since the exploration noise obeys the normal distribution with zero mean and is independent over time, and also independent with vector w_k $\forall k$. We have

$$E\left(n_k^{pe}(i) w_k(j)\right) = \begin{cases} 0 & i \neq j \\ \sigma_k^2 & i = j \end{cases} \quad \forall i, j = 0,1,2,...,l \tag{A.2a}$$

$$E\left(n_k^{pe}(i) n_{k-1}^{pe}(j)\right) = E(n_k^{pe}(i)) E(n_{k-1}^{pe}(j)) \ \forall i, j = 0,1,2,...,l \tag{A.2b}$$

According to the definition of ΔW_k, $E[\Delta W_k (\Delta W_k)^T] \ \forall k = k_0, k_0 + 1, ..., k_0 + \delta$ satisfies the inequality

$$\begin{aligned} E[\Delta W_k (\Delta W_k)^T] &\geq E[(\overline{\mathbf{n}}_k^{pe} - \overline{\mathbf{n}}_{k-1}^{pe})(\overline{\mathbf{n}}_k^{pe} - \overline{\mathbf{n}}_{k-1}^{pe})^T] \\ &\geq (\sigma_k^2 - \sigma_{k-1}^2)^2 \mathbf{I}_l \end{aligned} \tag{A.3}$$

where $\overline{\mathbf{n}}_k^{pe}$ are the Kronecker product quadratic polynomial stochastic indepdent basis vector based on exploration noise vector $\mathbf{n}_k^{pe}$, and $\mathbf{I}_l$ is $l \times l$ identity matrix.

Since $\sigma_k^2 \neq \sigma_{k-1}^2$, then there exists positive constant $\alpha_k = (\sigma_k^2 - \sigma_{k-1}^2)^2$ such that

$$\sum_{k=k_0}^{k_0+\delta} E[\Delta W_k (\Delta W_k)^T] \geq \sum_{k=k_0}^{k_0+\delta} (\sigma_k^2 - \sigma_{k-1}^2)^2 \mathbf{I}_l = \sum_{k=k_0}^{k_0+\delta} \alpha_k \mathbf{I}_l = \alpha \mathbf{I}_l \tag{A.4}$$

In the other words,

$$\sum_{k=k_0}^{k_0+\delta} E[\Delta W_k (\Delta W_k)^T] \geq \alpha \mathbf{I}_l \tag{A.5}$$

where $\alpha = \sum_{k=k_0}^{k_0+\delta} \alpha_k > 0$. Therefore, when exploration noise is added to the polynomial stochastic independent basis vector ΔW_k, the PE condition is satisfied.

On the other hand, based on the definition of Ω_k, we have

$$\sum_{k=k_0}^{k_0+\delta} E[\Omega_k^T \Omega_k] \geq \sum_{k=k_0}^{k_0+\delta} \begin{bmatrix} E[\Delta W_k^T \Delta W_k] & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & E[\Delta W_{k-j}^T \Delta W_{k-j}] \end{bmatrix}$$

$$\geq \sum_{k=k_0}^{k_0+\delta} \begin{bmatrix} \alpha_k^2 \mathbf{I}_l & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_{k-j}^2 \mathbf{I}_l \end{bmatrix} \geq \sum_{k=k_0}^{k_0+\delta} \rho_k^2 \mathbf{I}_{a=l\times j} \geq \rho^2 \mathbf{I}_a$$

where $\rho_k = \min\{\alpha_k, \alpha_{k-1}, \ldots, \alpha_{k-j}\}$ and $\rho^2 = \sum_{k=k_0}^{k_0+\delta} \rho_k^2$. Therefore, the PE condition of Ω_k is also satisfied.

Proof for Lemma 2. Using the Bellman equation and Q-learning with estimated stochastic cost function and matrix $\overline{H}_k$, we have

$$\hat{J}_k + e_{hk} = r(z_k, u(z_k)) + \hat{J}_{k+1} \tag{A.6}$$

Now consider (A.6) and

1) The left part of (A.6) can be expressed as

$$\begin{aligned}\hat{J}_k + e_{hk} = \hat{\bar{h}}_k^T \bar{w}_k + e_{hk} &= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \hat{\bar{H}}_k \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} + e_{hk} \\ &= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \hat{\bar{H}}_k^{zz} & \hat{\bar{H}}_k^{zu} \\ \hat{\bar{H}}_k^{uz} & \hat{\bar{H}}_k^{uu} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} + e_{hk}\end{aligned} \tag{A.7}$$

2) The right side of (A.6) can be shown as

$$r(z_k, u(z_k)) + \hat{J}_{k+1} = r(z_k, u(z_k)) + \underset{\tau,\gamma}{E}\left(w_{k+1}^T \hat{H}_{k+1} w_{k+1}\right) = r(z_k, u(z_k)) + \underset{\tau,\gamma}{E}\left(x_{k+1}^T \hat{P}_{k+1} x_{k+1}\right)$$

$$\begin{aligned}&= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} S_z & 0 \\ 0 & R_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} \\ &+ \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} \\ &= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}\end{aligned} \tag{A.8}$$

According to (A.7) and (A.8), (A.6) can be derived as

$$\begin{aligned}&\begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \hat{\bar{H}}_k^{zz} & \hat{\bar{H}}_k^{zu} \\ \hat{\bar{H}}_k^{uz} & \hat{\bar{H}}_k^{uu} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} + e_{hk} \\ &= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}\end{aligned}$$

Hence,

$$\begin{aligned}&\begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \hat{\bar{H}}_k^{zz} & \hat{\bar{H}}_k^{zu} \\ \hat{\bar{H}}_k^{uz} & \hat{\bar{H}}_k^{uu} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} = \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \\ &\begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix} - e_{hk} \\ &= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}\end{aligned}$$

$$-\begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \frac{e_{hk}}{tr\{z_k^T z_k\}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}^T \begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z - \frac{e_{hk}}{tr\{z_k^T z_k\}} & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \end{bmatrix}$$

and

$$\begin{bmatrix} \hat{\overline{H}}_k^{zz} & \hat{\overline{H}}_k^{zu} \\ \hat{\overline{H}}_k^{uz} & \hat{\overline{H}}_k^{uu} \end{bmatrix} = \begin{bmatrix} \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z - \frac{e_{hk}}{tr\{z_k^T z_k\}} & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{bmatrix} \tag{A.9}$$

where

$$\begin{cases} \hat{\overline{H}}_k^{zz} = \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) + S_z - \frac{e_{hk}}{tr\{z_k^T z_k\}} \\ \hat{\overline{H}}_k^{zu} = \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) \\ \hat{\overline{H}}_k^{uz} = \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) \\ \hat{\overline{H}}_k^{uu} = \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) + R_z \end{cases} \tag{A.10}$$

According to the estimation optimal control law (11), we have $\hat{u}_{1k} = -(\hat{\overline{H}}_k^{uu})^{-1} \hat{\overline{H}}_k^{uz} z_k$, which is expressed by using (23) as

$$\hat{u}_{1k} = -\underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk} + R)^{-1} \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) z_k \tag{A.11}$$

At the same time, according to the optimal control theory (Lewis and Syrmos, 1995) and (22) we know $\hat{J}_{k+1} = \underset{\tau,\gamma}{E}(z_{k+1}^T \ \hat{P}_{k+1} z_{k+1})$. Therefore, we can minimize of the stochastic cost function to get the optimal control as

$$
\begin{aligned}
\hat{u}_{2k} &= -\frac{1}{2} R_z^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \right) = -R_z^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} z_{k+1} \right) \\
&= -R^{-1} \mathop{E}_{\tau,\gamma}\left[B_{zk}^T \hat{P}_{k+1} \left(A_{zk} z_k + B_{zk} \hat{u}_{2k} \right) \right] \\
&= -R^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} A_{zk} \right) z_k - R^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} B_{zk} \right) \hat{u}_{2k}
\end{aligned}
\tag{A.12}
$$

The term $\hat{u}_{2k}$ can be solved by (15) as

$$
\left(I + R_z^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} B_{zk} \right) \right) \hat{u}_{2k} = -R_z^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} A_{zk} \right) z_k
$$

$$
\left(R_z + \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} B_{zk} \right) \right) \hat{u}_{2k} = -\mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} A_{zk} \right) z_k
$$

$$
\begin{aligned}
\hat{u}_{2k} &= -\left(R + \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} B_{zk} \right) \right)^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} A_{zk} \right) z_k \\
&= -\mathop{E}_{\tau,\gamma}\left(R + B_{zk}^T \hat{P}_{k+1} B_{zk} \right)^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} A_{zk} \right) z_k
\end{aligned}
\tag{A.13}
$$

According to (A.11) and (A.13), we have

$$
\begin{aligned}
\hat{u}_{1k} &= -\left(\hat{\overline{H}}_k^{uu} \right)^{-1} \hat{\overline{H}}_k^{uz} z_k \\
&= -\mathop{E}_{\tau,\gamma}\left(R + B_{zk}^T \hat{P}_{k+1} B_{zk} \right)^{-1} \mathop{E}_{\tau,\gamma}\left(B_{zk}^T \hat{P}_{k+1} A_{zk} \right) z_k = \hat{u}_{2k}
\end{aligned}
\tag{A.14}
$$

Therefore, $\tilde{u}_k = \hat{u}_{1k} - \hat{u}_{2k} = 0$ since $\hat{u}_{1k} = \hat{u}_{2k}$.

Proof for Theorem 2. Consider the following positive definite Lyapunov function candidate

$$
V = V_D\left(z_k\right) + V_J(\tilde{\overline{h}}_k) \tag{A.15}
$$

where $V_D\left(z_k\right)$ is defined in (30) and $V_J\left(\tilde{\overline{h}}_k\right)$ is defined as

$$
V_J(\tilde{\overline{h}}_k) = (\tilde{\overline{h}}_k \overline{w}_k - \tilde{\overline{h}}_k \overline{w}_{k-1})^2 = (\tilde{\overline{h}}_k \Delta W_{k-1})^2 \tag{A.16}
$$

The first difference of (A.15) can be expressed as $\Delta V = \Delta V_D\left(z_k\right) + \Delta V_J(\tilde{\overline{h}}_k)$, and considering that $\Delta V_J(\tilde{\overline{h}}_k) = (\tilde{\overline{h}}_{k+1} \Delta W_k)^2 - (\tilde{\overline{h}}_k \Delta W_{k-1})^2$ with the AE, we have

$$\Delta V_J(\tilde{\bar{h}}_k) = (\tilde{\bar{h}}_{k+1}\Delta W_k)^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2$$

$$= (\alpha_h \tilde{\bar{h}}_k \Delta W_{k-1})^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2$$

$$= -(1-\alpha_h^2)(\tilde{\bar{h}}_k \Delta W_{k-1})^2$$

$$\leq -(1-\alpha_h^2)\|\Delta W_{k-1}\|^2 \left\|\tilde{\bar{h}}_k\right\|^2 \quad \text{(A.17)}$$

Next, considering the first part $\Delta V_D(z_k) = z_{k+1}^T z_{k+1} - z_k^T z_k$ and applying the NCS and Cauchy-Schwartz inequality reveals

$$\begin{aligned}\Delta V_D(z_k) &\leq \|A_{zk}z_k + B_{zk}u_k - B_{zk}\tilde{u}_k\|^2 - z_k^T z_k \\ &\leq 2\|A_{zk}z_k + B_{zk}u_k\|^2 + 2\|B_{zk}\tilde{u}_k\|^2 - z_k^T z_k\end{aligned} \quad \text{(A.18)}$$

Applying the Lemma 3 (bounds on the optimal closed loop system in (29)) and recalling $\hat{u}_{1k} = \hat{u}_{2k}$ from Lemma 1 and (A.14), we know

$$\tilde{u}_k = \hat{u}_{1k} - \hat{u}_{2k} = -\left(\hat{\bar{H}}_k^{uu}\right)^{-1} \hat{\bar{H}}_k^{uz} z_k + \frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} = 0 \quad \text{(A.19)}$$

Therefore, $\Delta V_D(z_k)$ is expressed in terms as the adaptive estimator (AE) error dynamics of the matrix $\bar{H}_k$ and the relationship between $Q(z_k, u_k), \tilde{\bar{h}}_k$ and $\tilde{J}_k$, (A.18) revealing

$$\Delta V_D(z_k) \leq -\left(1-2k^*\right)\|z_k\|^2 + 2\|B_{zk}\tilde{u}_k\|^2$$

$$\leq -\left(1-2k^*\right)\|z_k\|^2 + 2B_M^2 \left\| -\left(\hat{\bar{H}}_k^{uu}\right)^{-1} \hat{\bar{H}}_k^{uz} z_k + \frac{1}{2} R^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \right\|^2$$

$$\leq -\left(1-2k^*\right)\|z_k\|^2 \quad \text{(A.20)}$$

At final step, combining the equation (A.18) and (A.20), we have

$$\Delta V \le -\left(1-2k^*\right)\|z_k\|^2 - \left(1-\alpha_h^2\right)\|\Delta W_{k-1}\|^2 \left\|\tilde{\tilde{h}}_k\right\|^2 \tag{A.21}$$

Since $0 < k^* < 1/2$ and $0 < \alpha_h < 1$, ΔV is negative definite (See Lemma 1 on PE condition) and V is positive definite. Note that $\left|\sum_{k=k_0}^{\infty} \Delta V_k\right| = |V_\infty - V_0| < \infty$ since $\Delta V < 0$ as long as (A.21) holds. Therefore, system state z_k and $\tilde{\tilde{h}}_k$ are all asymptotically stable in the mean. In other words, as $k \to \infty, z_k \to 0$, $\tilde{\tilde{h}}_k \to 0$, then $\hat{J}_k \to J_k$. Since optimal control $u_k^* = \frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial J_{k+1}}{\partial z_{k+1}}$, $\hat{u}_{2k} = \frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}}$ and $\hat{u}_{1k} = \hat{u}_{2k}$ (Lemma 2), then $\hat{u}_{2k} \to u_k^*$ and $\hat{u}_{1k} \to u_k^*$ when $\hat{J}_k \to J_k$.

Proof for Theorem 3. Consider the Lyapunov function candidate as

$$V^S = V_D^S(z_k) + V_J^S(\tilde{h}_k) \tag{A.22}$$

where $V_D^S(z_k) = z_k^T z_k$ and $V_J^S(\tilde{h}_k) = (\tilde{h}_k \Delta W_{k-1})^2$ with $\tilde{h}_k = h - \hat{h}_k$.

The first difference of (A.22) can be expressed as $\Delta V^S = \Delta V_D^S(z_k) + \Delta V_J^S(\tilde{\tilde{h}}_k)$, and considering that $\Delta V_J^S(\tilde{h}_k) = (\tilde{h}_{k+1}\Delta W_k)^2 - (\tilde{h}_k \Delta W_{k-1})^2$ with AE, we have

$$\begin{aligned}\Delta V_J(\tilde{h}_k) = (\tilde{h}_{k+1}\Delta W_k)^2 - (\tilde{h}_k \Delta W_{k-1})^2 \quad &= -(1-\alpha_s^2)(\tilde{h}_k \Delta W_{k-1})^2 \\ &\le -(1-\alpha_s^2)\|\Delta W_{k-1}\|^2 \left\|\tilde{h}_k\right\|^2\end{aligned} \tag{A.23}$$

Now considering $\Delta V_D^S(z_k) = z_{k+1}^T z_{k+1} - z_k^T z_k$ and deploying the NCS (32) and applying Cauchy-Schwartz inequality reveals

$$\begin{aligned}\Delta V_D^S(z_k) &\le \left\|A_z z_k + B_z u_k - B_z \tilde{u}_k\right\|^2 - z_k^T z_k \\ &\le 2\left\|A_z z_k + B_z u_k\right\|^2 + 2\left\|B_z \tilde{u}_k\right\|^2 - z_k^T z_k\end{aligned} \tag{A.24}$$

By using *Q*-learning, we know

$$\tilde{u}_k = -\left(\hat{H}_k^{uu}\right)^{-1}\hat{H}_k^{uz}z_k + \frac{1}{2}R^{-1}B_z^T\frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} = 0 \tag{A.25}$$

Therefore, the first difference of (A.22), ΔV^S, can be given as

$$\Delta V \le -\left(1-2k^*\right)\left\|z_k\right\|^2 - \left(1-\alpha_s^2\right)\left\|\Delta W_{k-1}\right\|^2\left\|\tilde{h}_k\right\|^2 \tag{A.26}$$

Since $0 < k^* < 1/2$ and $0 < \alpha_s < 1$, ΔV is negative definite and V is positive definite. Note that $\left|\sum_{k=k_0}^{\infty}\Delta V_k\right| = \left|V_\infty - V_0\right| < \infty$ since $\Delta V < 0$ as long as (A.26) holds. Therefore, system state z_k and $\tilde{h}_k$ are all asymptotically stable in the mean. In other words, as $k \to \infty, z_k \to 0$, $\tilde{h}_k \to 0$, then $\hat{J}_k \to J_k$. Since suboptimal control is given by $u_k^* = \frac{1}{2}R_z^{-1}B_z^T\frac{\partial J_{k+1}}{\partial z_{k+1}}$ and $\hat{u}_k = -(\hat{H}_k^{uu})^{-1}\hat{H}_k^{uz}z_k = \frac{1}{2}R_z^{-1}B_z^T\frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}}$ (Lemma 2), then it follows that $\hat{u}_k \to u_k^*$ when $\hat{J}_k \to J_k$.

PAPER II

STOCHASTIC OPTIMAL DESIGN FOR UNKNOWN LINEAR DISCRETE-TIME SYSTEM ZERO-SUM GAMES UNDER COMMUNICATION CONSTRAINTS

H. Xu and S. Jagannathan

Abstract - In this paper, stochastic optimal control strategy for unknown linear discrete-time system quadratic zero-sum games with communication imperfections (e.g. network-induced delays and packet losses), or referred to as networked control system (NCS) zero-sum games, related to H_∞ *optimal control problem is solved in forward-in-time manner. The proposed stochastic optimal approach, referred to as adaptive dynamic programming (ADP), estimates the cost or value function to solve the infinite horizon optimal regulation of unknown linear discrete-time system quadratic zero-sum games in the presence of network imperfections and subsequently optimal control and worst case disturbance inputs are derived based on the estimated value function. Update law for tuning the unknown parameters of the value function estimator is derived and Lyapunov theory is used to show that all signals are asymptotic stable (AS) in the mean and that the estimated control and disturbance signals converge to optimal control and disturbance inputs in the mean respectively. Simulation results are included to verify the theoretical claims.*

Key words— Linear discrete-time system, Networked control system, Adaptive estimation, Optimal control, Zero-sum games.

1. Introduction

Feedback control systems with control loops closed through a real-time network are called networked control system (NCS) (Halevi and Ray, 1988). In NCS, a communication packet carries the reference and control inputs, and plant outputs within a network among control system components such as sensors, controllers, and actuators. Though advantages of NCS are reduced system wiring, ease of system diagnosis and maintenance, and increased system agility, however, insertion of the communication network in the feedback loop brings many issues which have to be addressed before these benefits can be harvested.

First issue being the network-induced delay that occurs while exchanging data among devices connected to the shared wireless communication network. This delay, either constant or random, can degrade the performance of control system and even destabilize the linear system when the delay is not explicitly considered in design process. Second issue is the packet losses due to unreliable wireless communication network transmission which can cause a loss in control input resulting in instability. Therefore, recently Walsh et al. (1999) and Lian et al. (2001) consider stability and performance of NCS with constant delays. Krtolica et al. (1994) analyzes the stability of NCS with random delays while Wu and Chen (2007) study the stability of NCS with packet losses. Eventually, Zhang et al. (2001) conducts the stability analysis of NCS with communication imperfections (e.g. network delays and packet losses) and proposed a stability region.

On the other hand, Lian et al. (2003) introduces the optimal controller design (Lewis and Syrmos, 1995) for NCS without taking into account the disturbance input. By contrast, using the stochastic optimal control theory (Åstrom, 1970), Nilsson et al. (1998) introduces the optimal and suboptimal control design for linear NCS with random delays. Although these optimal and suboptimal controller designs have resulted in satisfactory performance, the NCS system dynamics and information on communication imperfections (e.g. network-induced delays and packet losses), which are difficult to know beforehand, need to be known accurately for the controller design while the optimality is achieved backward-in-time. However, these designs did not consider the optimality of the unknown NCS quadratic zero-sum games (Basar and Olsder, 1995; Littman 1994).

In contrast, adaptive/approximate dynamic programming (ADP) schemes proposed by Werbos (1990) and Watkins (1989), intend to solve optimal control design in forward-in-time manner for unknown nonlinear systems in contrast with traditional optimal control techniques (Lewis and Syroms, 1995) which work backward-in-time for known system dynamics. In ADP, one combines adaptive critics, a reinforcement learning technique, with dynamic programming where the optimal control is obtained through value and policy iterations. Recently, Tamimi et al. (2007) employs the *Q*-learning method to solve the optimal strategy for discrete-time linear time-invariant system quadratic zero-sum games without using the system matrices. Though the value and policy iteration-based approach works forward-in-time for optimal control (Tamimi et al. 2007) but it requires a large number of iterations within a sampling interval for convergence which can be a bottleneck for real-time control. Moreover, convergence of the algorithm is only shown while the stability of the overall system is not given.

By contrast, Dierks and Jagannathan (2009) uses two neural networks (NN) to solve the Hamilton-Jacobi-Bellman (HJB) equation forward-in-time for the optimal control of a class of general unknown nonlinear affine discrete-time systems. In this approach, value and policy iterations are not utilized; instead the value function and control policies are updated once per sampling interval by using past history of residual errors thus making the technique suitable for real-time control. However, these ADP-techniques are not suitable for NCS since they ignore the effects of communication imperfections (e.g. network-induced delays and packet losses). These communication imperfections can make the optimal design more involved (Tamimi et al., 2007) and cause instability (Zhang et al. 2001) if they are not properly accounted for.

Therefore, in this paper a time-based adaptive dynamic programming approach is undertaken to the stochastic optimal regulation of linear NCS quadratic zero-sum games with unknown system dynamics and communication imperfections (i.e. network-induced delays and packet losses) in order to solve the Bellman equation (Wonham, 1968) online and in forward-in-time manner. Using an initial stabilizing control, the value function is estimated online adaptively (Jagannathan, 2006) while its unknown parameters are tuned by using a novel update law since solving the Game Theoretic Riccati Equation (GRE) requires the system matrices. Then, using the idea of dynamic programming, the optimal control and worst case disturbance inputs which optimize the cost function can be calculated based on the information provided by the estimated value function. Thus the proposed time-based ADP scheme relaxes the need for system dynamics and information on communication imperfections (i.e. delay and packet losses) and it renders optimal solution without using value and policy iterations. Finally, the overall stability of the closed-loop system is demonstrated by using Lyapunov theory.

The importance of the paper stems from the fact that a game-theoretic adaptive system is proposed to create controllers for NCS quadratic zero-sum games that learn to co-exist with a L_2 -gain disturbance signal (Basar and Bernhard, 1995; Dragan and Morozan, 1997). In the control system design, this problem is defined as a two-player game that corresponds to the well-known H_∞ control. Next some background information is introduced.

2. Background

2.1. Linear NCS Quadratic Zero-sum Games

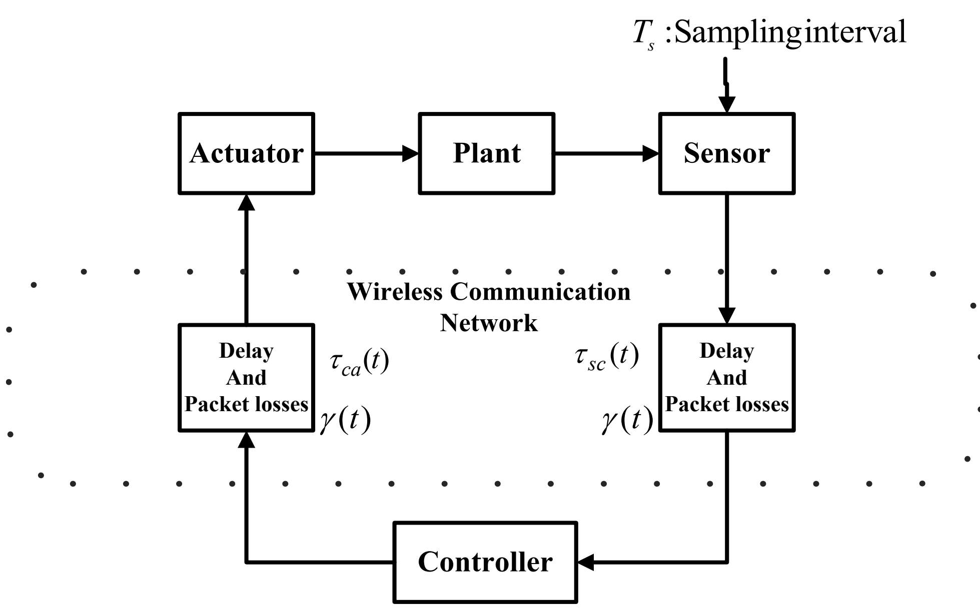


Fig 1. Networked Control System (NCS).

The basic structure of NCS considered in this paper is shown as Figure 1 where the feedback control loop is closed over a communication network and in particular a wireless communication network. Since wireless communication network is bandwidth limited, two types of network-induced delays and one type of packet losses are included in this model: (1) $\tau_{sc}(t)$: sensor-to-controller delay, (2) $\tau_{ca}(t)$: controller-to-actuator delay and (3) $\gamma(t)$: indicator of packet received at actuator. The following assumption is needed similar to other works (Liou and Ray, 1991; Hu and Zhu, 2003):

Assumption 1:

a). Sensor is time-driven while the controller and actuator are event-driven (Goldsmith, 2005).

b). Communication network is a wide area wireless communication network so that two types of network-induced delays are independent, ergodic and unknown while their probability distribution functions are assumed known (Liou and Ray, 1991; Hu and Zhu, 2003).

c). The sum of sensor-to-controller delay and controller-to-actuator delay is bounded (Liou and Ray, 1991) while the initial state of linear system is deterministic (Hu and Zhu, 2003).

Remark 1: The definition of “event-driven” implies that an action is taken in response to an event which may not be generated uniformly in time. In linear NCS quadratic zero-sum games, control and disturbance signals can be considered as the signals generated in response to the feedback sensor inputs; similarly, the actuator applies the control and disturbance inputs to the plant in response to the controller output. Since both a controller and an actuator respond upon receiving an event, they are referred to as “event-driven” controller and actuator in the NCS quadratic zero-sum game.

Consider the following linear time-invariant system with communication imperfections (i.e. network-induced delays and packet losses) which is given by

$$\dot{x}(t) = Ax(t) + \gamma(t)Bu(t-\tau(t)) + \gamma(t)Dd(t-\tau(t)) \tag{1}$$

where $\gamma(t) = \begin{cases} \mathbf{I}^{n\times n} & \text{if the control input is received at time } t \\ \mathbf{0}^{n\times n} & \text{if the control input lost at time } t \end{cases}$,with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $d(t) \in \mathbb{R}^l$ represent the system state, control input and disturbance vectors respectively, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{n\times m}$ denote the system matrices. From Assumption 1, it can be deduced that the sum of networked-induced delays is bounded above such that $\tau(t) = \tau_{sc}(t) + \tau_{ca}(t) < bT_s$ where b represents the delay bound while T_s being the sampling interval.

During a sampling interval $[kT_s, (k+1)T_s)\ \forall k$, the controller input $u(t)$ and disturbance input $d(t)$ to the plant are piecewise constants. According to Assumption 1, there are at most b number of current and previous control and disturbance inputs that can be received at the actuator. If many control and disturbance inputs are received at the same time, only the newest control and disturbance inputs are allowed to act on the controlled plant during any sampling interval $[kT_s, (k+1)T_s)\ \forall k$, and other previous control and disturbance inputs are deduced. Since control and disturbance inputs are based on event driven, the plant will implement control and disturbance inputs at these time instants $kT_s + t_i^k, i = 0,1,...,b$ and $t_i^k < t_{i-1}^k$ where $t_i^k = \tau_i^k - iT_s$ as illustrated in Figure 2 (Liou and Ray, 1991; Hu and Zhu, 2003).

For the event-driven controller and disturbance, the control and disturbance inputs become u_k, d_k in response to sensed signal x_k. Integration of (1) over a sampling interval $\left[kT_s, (k+1)T_s\right) \forall k$ yields

$$x_{k+1} = A_s + \sum_{i=0}^{b} \gamma_{k-i} B_i^k u_{k-i} + \sum_{i=0}^{b} \gamma_{k-i} D_i^k d_{k-i} \tag{2}$$

where $x_k = x(kT), A_s = e^{AT}$, $B_0^k = \int_{\tau_0^k}^{T} e^{A(T-s)} ds B \bullet \mathbf{1}\left(T - \tau_0^k\right)$,

$B_i^k = \int_{\tau_i^k - iT}^{\tau_{i-1}^k - (i-1)T} e^{A(T-s)} ds B \bullet \mathbf{1}\left(T + \tau_{i-1}^k - \tau_i^k\right)$
$\bullet \mathbf{1}\left(\tau_i^k - iT\right) \forall i = 1,2,\ldots,b;\; D_i^k = \int_{\tau_i^k - iT}^{\tau_{i-1}^k - (i-1)T} e^{A(T-s)} ds D \bullet \mathbf{1}\left(T + \tau_{i-1}^k - \tau_i^k\right) \bullet \mathbf{1}\left(\tau_i^k - iT\right) \; \forall i = 1,2,\ldots,b$
$D_0^i = \int_{\tau_0^k - kT}^{T} e^{A(T-s)} ds D \bullet \mathbf{1}\left((k+1)T - \tau_0^k\right);\; \mathbf{1}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$,

and $\gamma_{k-i} = \begin{cases} 1, \text{ if } u_{k-i} \text{ was received during } [kT_s, (k+1)T_s) \\ 0, \qquad \text{if } u_{k-i} \text{ was lost during } [kT_s, (k+1)T_s) \end{cases}$.

By using a new augment state variable $z_k = [x_k^T \, u_{k-1}^T \cdots u_{k-b}^T \, d_{k-1}^T \cdots d_{k-b}^T]^T$, equation (2) can be expressed as a linear time-varying discrete-time system described by

$$z_{k+1} = A_{zk} z_k + B_{zk} u_k + D_{zk} d_k \tag{3}$$

where the system matrices are a function of the unknown random delays, and packet losses which are given by

$$A_{zk} = \begin{bmatrix} A_s & \gamma_{k-1}B_1^k & \cdots & \gamma_{k-i}B_i^k & \cdots & \gamma_{k-b}B_b^k & \gamma_{k-1}D_1^k & \cdots & \gamma_{k-i}D_i^k & \cdots & \gamma_{k-b}D_b^k \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & I_m & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & I_m & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots & \vdots & 0 & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & I_m & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & I_l & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 & \cdots & I_l & 0 \end{bmatrix}$$

and $B_{zk} = [(\gamma_k B_0^k)^T \;\; I_m \;\; 0 \cdots 0]^T$, $D_{zk} = [0 \cdots (\gamma_k D_0^k)^T \;\; I_l \;\; 0 \cdots\cdots 0]^T$.

where I_m, I_l are $m \times m$ and $l \times l$ identity matrices.

Remark 2*:* It is assumed that wireless communication network changes more slowly (Goldsmith, 2005) when compared to the sampling rate. Therefore, the NCS system description (3) can be considered as a linear but slowly time-varying system with uncertain dynamics. The communication imperfections (i.e. network-induced delays and packet losses) are not accurately known except their upper bounds thus making the NCS dynamics uncertain. In this paper, the optimal strategy is proposed based on the slowly varying unknown linear NCS.

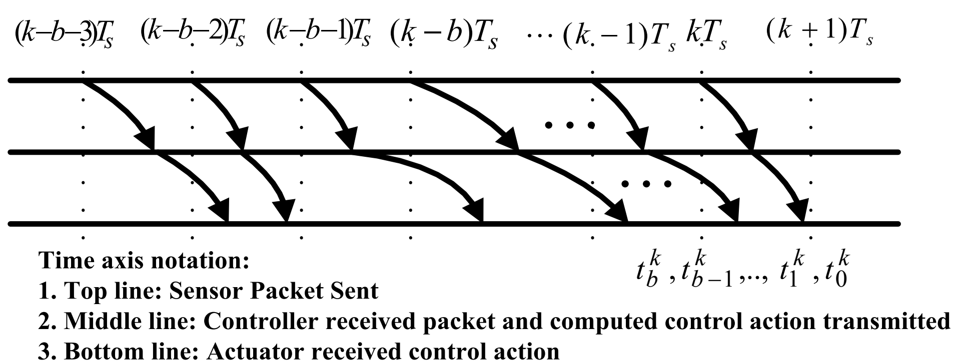


Fig 2. Timing diagram of signals transmitting in NCS.

Thus in this paper, based on optimal control theory (Lewis and Syrmos, 1995; Åstrom, 1970), stochastic cost function can be defined as

$$J_k = \underset{\tau,\gamma}{E}\left[\sum_{i=k}^{\infty}\left(x_i^T G x_i + u_i^T R u_i - d_i^T S d_i\right)\right] \quad \forall k = 0,1,2,... \tag{4}$$

where u_i, d_i are control and disturbance inputs respectively, G is a symmetric positive semi-definite matrix, R is a symmetric positive definite matrix, and S is a symmetric positive definite matrix defined equal to the square of upper bound γ on the desired L_2 gain disturbance attenuation (i.e. $S = \gamma^2 I$, I is identity matrix) (Basar and Bernhard, 1995), and $\underset{\tau,\gamma}{E}(\bullet)$ is the expectation operator (in this case the mean value) of $\sum_{i=k}^{\infty}\left(x_i^T Q x_i + u_i^T R u_i - d_i^T S d_i\right)$ based on the communication imperfections (i.e. network-induced delays and packet losses) at various time interval. After redefining the augment state variable z_k, original stochastic cost function, equation (4) can be expressed as

$$J_k = \underset{\tau,\gamma}{E}\left[\sum_{i=k}^{\infty}\left(z_i^T G_z z_i + u_i^T R_z u_i - d_i^T S_z d_i\right)\right] \quad \forall k = 0,1,2,... \tag{5}$$

where $G_z = \begin{bmatrix} Q & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, R_z = R, S_z = S$

Note that G_z is still symmetric positive semi-definite matrix while R_z, S_z are symmetric positive definite matrices respectively. Next traditional optimal control of discrete-time linear zero-sum games (Lewis and Syrmos, 1995; Basar and Bernhard, 1995) is introduced before presenting the proposed scheme.

2.2. Traditional Optimal Control Of Descrete-Time Systems

Consider the discrete-time linear time-varying zero-sum game dynamics described by

$$x_{k+1}= A_k x_k + B_k u_k + D_k d_k \tag{6}$$

where $x_k \in \mathbb{R}^n$ is the system states, $u_k \in \mathbb{R}^m$ is the control input, $d_k \in \mathbb{R}^l$ is the disturbance input and $A_k \in \mathbb{R}^{n\times n}$, $B_k \in \mathbb{R}^{n\times m}$, $D_k \in \mathbb{R}^{n\times l}$ are system dynamics matrix. Based on optimal control theory (Lewis and Syrmos, 1995; Basar and Bernhard, 1995), the infinite-horizon value function can be defined as

$$\begin{aligned} V^*(x_k) &= \min_{u_j} \max_{d_j} \sum_{j=k}^{\infty} r(x_j, u_j, d_j) \\ &= \min_{u_j} \max_{d_j} \sum_{j=k}^{\infty} (x_j^T Q x_j + u_j^T R u_j - d_j^T S d_j) \end{aligned} \tag{7}$$

with $r(x_k, u_k) = x_k^T Q x_k + u_k^T R u_k - d_k^T S d_k$, Q is symmetric positive semi-definite matrix, R and S are symmetric positive definite matrix.

Using dynamic programming, the optimization problem for discrete-time linear zero-sum game (6) and (7) can be derived as

$$V^*(x_k) = \min_{u_k} \max_{d_k} (r(x_k, u_k, d_k) + V^*(x_{k+1})) \tag{8}$$

Then Bellman equation can be represented as

$$0 = \min_{u_k} \max_{d_k} (r(x_k, u_k, d_k) + V^*(x_{k+1}) - V^*(x_k)) \tag{9}$$

Assuming that minimum on the right side of (9) exists and is unique then optimal strategy for linear zero-sum game can be expressed as (Lewis and Syrmos, 1995; Basar and Bernhard, 1995)

$$
\begin{aligned}
u_k^* &= -\frac{1}{2} R^{-1} B_k \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}} \\
d_k^* &= \frac{1}{2} S^{-1} D_k \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}}
\end{aligned}
\tag{10}
$$

Substituting optimal control policy (10) into Bellman equation, then Bellman equation with optimal strategy u_k^* and d_k^* can be derived as

$$
\begin{aligned}
0 = x_k^T Q x_k &+ \frac{1}{4} \frac{\partial V^{*T}(x_{k+1})}{\partial x_{k+1}} B_k^T R^{-1} B_k \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}} - \frac{1}{4} \frac{\partial V^{*T}(x_{k+1})}{\partial x_{k+1}} D_k^T S^{-1} D_k \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}} \\
&+ V^*(x_{k+1}) - V^*(x_k)
\end{aligned}
\tag{11}
$$

For linear system, assuming that zero-sum game has a value and is solvable, then value function (7) is known as a quadratic form of state and is represented as (Lewis and Syrmos, 1995, Basar and Olsder, 1995)

$$
V^*(x_k) = x_k^T P_k x_k \tag{12}
$$

where P_k is positive semi-definite matrix. Substituting (12) into (11), Bellman equation becomes Game-Theoretic Riccati equation (GRE) as

$$
0 = A_k^T P_{k+1} A_k + Q - P_k - [A_k^T P_{k+1} B_k \;\; A_k^T P_{k+1} D_k] \begin{bmatrix} B_k^T P_{k+1} B_k + R & B_k^T P_{k+1} D_k \\ D_k^T P_{k+1} B_k & D_k^T P_{k+1} D_k - S \end{bmatrix}^{-1} \begin{bmatrix} B_k^T P_{k+1} A_k \\ D_k^T P_{k+1} A_k \end{bmatrix} \tag{13}
$$

It is obvious that P_k is the solution of Riccati Equation. Meanwhile, optimal strategy can be expressed in terms of P_k and system matrix as

$$
\begin{aligned}
u_k^* &= (B_k^T P_{k+1} B_k + R - B_k^T P_{k+1} D_k (D_k^T P_{k+1} D_k - S)^{-1} D_k^T P_{k+1} B_k)^{-1} \\
&\quad \times (B_k^T P_{k+1} D_k (D_k^T P_{k+1} D_k - S)^{-1} D_k^T P_{k+1} A_k - B_k^T P_{k+1} A_k) x_k \\
d_k^* &= (D_k^T P_{k+1} D_k - S - D_k^T P_{k+1} B_k (B_k^T P_{k+1} B_k + R)^{-1} B_k^T P_{k+1} D_k)^{-1} \\
&\quad \times (D_k^T P_{k+1} B_k (B_k^T P_{k+1} B_k + R)^{-1} B_k^T P_{k+1} A_k - D_k^T P_{k+1} A_k) x_k
\end{aligned}
\tag{14}
$$

Remark 3: Traditional Game-Theoretic Riccati equation (13) is solved backward-in-time and optimal strategy (14) is obtained by using P_k, A_k, B_k and D_k. For linear zero-sum game with uncertain system dynamics, solving (13) and (14) is a challenge. Instead, policy and/or value iteration algorithm (Tamimi et al., 2007) have been implemented to approximate the value function in ADP and subsequently the control and disturbance inputs based on estimated value function using (10) are obtained so that the system dynamics are not needed. However, with the policy and value iteration-based schemes, it is not clear how to select number of iterations required for convergence and stability while keeping the hardware constraints. Inadequate number of policy and value iteration can lead to instability (Travis and Jagannathan, 2011).

Hence, in this paper, a time-based ADP method with adaptive estimation will be proposed to solve stochastic optimal strategy of NCS quadratic zero-sum games in forward-in-time manner and without using iteration methodology and known system dynamics as will be discussed in the next sections.

3. Stochastic Optimal Strategy for NCS Quadratic Zero-Sum Games

In this section, we use the idea of ADP (Watkins, 1989; Tamimi et al., 2007) and the concept of adaptive estimation of value function to develop stochastic optimal strategy for NCS quadratic zero-sum games with uncertain linear time-varying system dynamics that change slowly in comparison with the sampling interval due to communication imperfections (i.e. network-induced delays and packet losses). Thus, in this section, first, we introduce an adaptive estimation scheme to obtain the unknown value function for NCS quadratic zero-sum games with network imperfection. Second, a model-free online tuning of the parameters based on adaptive estimation and ADP algorithm will be proposed. Eventually the convergence proof is given.

3.1. Value Function Definition for NCS Quadratic Zero-Sum Games

In this section, we formulate Bellman's optimality principle for the NCS quadratic zero-sum games by using the concept of ADP under communication imperfections (i.e. network-induced delays and packet losses) described by (3). It is easy to verify that NCS quadratic zero-sum games has a unique equilibrium point, $z=0$, on a set Ω while the states are still measurable. According to these conditions, the stochastic optimal strategy which optimize the stochastic cost function J_k for NCS system (3) can be derived as (Tamimi et al., 2007; Basar and Olsder, 1995), $u_k^* = -K_k z_k, \quad d_k^* = -L_k z_k$ with K_k, L_k being the optimal Kalman gains for the control and disturbance inputs respectively.

If we assume that there exists a solution to the GRE, that is strictly feedback stabilizing, and then it can be shown (Basar and Olsder, 1995) that the policies attain a saddle-point equilibrium (Basar and Bernhard, 1995), which implies that minimax is equal to maximin, in the restricted class of feedback stabilizing policies. Assuming that the game has a value and is solvable, and then it is known that the value function is quadratic in the state and is given by (Lewis and Syrmos, 1995)

$$J_k = \underset{\tau,\gamma}{E}\left(z_k^T P_k z_k\right) \tag{15}$$

where matrix $P_k \geq 0$ is a solution to the GRE (Dragan and Morozan, 1997). The optimal action dependent value function of NCS quadratic zero-sum games is now defined to be

$$V\left(z_k, u_k, d_k\right) = \underset{\tau,\gamma}{E}\left\{z_k^T Q_z z_k + u_k^T R_z u_k - d_k^T S_z d_k + J_{k+1}\right\} \tag{16}$$

$$= \underset{\tau,\gamma}{E}\left\{r(z_k, u_k, d_k) + J_{k+1}\right\} = \underset{\tau,\gamma}{E}\{\left[z_k^T \; u_k^T \; d_k^T\right] H_k \left[z_k^T \; u_k^T \; d_k^T\right]^T\}$$

where $r(z_k, u_k, d_k) = z_k^T Q_z z_k + u_k^T R_z u_k - d_k^T S_z d$.

Since stochastic optimal control and disturbance inputs, u_k^*, d_k^*, are dependent on state z_k which is known at time k , value function can be expressed as $V(z_k, u_k, d_k) = [z_k^T \; u_k^T \; d_k^T]^T \; \underset{\tau,\gamma}{E}(H_k)[z_k^T \; u_k^T \; d_k^T]$. Then using Bellman equation and cost function, we can get

$$\begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix}^T \underset{\tau,\gamma}{E}(H_k) \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix} = \underset{\tau,\gamma}{E}\{r(z_k, u_k, d_k) + J_{k+1}\} \tag{17}$$

$$= \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix}^T \begin{bmatrix} G_z & 0 & 0 \\ 0 & R_z & 0 \\ 0 & 0 & -S_z \end{bmatrix} \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix} + \underset{\tau,\gamma}{E}\left\{ \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix}^T \begin{bmatrix} A_{zk} \\ B_{zk} \\ D_{zk} \end{bmatrix}^T P_{k+1} \begin{bmatrix} A_{zk} \\ B_{zk} \\ D_{zk} \end{bmatrix} \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix} \right\}$$

$$= \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix}^T \begin{bmatrix} G_z + \underset{\tau,\gamma}{E}(A_{zk}^T P_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T P_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T P_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} A_{zk}) & R_z + \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} D_{zk}) - S_z \end{bmatrix} \begin{bmatrix} z_k \\ u_k \\ d_k \end{bmatrix}$$

Therefore, $\underset{\tau,\gamma}{E}(H_k)$ can be written as

$$\overline{H}_k = \underset{\tau,\gamma}{E}(H_k) = \begin{bmatrix} \overline{H}_k^{zz} & \overline{H}_k^{zu} & \overline{H}_k^{zd} \\ \overline{H}_k^{uz} & \overline{H}_k^{uu} & \overline{H}_k^{ud} \\ \overline{H}_k^{dz} & \overline{H}_k^{du} & \overline{H}_k^{dd} \end{bmatrix} \tag{18}$$

$$= \begin{bmatrix} G_z + \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} B_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} D_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} A_{zk}\right) & R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} B_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} D_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(D_{zk}^T P_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(D_{zk}^T P_{k+1} B_{zk}\right) & \underset{\tau,\gamma}{E}\left(D_{zk}^T P_{k+1} D_{zk}\right) - S_z \end{bmatrix}$$

Then using (9), for the zero-sum game (Basar and Bernhard, 1995), the gain matrix associated with the optimal control and disturbance inputs can be expressed in terms of $\overline{H}_k$ as

$$\begin{aligned} K_k &= -(R_z + \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} B_{zk}) - \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} D_{zk})(\underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} D_{zk}) - S_z)^{-1} \\ &\times \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} B_{zk}))^{-1}(\underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} D_{zk})(\underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} D_{zk}) - S_z)^{-1} \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} A_{zk}) - \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} A_{zk})) \\ &= -(\overline{H}_k^{uu} - \overline{H}_k^{ud}(\overline{H}_k^{dd})^{-1}\overline{H}_k^{du})^{-1}(\overline{H}_k^{ud}(\overline{H}_k^{dd})^{-1}\overline{H}_k^{dz} - \overline{H}_k^{uz}) \end{aligned} \tag{19}$$

and

$$\begin{aligned} L_k &= -(\underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} D_{zk}) - S_z - \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} B_{zk})(\underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} B_{zk}) + R_z)^{-1} \\ &\times \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} D_{zk}))^{-1}(\underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} B_{zk})(\underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} B_{zk}) + R_z)^{-1} \times \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} A_{zk}) - \underset{\tau,\gamma}{E}(D_{zk}^T P_{k+1} A_{zk})) \\ &= -(\overline{H}_k^{du} - \overline{H}_k^{du}(\overline{H}_k^{uu})^{-1}\overline{H}_k^{ud})^{-1}(\overline{H}_k^{du}(\overline{H}_k^{uu})^{-1}\overline{H}_k^{uz} - \overline{H}_k^{dz}) \end{aligned} \tag{20}$$

Equations (19) and (20) represent time varying gains based on the solution of the GRE and hence some interesting observations can be stated using (19) and (20). If the matrix P_k is known in (11) and (12), then one still need the slowly time-varying system matrices to compute the controller gains. On the other hand, if the slowly time-varying matrix $\overline{H}_k$ can be learned online without the knowledge of NCS dynamics (3), the NCS system matrices are not required to compute the optimal strategy gains. This observation is consistent with the work of (Tamimi et al., 2007) where time invariant gains are derived for suboptimal control of time invariant linear discrete-time zero-sum games. While an adaptive estimator will be utilized to learn the time-varying matrix, $\overline{H}_k$, which in turn will be used to obtain the optimal gains.

Remark 4: It is important to note there are several differences between optimal design in this paper and Tamimi et al. (2007). First, Tamimi et al. (2007) method cannot even maintain the stability of NCS. By contrast, the proposed approach can be utilized for either NCS and uncertain linear time-varying or time-invariant discrete-time zero sum game without the communication imperfections. Second, equations (19) and (20) are based on slowly time varying system (3) and equations in (Tamimi et al., 2007) are only for time-invariant system. Third, Tamimi et al. (2007) uses value iteration within each sampling interval which in turn requires a significant number of iterations for convergence of the algorithm while the proposed scheme updates the value function and control policy once every sampling interval. Therefore, the proposed optimal strategy based on adaptive estimation of cost or value function is an online and forward-in-time approach and does not require policy and value iterations. Eventually, this paper derived closed-loop system stability which is not addressed in (Tamimi et al., 2007).

3.2. Model-Free Online Tuning Based on Adaptive Estimator

The proposed online tuning approach estimates the value-function (17) online. Since value-function includes the $\overline{H}_k$ matrix (18) which can be solved, the control signal and disturbance input can be obtain using (19) and (20). Next we make the following assumption since the NCS is a slowly linear time-varying unknown system (see Remark 3) and the delays are bounded above while the packet losses satisfy the Bernoulli distribution, and both of them change slowly (Goldsmith, 2005).

Assumption 2: The value-function, $V(z_k, u_k, d_k)$, can be expressed as linear in the unknown parameters (LIP)—a standard assumption in adaptive control (Jagannathan, 2006; Ioannou and Sun, 1996).

By using the stochastic adaptive control literature (Chen and Guo, 1991) and (16), the value-function can be represented in vector form as

$$V(z_k, u_k, d_k) = w_k^T \overline{H}_k w_k = \overline{h}_k^T \overline{w}_k \tag{21}$$

where $\overline{h}_k = vec(\overline{H}_k), w_k = [z_k^T \; u^T(z_k) \; d^T(z_k)]^T, w_k \in \mathbb{R}^q, q = n + (b+1)m + (b+1)l$, and $\overline{w}_k = (w_{k1}^2, ..., \; w_{k1}w_{kq}, w_{k2}^2, ..., w_{kq-1}w_{kq}, \; w_{kq}^2)$ is the Kronecker product quadratic polynomial stochastic indepdent basis vector. $\overline{h}_k = vec(\overline{H}_k)$ with the vector function acting on $q \times q$ matrices thus yielding a $q(q+1)/2 \times 1$ column vector.

Note: The $vec(\bullet)$ function is constructed by stacking the columns of matrix into one column vector with off-diagonal elements which can be combined as $H_{mn} + H_{nm}$. Therefore, the value-function can be expressed as target unknown parameter vector multiplied by the regression function $\overline{w}_k$.

The time-varying matrix $\overline{H}_k$ can be considered as slowly varying (Goldsmith, 2005). Then it can be expressed as a time-varying target parameter vector and a known regression function $\overline{w}_k$. Now, the value-function $V(z_k, u_k, d_k)$ estimation will be considered.

According to the definition of value-function (16) and relationship between value-function and stochastic cost function (Tamimi et al., 2007), we can use matrix $\overline{H}_k$ in (17) to express the stochastic cost function as

$$J_k(z) = V(z_k, u_k, d_k) = w_k^T \overline{H}_k w_k = \overline{h}_k^T \overline{w}_k \tag{22}$$

Then the value-function, $V(z_k, u_k, d_k)$, can be estimated by an adaptive estimator in terms of estimated parameter vector $\hat{\overline{h}}_k$ as

$$\hat{J}_k(z) = \hat{V}(z_k, u_k, d_k) = \hat{\overline{h}}_k^T \overline{w}_k \tag{23}$$

where $w_k = [z_k^T \ u^T(z_k) \ d^T(z_k)]^T$ and $\overline{w}_k$ is Kronecker product quadratic polynomial stochastic indepdent basis vector of w_k.

It is observed that Bellman Equation can be rewritten as $J_{k+1} - J_k + r(z_k, u_k, d_k) = 0$. This relationship, however, is not guaranteed to hold when we apply the estimated matrix $\hat{\overline{H}}_k$. Hence, using delayed values for convenience; the residual error associated with (15) can be expressed as $\hat{J}_{k+1} - \hat{J}_k + r(z_k, u_k, d_k) = e_{hk}$, i.e.

$$\begin{aligned} e_{hk} &= r(z_{k-1}, u_{k-1}, d_{k-1}) + \hat{\overline{h}}_k^T \overline{w}_k - \hat{\overline{h}}_k^T \overline{w}_{k-1} \\ &= r(z_{k-1}, u_{k-1}, d_{k-1}) + \hat{\overline{h}}_k^T (\overline{w}_k - \overline{w}_{k-1}) \\ &= r(z_{k-1}, u_{k-1}, d_{k-1}) + \hat{\overline{h}}_k^T \Delta W_{k-1} \end{aligned} \tag{24}$$

where $\Delta W_{k-1} = \overline{w}_k - \overline{w}_{k-1}$.

The dynamics of (24) are then rewritten as

$$e_{hk+1} = r(z_k, u_k, d_k) + \hat{\overline{h}}_{k+1}^T \Delta W_k \tag{25}$$

Next, we define an auxiliary residual error vector as

$$\Sigma_{hk} = \Gamma_{k-1} + \hat{\overline{h}}_k^T \Omega_{k-1} \in \mathbb{R}^{1\times(1+i)} \tag{26}$$

where $\Gamma_{k-1} = [r(z_{k-1}, u_{k-1}, d_{k-1}) \ r(z_{k-2}, u_{k-2}, d_{k-2}) \cdots r(z_{k-1-i}, u_{k-1-i}, d_{k-1-i})]$, and $\Omega_{k-1} = [\Delta W_{k-1} \ \Delta W_{k-2} \cdots \Delta W_{k-1-i}], 0 < i < k-1 \in \mathbb{N}$ with $\mathbb{N}$ being the set of natural real numbers. It is important to note that (18) indicates a time history of the previous $i+1$ residual errors (16) recalculated by using the most recent $\hat{\overline{h}}_k$. The time history of previous residual errors allows one to overcome the need for any iterative-based value and policy update schemes while still rendering optimal control solution. Therefore, the proposed approach can be referred to as time-based ADP.

Next the dynamics of the auxiliary vector (26) are generated similar to (25) and revealed to be

$$\Sigma_{hk+1} = \Gamma_k + \hat{\bar{h}}_{k+1}^T \Omega_k \tag{27}$$

Now define the update law of the slowly time-varying matrix $\overline{H}_k$ as

$$\hat{\bar{h}}_{k+1} = \Omega_k \left(\Omega_k^T \Omega_k\right)^{-1} \left(\alpha_h \Sigma_{hk}^T - \Gamma_k^T\right) \tag{28}$$

where $0 < \alpha_h < 1$. Substituting (28) into (27) results in

$$\Sigma_{hk+1} = \alpha_h \Sigma_{hk} \tag{29}$$

Remark 5: It is observed that the cost function J_k and adaptive estimation (23) will become zero only when $z_k = 0$. Hence, when the system states have converged to zeros, the value-function approximation is no longer updated. It can be seen as a persistency of excitation (PE) requirement for the inputs to the value function estimator wherein the system states must be persistently existing long enough for the adaptive estimator to learn the optimal stochastic cost function. Therefore exploration noise is added to the control and disturbance inputs in order to satisfy the PE condition (Tamimi et al., 2007) which is given next.

Definition 1: (Persistence of excitation) A stochastic vector $\beta_k \in \mathbb{R}^p$ is said to be PE if there exists positive constants δ, α and $k_0 \geq 1$, such that

$$\sum_{k=k_0}^{k_0+\delta} E[\beta_k \beta_k^T)] \geq \alpha \mathbf{I}$$

where $\mathbf{I}$ is identity matrix, $E\{\bullet\}$ is the mean value of $\{\bullet\}$.

Lemma 1: Persistence of excitations of vector ΔW_k (24) and Ω_k can be satisfied by adding exploration noise.

Proof: Refer to Xu, Jagannathan and Lewis (2011).

Now define the parameter estimation error as $\tilde{\bar{h}}_k = \bar{h}_k - \hat{\bar{h}}_k$. Rewrite Bellman equation using an adaptive estimation with target parameters (21) revealing $r(z_k, u_k, d_k) + \bar{h}_{k+1}^T \bar{w}_{k+1} = \bar{h}_{k+1}^T \bar{w}_k$, which can be expressed as

$$r(z_k, u_k, d_k) = \bar{h}_{k+1}^T \bar{w}_k - \bar{h}_{k+1}^T \bar{w}_{k+1} = -\bar{h}_{k+1}^T \Delta W_k \tag{30}$$

Substituting $r(z_k, u_k, d_k)$ into (25) and utilizing (24) with $e_{hk+1} = \alpha_h e_{hk}$ from (29) yields

$$\tilde{\bar{h}}_{k+1}^T \Delta W_k = -\alpha_h r(z_{k-1}, u_{k-1}, d_{k-1}) - \alpha_h \hat{\bar{h}}_k^T \Delta W_{k-1} \tag{31}$$

Using the similar method as $r(z_k, u_k, d_k)$, we can form $r(z_{k-1}, u_{k-1}, d_{k-1})$, and substituting this expression into (31), we have

$$\tilde{\bar{h}}_{k+1}^T \Delta W_k = \alpha_h \tilde{\bar{h}}_k^T \Delta W_{k-1} \tag{32}$$

Next, the convergence of the cost function errors with adaptive estimation error dynamics $\tilde{\bar{h}}_k$ given by (32) is demonstrated for an initial admissible control (Jagannathan, 2006) policy. The NCS slowly time-varying system dynamics are known to be asymptotically stable in the mean if an initial admissible control policy can be applied provided the system matrices are known. However, introducing the estimated value-function results in estimation errors for the stochastic cost function J_k, and therefore stability of estimated stochastic cost function needs to be studied. Similarly as (Xu, Jagannathan and Lewis, 2011), cost adaptive estimator errors can be proven to be asymptotic stable in the mean, i.e. $\hat{J}_k \to J_k^*$ and $\tilde{\bar{h}}_k \to 0$ when $k \to \infty$. Subsequently, asymptotic stability of cost adaptive estimation errors will be used for proving the overall closed-loop system stability in Theorem 1 by using an initial admission control policy.

Next, we show that the estimated control and disturbance input based on this estimated matrix will indeed converge to the optimal control input and disturbance input.

3.3. Estimation of the Optimal Feedback Control and Disturbance Signal

There are two ways to estimate the optimal control and disturbance signal inputs for regulating the NCS quadratic zero-sum games. One is based on time-varying matrix $\overline{H}_k$ while the other is based on standard optimal theory by minimizing the cost function. The difference being that the latter method requires the system dynamics and it solves the optimal controller backward. However, it is shown next that ultimately both are equivalent and can be used in the proofs.

Method I: As mentioned before, slowly time varying matrix $\overline{H}_k$ can be estimated by using an adaptive estimator. According to ADP scheme and equation (19)(20), the estimated optimal NCS control and disturbance inputs can be expressed by the adaptive estimation $\overline{H}_k$ as

$$\begin{aligned} \hat{u}_{1k} &= -\hat{K}_k z_k = \left(\hat{\overline{H}}_k^{uu} - \hat{\overline{H}}_k^{ud}\left(\hat{\overline{H}}_k^{dd}\right)^{-1}\hat{\overline{H}}_k^{du}\right)^{-1}\left(\hat{\overline{H}}_k^{ud}\left(\hat{\overline{H}}_k^{dd}\right)^{-1}\hat{\overline{H}}_k^{dz} - \hat{\overline{H}}_k^{uz}\right) z_k \\ \hat{d}_{1k} &= -\hat{L}_k z_k = \left(\hat{\overline{H}}_k^{du} - \hat{\overline{H}}_k^{du}\left(\hat{\overline{H}}_k^{uu}\right)^{-1}\hat{\overline{H}}_k^{ud}\right)^{-1}\left(\hat{\overline{H}}_k^{du}\left(\hat{\overline{H}}_k^{uu}\right)^{-1}\hat{\overline{H}}_k^{uz} - \hat{\overline{H}}_k^{dz}\right) z_k \end{aligned} \tag{33}$$

Method II: Alternatively, the estimated optimal control and disturbance signal which optimize the estimated cost function (22) with adaptive estimation $\hat{\overline{H}}_k$ as

$$\begin{aligned} \hat{u}_{2k} &= -\frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \\ \hat{d}_{2k} &= \frac{1}{2} S_z^{-1} D_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \end{aligned} \tag{34}$$

where $J_{k+1} = \underset{\tau,\gamma}{E}\left(w_{k+1}^T \hat{H}_{k+1} w_{k+1}\right) = \underset{\tau,\gamma}{E}\left(z_{k+1}^T \hat{P}_{k+1} z_{k+1}\right)$. Next, it will be shown that the optimal control and disturbance input obtained by method I and II are equivalent.

Lemma 2: The optimal control and disturbance estimations calculated with the adaptive estimation of $V(z_k, u_k, d_k)$ are equal to the optimal control and disturbance inputs obtained by optimizing the cost function J_k, i.e. $\hat{u}_{1k} = \hat{u}_{2k}, \hat{d}_{1k} = \hat{d}_{2k}$.

Proof: Use the Bellman equation and ADP algorithm with estimated stochastic cost function and matrix $\overline{H}_k$, we have

$$\hat{J}_k + e_{hk} = r\left(z_k, u(z_k), d(z_k)\right) + \hat{J}_{k+1} \tag{35}$$

1) Left side of (35) can be expressed as

$$\hat{J}_k + e_{hk} = \hat{\overline{h}}_k^T \overline{w}_k + e_{hk} = \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \hat{\overline{H}}_k \begin{bmatrix} z_k \\ u(z_k) \\ w(z_k) \end{bmatrix} + e_{hk}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} \hat{\overline{H}}_k^{zz} & \hat{\overline{H}}_k^{zu} & \hat{\overline{H}}_k^{zd} \\ \hat{\overline{H}}_k^{uz} & \hat{\overline{H}}_k^{uu} & \hat{\overline{H}}_k^{ud} \\ \hat{\overline{H}}_k^{dz} & \hat{\overline{H}}_k^{du} & \hat{\overline{H}}_k^{dd} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} + e_{hk} \tag{36}$$

2) Right side of (35) can be shown as

$$\underset{\tau,\gamma}{E}\{r(z_k, u_k, d_k) + \hat{J}_{k+1}\} = z_k^T G_z z_k + u_k^T R_z u_k - d_k^T S_z d_k + \underset{\tau,\gamma}{E}(z_{k+1}^T P_{k+1} z_{k+1}) \tag{37}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} G_z & 0 & 0 \\ 0 & R_z & 0 \\ 0 & 0 & -S_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} + \underset{\tau,\gamma}{E}\left\{ \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} A_{zk} \\ B_{zk} \\ D_{zk} \end{bmatrix}^T \hat{P}_{k+1} \begin{bmatrix} A_{zk} \\ B_{zk} \\ D_{zk} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} \right\}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} G_z + \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) & R_z + \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}$$

According to the (36) and (37), equation (35) can be derived as

$$\begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} \hat{H}_k^{zz} & \hat{H}_k^{zu} & \hat{H}_k^{zd} \\ \hat{H}_k^{uz} & \hat{H}_k^{uu} & \hat{H}_k^{ud} \\ \hat{H}_k^{dz} & \hat{H}_k^{du} & \hat{H}_k^{dd} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} + e_{hk} \tag{38}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} G_z + \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) & R_z + \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}$$

Hence

$$\begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} \hat{H}_k^{zz} & \hat{H}_k^{zu} & \hat{H}_k^{zd} \\ \hat{H}_k^{uz} & \hat{H}_k^{uu} & \hat{H}_k^{ud} \\ \hat{H}_k^{dz} & \hat{H}_k^{du} & \hat{H}_k^{dd} \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} + e_{hk}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} G_z + \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T \hat{P}_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) & R_z + \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk}) \\ \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk}) & \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z \end{bmatrix}$$

$$\times \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} - \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} e_{hk}/tr\left(z_k^T z_k\right) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix} \tag{39}$$

$$= \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}^T \begin{bmatrix} G_z + \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} A_{zk}\right) - \dfrac{e_{hk}}{tr\left(z_k^T z_k\right)} & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} B_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T \hat{P}_{k+1} D_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) & R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) & \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} D_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} B_{zk}\right) & \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} D_{zk}\right) - S_z \end{bmatrix} \begin{bmatrix} z_k \\ u(z_k) \\ d(z_k) \end{bmatrix}$$

Hence, we have

$$\hat{u}_{1k} = \left(\hat{H}_k^{uu} - \hat{H}_k^{ud} \left(\hat{H}_k^{dd} \right)^{-1} \hat{H}_k^{du} \right)^{-1} \left(\hat{H}_k^{ud} \left(\hat{H}_k^{dd} \right)^{-1} \hat{H}_k^{dz} - \hat{H}_k^{uz} \right) z_k$$
$$= [(R_z + \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) - \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk})(\underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z)^{-1}$$
$$\times \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk}))^{-1} (\underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk})(\underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z)^{-1}$$
$$\times \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk}) - \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk})] z_k \tag{40}$$

$$\hat{d}_{1k} = \left(\hat{H}_k^{dd} - \hat{H}_k^{du} \left(\hat{H}_k^{uu} \right)^{-1} \hat{H}_k^{ud} \right)^{-1} \left(\hat{H}_k^{du} \left(\hat{H}_k^{uu} \right)^{-1} \hat{H}_k^{uz} - \hat{H}_k^{dz} \right) z_k$$
$$= [(\underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z - \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk})(\underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) + R_z)^{-1}$$
$$\times \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk}))^{-1} (\underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk})(\underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) + R_z)^{-1}$$
$$\times \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) - \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk})] z_k$$

On the other hand, the optimal control and disturbance signals $\hat{u}_{2k}, \hat{d}_{2k}$ generated by method II can be expressed as:

$$\begin{aligned} \hat{u}_{2k} &= -\frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} = -R_z^{-1} B_{zk}^T \hat{P}_{k+1} z_{k+1} \\ \hat{d}_{2k} &= \frac{1}{2} S_z^{-1} D_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} = S_z^{-1} D_{zk}^T \hat{P}_{k+1} z_{k+1} \end{aligned} \tag{41}$$

Since $z_{k+1} = A_{zk} z_{k+1} + B_{zk} u_k + D_{zk} d_k$, (41) can be derived as

$$\hat{u}_{2k} = -R_z^{-1} B_{zk}^T \hat{P}_{k+1} z_{k+1}$$
$$= -R_z^{-1} \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) z_k - R_z^{-1} \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}) \hat{u}_{2k} - R_z^{-1} \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk}) \hat{d}_{2k}$$

$$\hat{d}_{2k} = S_z^{-1} D_{zk}^T \hat{P}_{k+1} z_{k+1} \tag{42}$$

$$= S_z^{-1} \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk}) z_k + S_z^{-1} \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk}) \hat{u}_{2k} + S_z^{-1} \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) \hat{d}_{2k}$$

It is obvious that equation (42) is an equation with $\hat{u}_{2k}, \hat{d}_{2k}$ which can be simplified as

$$\begin{cases} \left(I + R_z^{-1} \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right)\right)\hat{u}_{2k} = -R_z^{-1} E\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) z_k \\ \qquad\qquad\qquad\qquad\qquad - R_z^{-1} \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} D_{zk}\right)\hat{d}_{2k} \\ \left(I - S_z^{-1} \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} D_{zk}\right)\right)\hat{d}_{2k} = S_z^{-1} E\left(D_{zk}^T \hat{P}_{k+1} A_{zk}\right) z_k \\ \qquad\qquad\qquad\qquad\qquad + S_z^{-1} \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} B_{zk}\right)\hat{u}_{2k} \end{cases}$$

In the other words,

$$\left(R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right)\right)\hat{u}_{2k} = -\underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} A_{zk}\right) z_k - \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} D_{zk}\right)\hat{d}_{2k}$$

and (43)

$$\left(S_z - \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} D_{zk}\right)\right)\hat{d}_{2k} = \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} A_{zk}\right) z_k + \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} B_{zk}\right)\hat{u}_{2k}$$

This equation (43) can be solved to obtain

$$\hat{u}_{2k} = (R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} B_{zk}\right) - \underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} D_{zk}\right)\left(\underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} D_{zk}\right) - S_z\right)^{-1}$$

$$\times \underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} B_{zk}\right))^{-1} (\underset{\tau,\gamma}{E}\left(B_{zk}^T \hat{P}_{k+1} D_{zk}\right)\left(\underset{\tau,\gamma}{E}\left(D_{zk}^T \hat{P}_{k+1} D_{zk}\right) - S_z\right)^{-1}$$

$$\times \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk}) - \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk})) z_k \tag{44}$$

$$= \left(\hat{\hat{H}}_k^{uu} - \hat{\hat{H}}_k^{uw}\left(\hat{\hat{H}}_k^{ww}\right)^{-1} \hat{\hat{H}}_k^{wu}\right)^{-1} \left(\hat{\hat{H}}_k^{uw}\left(\hat{\hat{H}}_k^{ww}\right)^{-1} \hat{\hat{H}}_k^{wz} - \hat{\hat{H}}_k^{uz}\right) z_k = \hat{u}_{1k}$$

$$\hat{d}_{2k} = (\underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} D_{zk}) - S_z - (R_z + \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}))^{-1}$$

$$\times \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} D_{zk}))^{-1} (\underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} B_{zk})(R_z + \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} B_{zk}))^{-1}$$

$$\times \underset{\tau,\gamma}{E}(B_{zk}^T \hat{P}_{k+1} A_{zk}) - \underset{\tau,\gamma}{E}(D_{zk}^T \hat{P}_{k+1} A_{zk})) z_k \tag{45}$$

$$= \left(\hat{\hat{H}}_k^{ww} - \hat{\hat{H}}_k^{wu}\left(\hat{\hat{H}}_k^{uu}\right)^{-1} \hat{\hat{H}}_k^{uw}\right)^{-1} \left(\hat{\hat{H}}_k^{wu}\left(\hat{\hat{H}}_k^{uu}\right)^{-1} \hat{\hat{H}}_k^{uz} - \hat{\hat{H}}_k^{wz}\right) z_k = \hat{d}_{1k}$$

Therefore, $\widetilde{u}_k = \hat{u}_{1k} - \hat{u}_{2k} = 0, \widetilde{d}_k = \hat{d}_{1k} - \hat{d}_{2k} = 0$ since $\hat{u}_{1k} = \hat{u}_{2k}$, $\hat{d}_{1k} = \hat{d}_{2k}$.

Since the equality proven in this lemma is in both ways and noting the drawback of second method, we use the first method to solve the optimal strategy design for the NCS. However, we will use the Lemma 2 to complete the convergence proof since they are equivalent. Next, the stability of the cost function, control estimation, and adaptive estimation error dynamics are considered.

3.4. Closed-Loop System Stability

In this section, it will be shown that slowly time-varying matrix $\overline{H}_k$ and related value function estimation errors dynamics are asymptotically stable in the mean. Further, the estimated control and disturbance input for NCS (33) will approach their optimal control signal asymptotically. The block diagram representation of stochastic optimal regulator of NCS quadratic zero-sum games with unknown system dynamics is shown in Figures 3 and 4 presents the flowchart of proposed stochastic optimal strategy for NCS quadratic zero-sum games wherein the optimal strategy are obtained without using value and policy iterations.

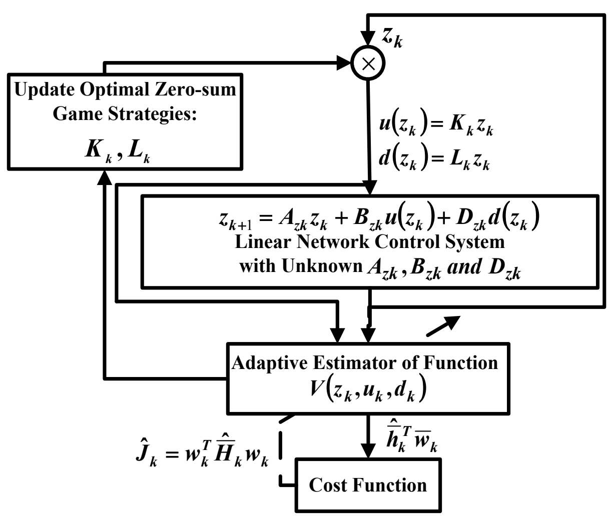


Fig 3. Stochastic optimal regulator block diagram.

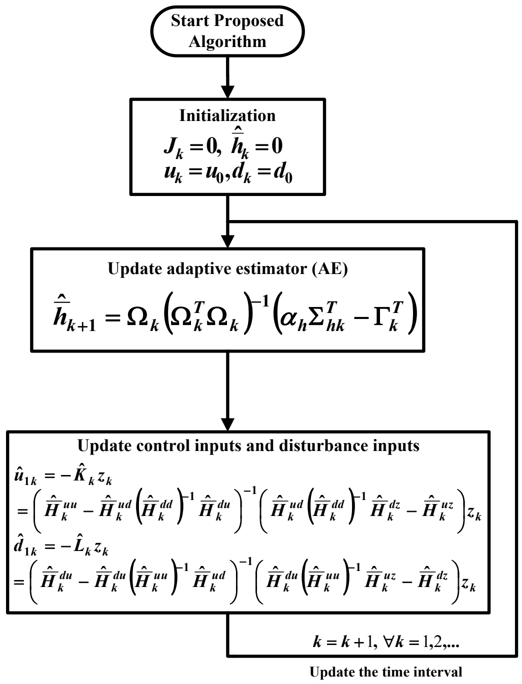


Fig 4. The flowchart of proposed stochastic optimal scheme.

Next, the initial system states are considered to reside in a set when the initial stabilizing control and disturbance inputs u_{0k}, d_{0k} are being utilized. Further sufficient condition for the adaptive estimator tuning gain α_h is derived to ensure the all future states will converge to zero. Then it can be shown that the actual control and disturbance input approach the optimal strategy asymptotically.

Before convergence proof, the following result is needed to establish bounds on the closed loop dynamics when the optimal control and disturbance inputs are applied to the NCS system (3) with communication imperfections (i.e. network-induced delays and packet losses).

Lemma 3*:* There exists admissible control and disturbance policies be applied to the unknown NCS such that the system dynamics are satisfying

$$\left\|A_{zk}z_k + B_{zk}u_k^* + D_{zk}d_k^*\right\|^2 \le k^*\left\|z_k\right\|^2 \tag{46}$$

where $0 < k^* < \dfrac{1}{3}$ is a constant.

Proof: Consider the Lyapunov function candidate

$$V_D(z_k) = z_k^T z_k \tag{47}$$

whose first difference of $V_D(z_k)$ is given by $\Delta V_D(z_k) = z_{k+1}^T z_{k+1} - z_k^T z_k$. Note that since u_k^*, d_k^* are admissible control and disturbance policy, it follows from the definition of admissible control and disturbance that the NCS dynamics (3) with optimal control and disturbance applied are asymptotically stable in the mean, and the sequence $\left\|z_k\right\|,\ k = 1,2,...,\infty$ monotonically decreases until it reaches zero. This result directly implies that $z_{k+1}^T z_{k+1} - z_k^T z_k < 0$ or $\Delta V_D(z_k) < 0$. Using the fact $\Delta V_D(z_k) < 0$, it is clear that $z_{k+1}^T z_{k+1} < z_k^T z_k$. Substituting system dynamics $z_{k+1} = A_{zk}z_k + B_{zk}u_k^* + D_{zk}d_k^*$ yields

$$\begin{aligned}\Delta V_D(z_k) &= (A_{zk}z_k + B_{zk}u_k^* + D_{zk}d_k^*)^T(A_{zk}z_k + B_{zk}u_k^* + D_{zk}d_k^*) - z_k^T z_k \\ &\le \left\|A_{zk}z_k + B_{zk}u_k^* + D_{zk}d_k^*\right\|^2 - \left\|z_k\right\|^2 < 0\end{aligned} \tag{48}$$

Eventually, we must identify a bound on $\left\|A_{zk}z_k + B_{zk}u_k^* + D_{zk}d_k^*\right\|$ which guarantees the sufficient condition for stability $\Delta V_D(z_k) < 0$ is still met. Selecting the bound shown in optimal control policy, (46) reveals $\Delta V_D(z_k) < -(1 - k^*)z_k^T z_k < 0$ as required.

Theorem 1*(Convergence of the Optimal Control and Disturbance Signals):* Given the initial conditions for the system state z_0, cost function and adaptive estimator parameter vectors $\hat{\bar{h}}_0$ be bounded in the set $\mathbf{\Omega}$, let u_{0k}, d_{0k} be any initial admissible control and disturbance policies for the NCS with communication imperfections (i.e. network-induced delays and packet losses), which can maintain initial system condition to be bounded in the set $\boldsymbol{\Omega}$ while satisfying the bounds given by (46) for $0 < k^* < 1/3$. Let the adaptive estimation parameter vector be tuned and estimation control and disturbance policy be provided by (28) and (33) respectively. Then, there exist positive constants α_h given by Theorem 1 such that the system states z_k and stochastic cost function parameter estimator errors $\tilde{\bar{h}}_k$ are all asymptotic stable (AS) in the mean. In other words, as $k \to \infty, z_k \to 0, \tilde{\bar{h}}_k \to 0, \hat{J}_k \to J_k$ and $\hat{u}_{2k} \to u_k^*, \hat{u}_{1k} \to u_k^*, \hat{d}_{2k} \to d_k^*, \hat{d}_{1k} \to d_k^*$.

Proof: Consider the following positive definite Lyapunov function candidate

$$V = V_D(z_k) + V_J(\tilde{\bar{h}}_k) \tag{49}$$

where $V_D(z_k)$ is defined in (39) and $V_J(\tilde{\bar{h}}_k)$ is defined as

$$V_J(\tilde{\bar{h}}_k) = (\tilde{\bar{h}}_k \overline{w}_k - \tilde{\bar{h}}_k \overline{w}_{k-1})^2 = (\tilde{\bar{h}}_k \Delta W_{k-1})^2 \tag{50}$$

The first difference of (49) can be expressed as $\Delta V = \Delta V_D(z_k) + \Delta V_J(\tilde{\bar{h}}_k)$, and considering that $\Delta V_J(\tilde{\bar{h}}_k) = (\tilde{\bar{h}}_{k+1} \Delta W_k)^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2$ with the adaptive estimator, we have

$$\begin{aligned} \Delta V_J(\tilde{\bar{h}}_k) &= (\tilde{\bar{h}}_{k+1} \Delta W_k)^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2 = (\alpha_h \tilde{\bar{h}}_k \Delta W_{k-1})^2 - (\tilde{\bar{h}}_k \Delta W_{k-1})^2 \\ &= -(1-\alpha_h^2)(\tilde{\bar{h}}_k \Delta W_{k-1})^2 \le -(1-\alpha_h^2)\Delta W_{\min}^2 \left\| \tilde{\bar{h}}_k \right\|^2 \end{aligned} \tag{51}$$

Next, considering the first part $\Delta V_D(z_k) = z_{k+1}^T z_{k+1} - z_k^T z_k$ and applying the NCS quadratic zero-sum games and Cauchy-Schwartz inequality reveals

$$\begin{aligned} \Delta V_D(z_k) &\le \left\| A_{zk} z_k + B_{zk} u_k^* + D_{zk} d_k^* - B_{zk} \tilde{u}_k - D_{zk} \tilde{d}_k \right\|^2 - z_k^T z_k \\ &\le 3\left\| A_{zk} z_k + B_{zk} u_k^* + D_{zk} d_k^* \right\|^2 + 3\left\| B_{zk} \tilde{u}_k \right\|^2 + 3\left\| D_{zk} \tilde{d}_k \right\|^2 - z_k^T z_k \end{aligned} \tag{52}$$

Applying the Lemma 3 (bounds on the optimal closed loop system in (46)) and recalling $\hat{u}_{1k} = \hat{u}_{2k}, \hat{d}_{1k} = \hat{d}_{2k}$ from Lemma 2 and (45), we know

$$\tilde{u}_k = \hat{u}_{2k} - \hat{u}_{1k} = \frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}}$$
$$+ \left(\hat{\bar{H}}_k^{uu} - \hat{\bar{H}}_k^{ud} \left(\hat{\bar{H}}_k^{dd} \right)^{-1} \hat{\bar{H}}_k^{du} \right)^{-1} \left(\hat{\bar{H}}_k^{ud} \left(\hat{\bar{H}}_k^{dd} \right)^{-1} \hat{\bar{H}}_k^{dz} - \hat{\bar{H}}_k^{uz} \right) z_k = 0$$

$$\tilde{d}_k = \hat{d}_{2k} - \hat{d}_{1k} = -\frac{1}{2} S_z^{-1} D_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \tag{53}$$

$$+ \left(\hat{\bar{H}}_k^{dd} - \hat{\bar{H}}_k^{du} \left(\hat{\bar{H}}_k^{uu} \right)^{-1} \hat{\bar{H}}_k^{ud} \right)^{-1} \left(\hat{\bar{H}}_k^{du} \left(\hat{\bar{H}}_k^{uu} \right)^{-1} \hat{\bar{H}}_k^{uz} - \hat{\bar{H}}_k^{dz} \right) z_k = 0$$

Therefore, $\Delta V_D(z_k)$ is expressed in terms as the adaptive estimator error dynamics of the matrix $\bar{H}_k$ and the relationship between $Q(z_k, u_k, d_k), \tilde{\bar{h}}_k$ and $\tilde{J}_k$, (52) revealing

$$\Delta V_D(z_k) \le -(1 - 3k^*)\|z_k\|^2 + 3\|B_{zk}\tilde{u}_k\|^2 + 3\|D_{zk}\tilde{w}_k\|^2 \le -(1 - 2k^*)\|z_k\|^2$$
$$+ 3B_M^2 \left\| \left(\hat{\bar{H}}_k^{uu} - \hat{\bar{H}}_k^{ud} \left(\hat{\bar{H}}_k^{dd} \right)^{-1} \hat{\bar{H}}_k^{du} \right)^{-1} \left(\hat{\bar{H}}_k^{ud} \left(\hat{\bar{H}}_k^{dd} \right)^{-1} \hat{\bar{H}}_k^{dz} - \hat{\bar{H}}_k^{uz} \right) z_k \right.$$
$$\left. + \frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \right\|^2 + 3D_M^2 \left\| \left(\hat{\bar{H}}_k^{dd} - \hat{\bar{H}}_k^{du} \left(\hat{\bar{H}}_k^{uu} \right)^{-1} \hat{\bar{H}}_k^{ud} \right)^{-1} \right.$$
$$\left. \times \left(\hat{\bar{H}}_k^{du} \left(\hat{\bar{H}}_k^{uu} \right)^{-1} \hat{\bar{H}}_k^{uz} - \hat{\bar{H}}_k^{dz} \right) z_k - \frac{1}{2} S_z^{-1} D_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}} \right\|^2$$
$$\le -(1 - 3k^*)\|z_k\|^2 \tag{54}$$

At final step, combining the equation (51) and (52), we have

$$\Delta V \le -(1 - 3k^*)\|z_k\|^2 - (1 - \alpha_h^2)\Delta W_{\min}^2 \left\| \tilde{\bar{h}}_k \right\|^2 \tag{55}$$

Since $0 < k^* < 1/3$ and $0 < \alpha_h < 1$, ΔV is negative definite with V being positive definite. Also, observe that $\left|\sum_{k=k_0}^{\infty} \Delta V_k\right| = |V_\infty - V_0| < \infty$ since $\Delta V < 0$ as long as (55) holds. Now, taking the limit as $k \to \infty$, the system states z_k and $\tilde{\tilde{h}}_k$ converges to zero asymptotically. In other words, as $k \to \infty, z_k \to 0, \tilde{\tilde{h}}_k \to 0$, then $\hat{J}_k \to J_k^*$. Since optimal control $u_k^* = -\frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial J_{k+1}}{\partial z_{k+1}}$, $\hat{u}_{2k} = -\frac{1}{2} R_z^{-1} B_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}}$ and $\hat{u}_{1k} = \hat{u}_{2k}$ (Lemma 2), then $\hat{u}_{2k} \to u_k^*$ and $\hat{u}_{1k} \to u_k^*$ when $\hat{J}_k \to J_k$. Also since optimal disturbance input $d_k^* = \frac{1}{2} S_z^{-1} D_{zk}^T \frac{\partial J_{k+1}}{\partial z_{k+1}}$, $\hat{d}_{2k} = \frac{1}{2} S_z^{-1} D_{zk}^T \frac{\partial \hat{J}_{k+1}}{\partial z_{k+1}}$ and $\hat{d}_{1k} = \hat{d}_{2k}$ (Lemma 2), then $\hat{d}_{2k} \to d_k^*$ and $\hat{d}_{1k} \to d_k^*$ when $\hat{J}_k \to J_k$.

Remark 6: In traditional ADP (Tamimi et al., 2007), policy and value iteration methods are employed during a fixed sampling interval, and system states and inputs are recalculated and stored for learning optimal strategy. For example, during time $[kT_s, (k+1)T_s]$ the system states J_{k+1}^i and u_k^i will be recalculated and stored for learning optimal strategy J_k^* and u_k^* when iteration index changes from 1 to ∞, i.e. $i = 1,2,...,\infty$. Consequently, traditional ADP value and policy iterations can consume significant amount of time which may not be practically viable in real-time environment. However, the proposed novel stochastic optimal design does not require value and policy iterations while the cost function and control input are updated once every sampling interval and therefore will be referred to as time-based ADP. Only the measured real-time data is used to tune the cost function, optimal control inputs and optimal disturbance inputs, i.e. when $k \to \infty$, $\hat{J}_k \to J_k^*$, $\hat{u}_k \to u_k^*$, and $\hat{d}_k \to d_k^*$.

4. Simulation Results

In this section, stochastic optimal control of NCS quadratic zero-sum games is evaluated. At the same time, the standard optimal strategy of NCS quadratic zero-sum games with known dynamics is also simulated for comparisons.

Example: The continuous-time version of a batch reactor system dynamics is given as (Dacic et al., 2007)

$$\dot{x} = \begin{bmatrix} 1.38 & -0.2077 & 6.715 & -5.676 \\ -0.5814 & -4.29 & 0 & 0.675 \\ 1.067 & 4.273 & -6.654 & 5.893 \\ 0.048 & 4.273 & 1.343 & -2.104 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 5.679 & 0 \\ 1.136 & -3.146 \\ 1.136 & 0 \end{bmatrix} u$$

$$+ \begin{bmatrix} 10 & 0 & 10 & 0 \\ 0 & 5 & 0 & 5 \end{bmatrix}^T d \tag{56}$$

where $x \in \mathbb{R}^{4\times 1}$, $u \in \mathbb{R}^{2\times 1}$ and $d \in \mathbb{R}^{2\times 1}$. It is important to note that this example has developed over the years as a benchmark example for NCS, see e.g. (Dacic et al., 2007; Dacic and Nesic, 2007; Walsh et al., 1999).

The parameters of this NCS quadratic zero-sum games are selected as (Hu and Zhu, 2003)

1. The sampling time: $T_s = 0.8\sec$;

2. The bound of delay is two, i.e. $b = 2$;

3. The mean value of random delays: $E(\tau_{sc}) = 0.5\sec$, $E(\tau) = 1.1\sec$;

4. Packet losses follow Bernoulli distribution with $p = 0.3$. These values can be changed.

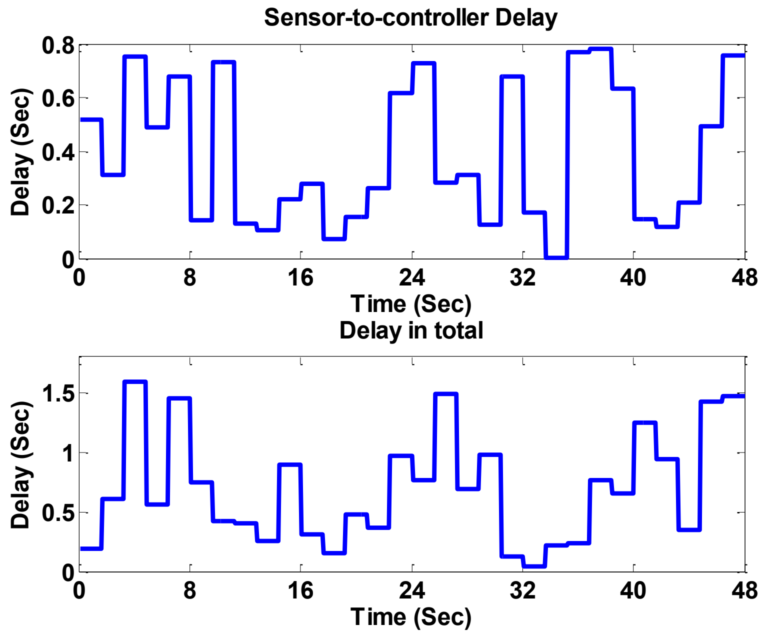


Fig 5. The distribution of random delays in NCS

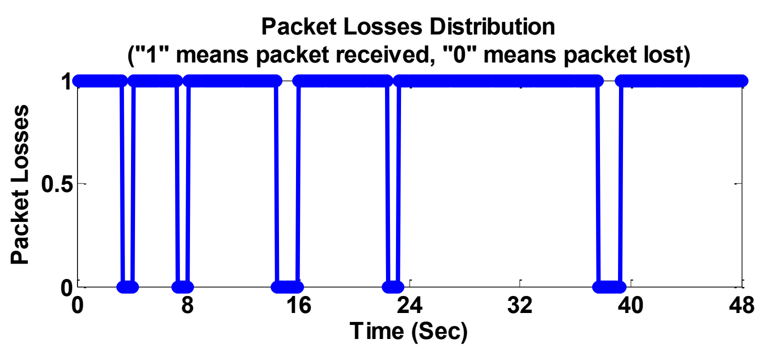


Fig 6. The distribution of packet losses

The distribution of random delays, includes sensor-to-controller delay τ_{sc} with a total delay of τ, are shown in Figure 5 and the packet losses are shown in Figure 6. Incorporating the random delays $\tau(t)$ and packet losses $\gamma(t)$ to batch reactor system (56), the original time-invariant system (53) was represented as a slowly time-varying linear NCS given by (3). For instance, when $k = 20, t = kT_s = 16\sec$, networked-induced delay $\tau_k = 0.41\sec, \tau_{k-1} = 0.45\sec, \tau_{k-2} = 0.85\sec$, packet losses $\gamma_k = 1, \gamma_{k-1} = 1,\ \gamma_{k-2} = 0$, the NCS dynamics can be calculated based on (3) as

$$z_{k+1} = \begin{bmatrix} 4.82 & 0.89 & 3.45 & -1.71 & 0.93 & -1.53 & 0 & 0 & 11.68 & -0.66 & 0 & 0 \\ -0.47 & 0.08 & -0.30 & 0.28 & 0.10 & 0.13 & 0 & 0 & -1.07 & 0.30 & 0 & 0 \\ 0.19 & 0.86 & 0.30 & 0.60 & 0.97 & -0.16 & 0 & 0 & 0.79 & 1.30 & 0 & 0 \\ -0.29 & 0.78 & -0.05 & 0.79 & 0.89 & 0.0004 & 0 & 0 & -0.39 & 1.30 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} z_k + \begin{bmatrix} 0.93 & -2.44 \\ 1.44 & 0.15 \\ 2.80 & -0.79 \\ 2.73 & -0.19 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} u_k + \begin{bmatrix} 20.03 & -1.90 \\ -1.45 & 1.67 \\ 3.31 & 3.52 \\ 0.26 & 4.32 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} d_k$$

where $z_k = [x_k \; u_{k-1} \; u_{k-2} \; d_{k-1} \; d_{k-2}]^T \in \mathbb{R}^{12\times 1}$. When $k = 54, t = kT_s = 43.2\sec$, networked-induced delay $\tau_k = 0.39\sec, \tau_{k-1} = 0.91\sec, \tau_{k-2} = 0.87\sec$, packet losses $\gamma_k = 1, \gamma_{k-1} = 1, \gamma_{k-2} = 1$, the NCS dynamics become

$$z_{k+1} = \begin{bmatrix} 4.82 & 0.89 & 3.45 & -1.71 & 1.76 & -3.43 & 0.03 & -0.04 & 26.48 & -1.99 & 0.32 & -0.02 \\ -0.47 & 0.08 & -0.30 & 0.28 & 0.60 & 0.27 & 0.001 & 0.004 & -2.27 & 1.06 & -0.03 & 0.007 \\ 0.19 & 0.86 & 0.30 & 0.60 & 2.91 & -0.49 & 0.02 & -0.004 & 2.44 & 3.69 & 0.02 & 0.03 \\ -0.29 & 0.78 & -0.05 & 0.79 & 2.76 & -0.11 & 0.02 & -0.0005 & -0.34 & 3.99 & -0.01 & 0.03 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} z_k + \begin{bmatrix} 0.07 & -0.49 \\ 0.94 & 0.01 \\ 0.84 & -0.45 \\ 0.84 & -0.08 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} u_k + \begin{bmatrix} 4.91 & -0.56 \\ -0.22 & 0.91 \\ 1.65 & 1.02 \\ 0.22 & 1.60 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} d_k$$

First, Figure 7 depicts the performance of the conventional stochastic optimal strategy for NCS quadratic zero-sum games with known dynamics and information of communication imperfections (i.e. network-induced delays and packet losses) obtained by solving the Game-theoretic Riccati Equation (GRE) in backward-in-time manner. The control and disturbance inputs can force the state regulation errors converge to zeros while ensuring the NCS stable in the mean when communication imperfections (i.e. network-induced delays and packet losses) are accurately known.

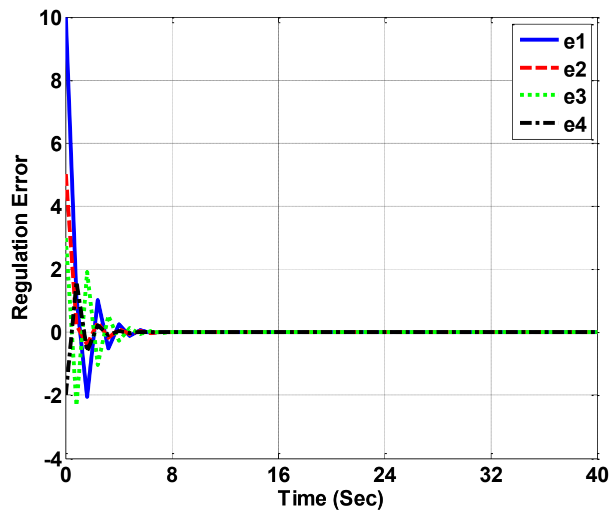


Fig 7. Performance of the conventional stochastic optimal control scheme with known system dynamics and communication imperfections.

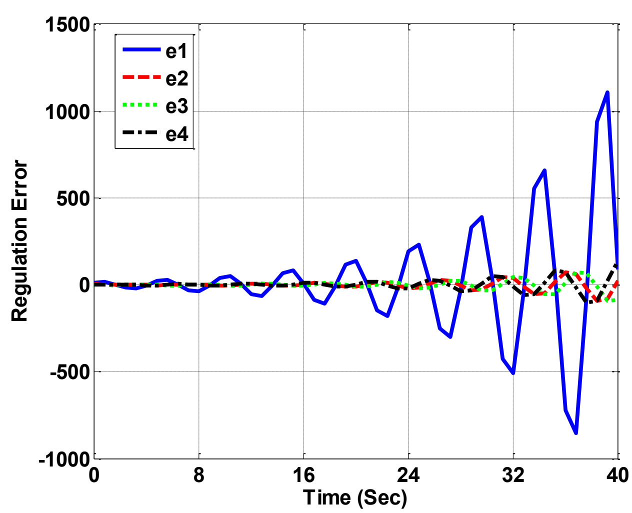


Fig 8. Performance of the ADP value iteration-based scheme [14] in the presence of communication imperfections.

Next, by using the ADP value iteration (VI) method and modifying the strategy (Tamimi et al., 2007), the control and disturbance inputs, $u_k = \begin{bmatrix} -0.019 & 0.186 & 0.026 & 0.172 \\ -0.040 & -0.011 & -0.029 & 0.012 \end{bmatrix} x_k$ and $d_k = \begin{bmatrix} -0.149 & -0.063 & -0.114 & 0.023 \\ 0.077 & -0.273 & -0.003 & -0.269 \end{bmatrix} x_k$, are designed. The ADP VI scheme normally does not require any system dynamics and information of communication imperfections. However, the ADP VI based control cannot maintain the batch reactor system stable in the mean in the presence of communication imperfections (i.e. network-induced delays and packet losses) as shown in Figure 8.

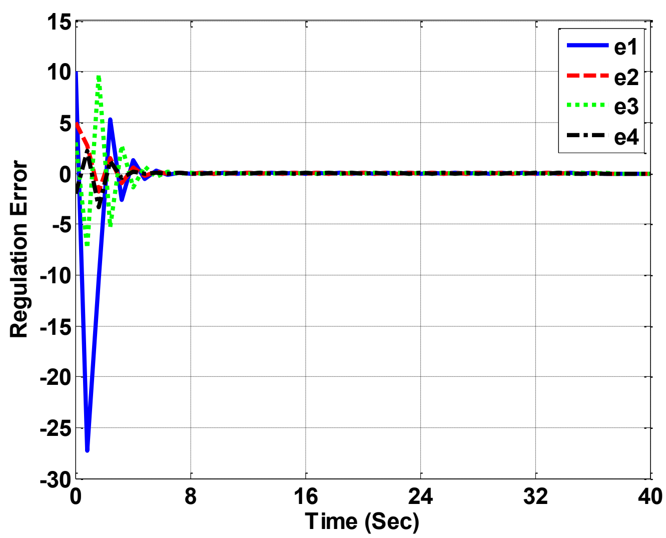
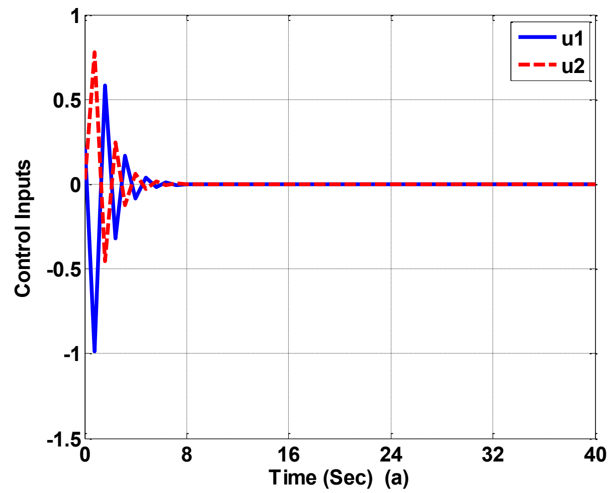


Fig 9. Performance of the stochastic optimal controller for NCS with uncertain system dynamics and with communication imperfections.

Finally, the proposed adaptive stochastic optimal strategy is implemented for the NCS quadratic zero-sum games with unknown system dynamics in presence of communication imperfections (i.e. network-induced delays and packet losses). The augment state is generated as $z_k = [x_k \; u_{k-1} \; u_{k-2} \; d_{k-1} \; d_{k-2}]^T \in \mathbb{R}^{12\times 1} \; \forall k$ or $w_k = [z_k \; u_k \; d_k]^T \in \mathbb{R}^{16\times 1}$. The initial stabilizing policy and disturbance input for the algorithm were selected as $u_0(z_k)=\begin{bmatrix} 0.12 & -0.02 & 0.08 & -0.06 & -0.12 & -0.05 & -0.024 & -0.029 & 0.46 & -0.214 & 0.236 & -0.064 \\ -0.045 & 0.015 & -0.03 & 0.035 & 0.06 & 0.008 & 0.027 & 0.006 & -0.13 & 0.108 & -0.073 & 0.051 \end{bmatrix} z_k$ and $d_0(z_k)=\begin{bmatrix} -0.95 & -0.25 & -0.70 & 0.28 & -0.51 & 0.61 & -0.234 & 0.274 & -4.591 & 0.055 & -2.07 & 0.02 \\ 0.41 & -0.12 & 0.25 & -0.297 & -0.77 & -0.135 & -0.253 & -0.077 & 1.438 & -1.33 & 0.72 & -0.47 \end{bmatrix} z_k$ while the regression function for value-function was generated as $\{w_1^2, w_1 w_2, w_1 w_3, \ldots, w_2^2, \ldots, w_{15}^2, \ldots, w_{16}^2\}$ as (21).



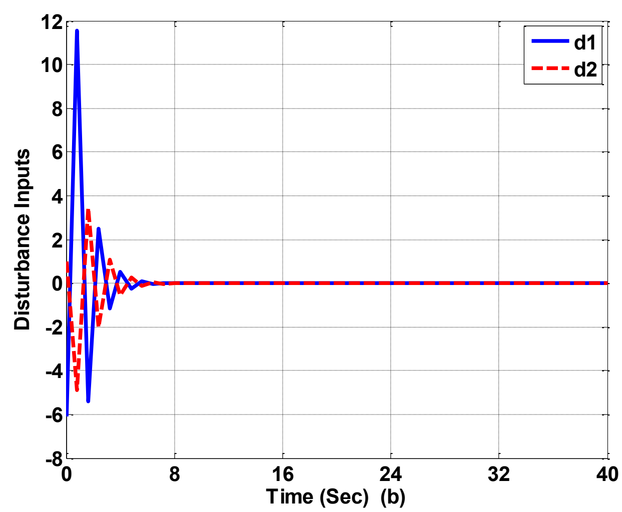


Fig 10. Performance of the proposed optimal controller: (a) Control inputs $u=(u_1\ u_2)^T \in \mathbb{R}^{2\times 1}$; (b) Disturbance input $d=(d_1\ d_2)^T \in \mathbb{R}^{2\times 1}$.

The design parameter for the value-function $V(z_k, u_k, d_k)$ was selected as $\alpha_h = 10^{-6}$ while initial parameters for the adaptive estimator were set to zeros at the beginning of the simulation. The simulation was run for 40 seconds, and for the first 22 seconds, exploration noise with mean zero and variance 0.08 was added to the system in order to ensure the persistency of excitation (PE) condition holds (Lemma 1).

In Figures 9 through 11, the performance of proposed adaptive estimation based optimal strategy is evaluated. As shown in Figure 8, the proposed adaptive estimation based optimal strategy can also force the NCS quadratic zero-sum games state regulation errors converges to zero even when the NCS dynamics are unknown which implies that the proposed strategy can make the NCS closed-loop system stable in the mean. Due to an initial online tuning phase needed to learn optimal control and disturbance inputs, there is a slight overshoot at the beginning. In Figure 10 (a), (b), the control and disturbance inputs of proposed adaptive estimation based optimal strategy are shown. Proposed adaptive estimation based optimal control and disturbance inputs can make the NCS states converge to zero quickly.

Estimated value-function for NCS quadratic zero-sum games are shown in Figure 11. Estimated value-function is defined in (23) as $\hat{V}(z_k, u_k, d_k) = [z_k^T \ u_k^T \ d_k^T]^T \hat{\bar{H}}_k [z_k^T \ u_k^T \ d_k^T]$. If all the states are equal to zeros except z_1, z_2, the estimated value-function is shown as Figure 11 (a), while Figure 11 (b) illustrates the estimated value-function when all the states are equal to zeros except z_3, z_4. It is important to note two key points. First, based on definition of estimated value-function, if all the states are equal to zero, the estimated value-function can be zero. Otherwise, estimated value-function should be a quadratic positive value. This is why a valley is observed in Figure 11 (a)(b). Second, the proposed stochastic optimal strategy is designed to minimize the estimated value-function.

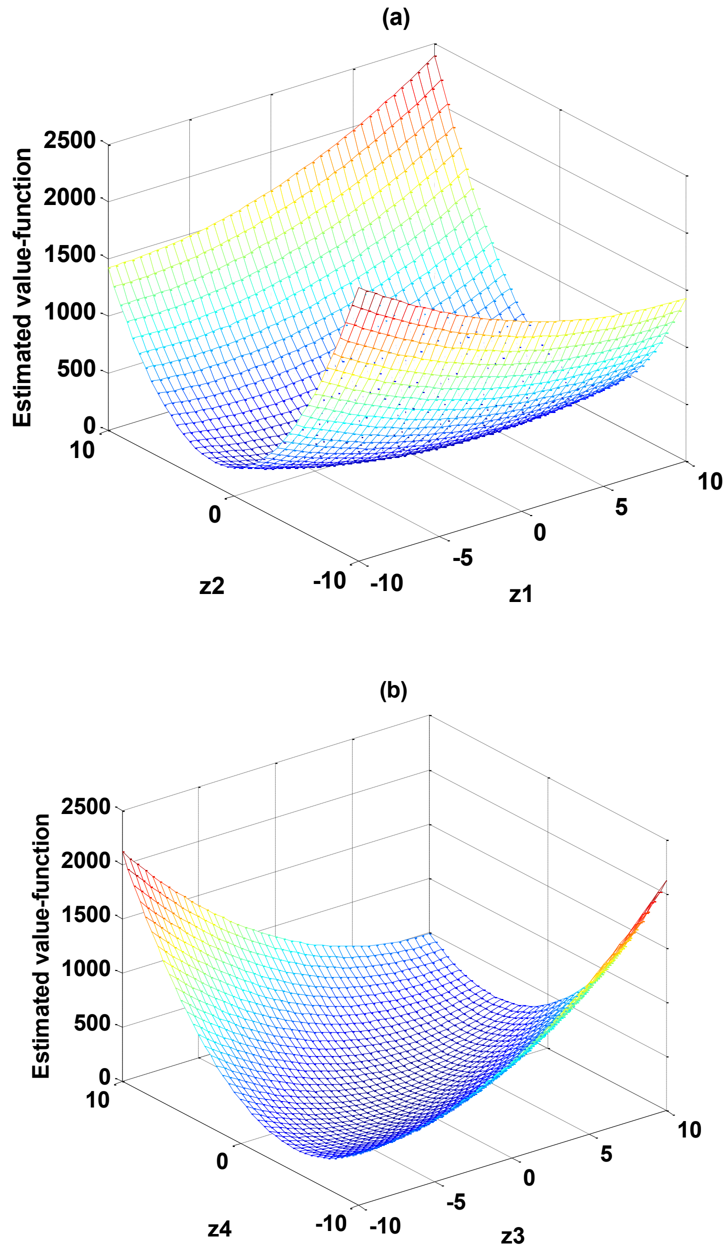


Fig 11. Estimated value-function: (a) in the z_1, z_2 direction (b) in the z_3, z_4 direction

Based on the results presented in Figures 6 through 10, and after a short initial tuning time, the proposed adaptive estimation based stochastic optimal strategy for NCS quadratic zero-sum games with uncertain dynamics and imprecise information on communication imperfections will have nearly the same performance as that of the conventional optimal strategy for NCS quadratic zero-sum games when system dynamics and communication imperfection are known.

5. Conclusions

In this work, a direct adaptive dynamics programming scheme is proposed which combines the adaptive estimation and the concept of ADP to solve the Bellman equation in real time for the stochastic optimal regulation of NCS quadratic zero-sum games with communication imperfections (i.e. network-induced delays and packet losses). The availability of past state values ensured that NCS quadratic zero-sum games dynamics were not needed when an adaptive estimator generates an estimated value-function and the novel stochastic optimal control and disturbance laws based on the estimation of $V(z_k, u_k, d_k)$. An initial admissible control and disturbance policies ensured that the adaptive estimator learns the value-function $V(z_k, u_k, d_k)$ and the matrix $\underset{\tau,\gamma}{E}(H_k)$, cost function and optimal control and disturbance signal online. Initial overshoots are observed due to the online learning phase while they quickly die with time. All adaptive estimation parameters were tuned online using proposed update laws and Lyapunov theory demonstrated the asymptotic stability (AS) of the closed-loop system.

References

[1] Al-Tamimi, A., Lewis, F.L., and Abu-Khalaf, M. (2007). Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control. Automatica, vol. 3, pp. 471-481.

[2] Åstrom, K.J. (1970). Introduction to Stochastic Control Theory. Academic, New York.

[3] Basar, T., and Bernhard, P. (1995). H-infinite Optimal Control and Related Minimax Design Problems. Birkhauser, Basel.
[4] Basar, T., and Olsder, G. J. (1995). Dynamic Non-cooperative Game Theory. 2nd Edition, Academic, New York.
[5] Carnevale, D., Teel, A.R., and Nesic, D. (2007). A Lyapnov proof of improved maximum allowable transfer interval for networked control systems. IEEE Trans. Autom. Contr., vol. 52, pp. 892-897.
[6] Chen, H. F., and Guo, L. (1991). Identification and Stochastic Adaptive Control. Birkauser, Cambridge, MA.
[7] Dacic, D.B., and Nesic, D. (2007). Quadratic stabilization of linear networked control systems via simultaneous protocol and controller design. Automatica, vol. 43, pp. 1145-1155.
[8] Dierks, T., and Jagannathan, S. (2009). Optimal control of affine nonlinear discrete-time systems with unknown internal dynamics. in Proc. Conf. on Decision and Contr., pp. 6750-6755.
[9] Dierks, T., and Jagannathan, S. (2011). Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update. IEEE Trans. Neur. Net., conditionally accepted.
[10] Dragan, V., and Morozan, R. (1997). Global solution to game-theoretic riccati equation of stochastic control. J. Diff. Equ., vol. 138, pp. 328-350.
[11] Goldsmith, A. (2003). Wireless Communications. Cambridge University Press, Cambridge, U.K.
[12] Halevi, Y., and Ray, A. (1988). Integrated communication and control systems: part I—analysis. J. Dynamic Syst., Measure. Contr., vol. 110, pp. 367-373.
[13] Hu, S.S., and Zhu, Q.X. (2003). Stochastic optimal control and analysis of stability of networked control systems with long delay. Automatica, vol. 39, pp. 1877-1884.
[14] Ioannou, P.A., and Sun, J. (1996). Robust adaptive control. Upper Saddle River, NJ: Prentice-Hall.
[15] Jagannathan, S. (2006). Neural Network Control of Nonlinear Discrete-time Systems, CRC Press.
[16] Kreisselmeier, G. (1986). Adaptive control of a class of slowly time-varying plants. Syst. Control Lett., vol. 8, pp. 97-103.
[17] Krtolica, R., Ozguner, U., Chan, H., Goktas, H., Winkelman, J. and Liubakka, M. (1994). Stability of linear feedback systems with random communication delays. Int. J. Contr., vol. 59, pp. 925-953.
[18] Lewis, F.L., and Syrmos, V.L. (1995). Optimal Control, 2nd Edition, Wiley, New York.
[19] Lian, F., Moyne, J. and Tilbury, D. (2003). Modeling and optimal controller design of networked control systems with multiple delays. Int. J. Contr., vol. 76, pp. 591-606.
[20] Lian, F., Moyne, J. and Tilbury, D. (2001). Analysis and modeling of networked control systems: MIMO case with multiple time delays. in Proc. Amer. Contr. Conf., pp. 4306-4312.

[21] Liou, L.W., and Ray, A. (1991). A stochastic regulator for integrated communication and control systems: part I—formulation of control law. ASME Jour. of Dynamic Syst., Measureme. and Contr., vol. 4, pp. 604-611.
[22] Littman, M. L. (1994). Markov games as a framework for multiagent reinforcement learning. in Proc. Int. Conf. on Machine Learni., pp. 157-163.
[23] Middleton, R.H., and Goodwin, G.C. (1988). Adaptive control of time-varying linear systems. IEEE Trans. Automat. Contr., vol. 33, pp. 150-155.
[24] Nilsson, J., Bernhardsson, B., and Wittenmark, B. (1998). Stochastic analysis and control of real-time systems with random time delays. Automatica, vol. 34, pp. 57-64.
[25] Walsh, G.C., Beldiman, O. and Bushnell, L. (1999). Asymptotic behavior of networked control systems. in Proc. IEEE international Conf. on Control App., pp. 1448-1453.
[26] Walsh, G.C., Ye, H., and Bushnell, L. (1999). Stability analysis of networked control systems. in Proc. Amer. Control Conf., pp. 2876-2880.
[27] Watkins, C., (1989). Learning from delayed rewards, Ph.D. Thesis, Cambridge,England: Cambridge University.
[28] Werbos, P.J. (1991). A menu of designs for reinforcement learning over time. MA: MIT Press.
[29] Wonham, W.M. (1968). On a matrix Riccati equation of stochastic control. SIAM J. Contr., vol.6, pp. 681-697.
[30] Wu, J. and Chen, T. (2007). Design of networked control systems with packet dropouts. IEEE Trans. Automat. Control, vol. 52, pp. 1314-1319.
[31] Xu, H., and Jagannathan, S. (2011). Stochastic optimal control of unknown linear networked control system using Q-learning methodology. in Proc. Amer. Contr. Conf., pp. 2819-2824.
[32] Xu, H., Jagannathan, S., and Lewis, F. L., (2011). Stochastic optimal control of unknown linear networked control system in the presence of random delays and packet losses. To appear in Automatica.
[33] Zhang, W., Branicky, M.S., and Phillips, S. (2001). Stability of networked control systems. IEEE Contr. Syst. Mag., vol. 21, pp. 84-99.

PAPER III

STOCHASTIC OPTIMAL CONTROLLER DESIGN FOR UNCERTAIN NONLINEAR NETWORKED CONTROL SYSTEM VIA NEURO DYNAMIC PROGRAMMING

H. Xu and S. Jagannathan

Abstract - The stochastic optimal control design for the nonlinear networked control system (NNCS) with uncertain system dynamics is a challenging problem due to the presence of both system nonlinearities and communication network imperfections such as random delays and packet losses, which are considered unknown. In the recent literature, neuro dynamic programming (NDP) techniques, based on value and policy iterations, have been widely used to solve the optimal control of general affine nonlinear systems with known partial knowledge of system dynamics. However, for real-time control, value and policy iterations-based methodology is not suitable and time-based NDP techniques are preferred. In addition, output feedback based controller designs are preferred for implementation. Therefore, in this paper, a novel NNCS representation incorporating the system uncertainties and network imperfections is introduced first by using input and output measurements for facilitating output feedback. Then, an online neural network (NN) identifier is introduced to estimate the control coefficient matrix. Subsequently, the critic and action NNs are employed along with the NN identifier to determine the forward-in-time, time-based stochastic optimal control of NNCS without using value and policy iterations. Here, the value function and control inputs are updated once every sampling instant. By using novel NN weight update laws, Lyapunov theory is used to show that all the closed-loop signals and NN weights are uniformly ultimately bounded (UUB) in the mean while the approximated control input converges close to its

target value with time in the mean. Simulation results are included to show the effectiveness of the proposed scheme.

I. INTRODUCTION

Feedback control systems with control loops closed through a real-time communication network are called networked control systems (NCS) [1]. In NCS, a communication packet carries the reference input, plant output, and control input which are exchanged by using a communication network among control system components such as sensor, controller and actuators. A NCS results in reduced system wiring with ease of system diagnosis and maintenance, and has increased system agility. Adding a communication network in the feedback control loop, however, brings challenging issues.

First main issue is the network-induced delay in the control loop that occurs when exchanging data among devices connected to the shared medium. The delay, either constant or random, can degrade the performance of the control system and even destabilize the system when the delay is not explicitly considered in the design process. Second main issue is the packet losses in the communication network due to unreliable path transmission which can cause a loss in control input resulting in instability.

Recently, Walsh [2] proposed a scheduling protocol and analyzed the asymptotic behavior of nonlinear NCS (or NNCS). Polushin [3] proposed a model-based stabilizing control for NNCS. Using model predictive control theory [4], Liu [5] proposed a predictive stable control for NNCS. However, the only objective of these controller designs [2-5] is to make the NNCS stable when the dynamics are considered known. In general, optimality is generally preferable for NCS and especially for NNCS, which is very difficult to attain. The unknown dynamics and network imperfections in the case of NNCS further complicates the optimal controller design.

List of Abbreviations

Symbol	Quantity
τ_{sc}	Sensor-to-controller delay
τ_{ca}	Controller-to-actuator delay
γ	Indicator of packet losses
T_s	Sampling time
$\bar{d}T_s$	Upper bound on delay
z_k	Augmented states of NCS at time *k*
y_k^o	Modified state vector with current output and pervious inputs
V_k	Stochastic value function at time *k*
W_C	Target weights of NN-identifier
$\hat{W}_{Ck}$	Estimated weights of NN-identifier at time *k*
e_{yk}	Identification errors at time *k*
W_V	Target weights of Critic NN
$\hat{W}_V$	Estimated weights of Critic NN at time *k*
e_{Vk}	Residual error
W_u	Target weights of Action NN
$\hat{W}_{uk}$	Estimated weights of Action NN at time *k*
e_{Vk}	Action NN estimation error
$\alpha_C, \alpha_V, \alpha_u$	Tuning parameters for NN-identifier, Critic NN and Action NN respectively
$\varepsilon_{Ck}, \varepsilon_{Vk}, \varepsilon_{uk}$	Reconstruction errors for NN-identifier, Critic NN and Action NN respectively

Table 1. List of abbreviations for NNCS

On the other hand, Neuro dynamics programming (NDP) and adaptive/approximate dynamics programming (ADP) techniques proposed by Bertsekas and Tsitsiklis [8] and Werbos [6] respectively, intend to solve optimal control problem forward-in-time similar to a standard Riccati equation-based backward-in-time solution for linear systems. In NDP and ADP, one combines adaptive critics, a reinforcement learning technique, with dynamic programming [22-24]. Zhang et al. [25] introduced near-optimal control of affine nonlinear discrete-time systems with control constraints by using iterative ADP algorithm. Greedy ADP iteration algorithm is derived to obtain optimal tracking control for discrete-time nonlinear system in [26]. Recently Lewis and Vrabie [9] introduced the methods of reinforcement learning and NDP for feedback control to obtain the optimal controller for both linear time-invariant and nonlinear system with partially unknown dynamics by using value and policy iterations.

In contrast, in [10], neural networks (NNs) are utilized to solve the optimal regulation of a discrete-time nonlinear system in an offline manner by assuming that there are no reconstruction errors. Besides ignoring the online approximator (OLA) reconstruction errors, complete dynamics are needed to implement offline NN training. To overcome the iterative offline training methodology, authors in [11] used two NNs to solve the Hamilton Jacobi Bellman (HJB) equation in forward-in-time for time-based optimal control of a class of general nonlinear affine discrete-time systems. However, these papers [9-11, 25-26] are not applicable for NNCS since the effects of delays and packet losses are not considered while state measurement is assumed. Moreover, value and policy based schemes are not suitable for hardware implementation.

The network imperfections such as delays and packet losses can cause instability [1] if they are not considered carefully which in turn make the optimal controller design for NNCS more involved and different than [9-11]. Although NDP is an effective technique to solve the optimal control of NNCS, traditional NDP techniques

[9,16] require partial knowledge of the system dynamics which becomes a problem for NNCS due to the presence of unknown random delays and packet losses. In addition, an NDP technique using value and policy iterations [9,16] is not suitable for real-time control since the number of iterations needed for convergence within a sampling interval is unknown. Also, in some cases [9,17], a model may be needed to iterate the value and policies. Therefore, the standard HDP value and policy iteration methods [9,17] cannot be utilized for NNCS and a novel scheme is necessary.

Besides the need to relax value and policy iterations, it would be desirable to be able to convert the system dynamics in state space form to the dynamics in terms of input/output since the system states are normally not measurable. Such techniques belong to the field of data-based control techniques [12], where the control input depends on output/input data measured directly from the plant. To the best knowledge of the authors, there are no known NDP methods developed in the literature for the control of unknown nonlinear NCS in the input-output form.

Thus, in this paper, a novel time-based NDP algorithm is derived for NNCS with uncertain dynamics and in the presence of network imperfections such as random delays and packet losses which are normally unknown. To learn the partial dynamics of NNCS, an online neural network (NN) identifier is introduced first. Then by using an initial stabilizing control, a critic NN is tuned online to learn the value function of NNCS since solving the discrete-time Hamilton-Jacobi-Bellman (HJB) equation requires system dynamics. Subsequently, an action NN is utilized to minimize the value function based on the information provided by the critic NN and NN identifier. Therefore, the proposed novel input-output feedback-based NDP algorithm relaxes the need for system dynamics and information on random delays and packet losses. Value and policy iterations are not used and the value function and control input are updated once a sampling instant making the proposed NDP scheme a time-based model-free optimal controller for NNCS.

The main contribution of this work includes a time-based NDP optimal control scheme using output feedback without utilizing value and policy iterations for uncertain NNCS. Closed-loop stability is demonstrated by selecting novel NN update laws. This paper is organized as follows. First, Section II presents the NNCS background and output/input system representation for NNCS. A novel online optimal control scheme with online identifier is proposed in Section III for unknown NNCS with network imperfections while the stability of this optimal control scheme is verified by using Lyapunov theory. Section IV illustrates the effectiveness of proposed schemes via numerical simulations and Section V provides concluding remarks.

II. NONLINEAR NETWORKED CONTROL SYSTEM BACKGROUND

A. NNCS Structure

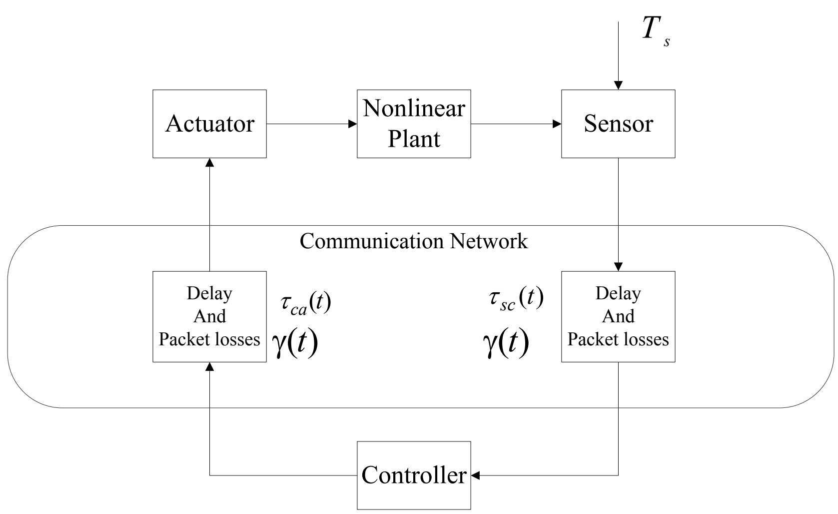


Fig 1. Nonlinear Networked Control System (NNCS)

The NNCS structure considered in this paper is shown in Figure 1 where the feedback control loop is closed over the communication network. Due to unreliable communication network, networked-induced delays and packet losses are included in this structure such as: (1) $\tau_{sc}(t)$: sensor-to-controller delay, (2) $\tau_{ca}(t)$: controller-to-actuator delay, and (3) $\gamma(t)$: indicator of packet losses at the actuator.

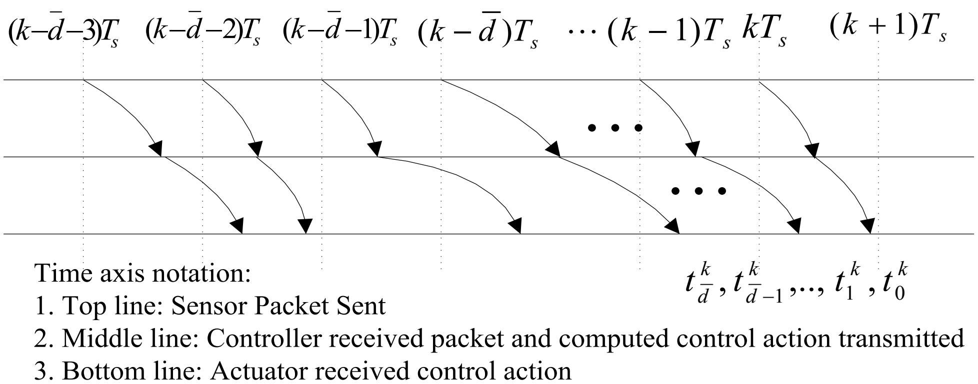


Fig 2. Timing diagram of signals transmitting in NCS.

Next the following assumption is needed that is consistent with the literature in NCS [13,19]:

Assumption 1:

a). Sensor is time-driven while the controller and actuator are event-driven [14].

b). Communication network is a wide area network so that the two network-induced delays are considered independent, ergodic and unknown whereas their probability distribution functions are considered known [13,19].

c). The total delay (sum of both types) is bounded [13] while the initial state of the nonlinear system is deterministic [19].

B. NNCS System Dynamics Representation

In this paper, a continuous-time affine nonlinear system of the form $\dot{x} = f(x) + g(x)u$ and $y = Cx$ is considered, where x, y and u denotes system state, output and input vector while $f(\bullet)$ and $g(\bullet)$ are smooth nonlinear functions of the state and C is the output matrix. When the random delays and packet losses of the communication network are considered, the control input $u(t)$ is delayed and can be lost at times due to packet losses. Therefore the nonlinear system after the incorporation of delay and packet loss effects can be expressed as

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + \gamma(t) g(x(t)) u(t - \tau(t)) \\ y(t) &= Cx(t) \end{aligned} \tag{1}$$

where $\gamma(t) = \begin{cases} \mathbf{I}^{n\times n} & \text{if control input is received by the actuator at time t} \\ \mathbf{0}^{n\times n} & \text{if control input is lost at time t} \end{cases}$ with $\mathbf{I}^{n\times n}$ is identity matrix, $u(t-\tau(t))$ is the delayed control inputs $x(t) \in \Re^n, u(t) \in \Re^m, y(t) \in \Re^n$, $f(x) \in \Re^n, g(x) \in \Re^{n\times m}$ and $C \in \Re^{n\times n}$ being invertible. From Assumption 1, sum of network-induced delays is considered bounded above, i.e. $\tau(t) = \tau_{sc}(t) + \tau_{ca}(t) < \bar{d}T_s$ where $\bar{d}$ represents the delay bound with T_s being the sampling interval.

For wireless network-based NNCS, the controller has to convert the control inputs into packets [21] and transmit them to the actuator through the communication network. Then actuator applies the control inputs in response to a received packet from the controller. Consequently, the controller for NNCS is normally referred to as event-driven and the control input $u(t)$ to the plant is considered as a piecewise constant [1,7] during a sampling interval. This actual control input during the sampling interval $[kT_s, (k+1)T_s)$

without delays can be obtained as $u(t) = u_k[\delta(t - kT_s) - \delta(t - (k+1)T_s)]$ where $\delta(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$. This means that a piecewise constant control input u_k is applied to the continuous-time plant during the time interval $[kT_s, (k+1)T_s)$ (i.e. $u(t) = u_k$, $t \in [kT_s, (k+1)T_s), \forall k$).

For general representation of NCS, Consider the communication imperfections and according to Assumption 1, there are at most $\bar{d}$ number of current and previous control input values to arrive at the actuator in the form of packets. If several control inputs arrive at the same time, only the latest control input will be selected during any sampling interval $[kT_s, (k+1)T_s)$ $\forall k$ while the others are ignored. It is possible that some packets carrying the control inputs arrive without any delay. System states change at time instants $kT_s + t_i^k, \quad i = 0,1,\ldots,\bar{d}-1$ and $t_i^k < t_{i-1}^k$ where $t_i^k = \tau_i^k - iT_s$ as illustrated in Figure 2 [13,19].

Since the controller is event-driven [14], (the controller updates the control input upon the receipt of the sensor measurement), the term u_k can be used to express the controller when the sensor signal x_k is transmitted to the controller. Thus, integration (1) over a sampling interval $[kT_s, (k+1)T_s)$ yields

$$
\begin{aligned}
&\int_{kT_s}^{(k+1)T_s} \dot{x}(t)dt = \int_{kT_s}^{(k+1)T_s} f(x(t))dt \\
&+ \gamma((k-\bar{d})T_s)\left(\int_{kT_s}^{\tau_d - \bar{d}T_s} g(x(t))dt\right)u((k-\bar{d})T_s) \\
&+ \cdots + \gamma((k-i)T_s)\left(\int_{\tau_{i+1}-(i+1)T_s}^{\tau_i - iT_s} g(x(t))dt\right)u((k-i)T_s) \\
&+ \cdots + \gamma(kT_s)\left(\int_{\tau_0}^{(k+1)T_s} g(x(t))dt\right)u(kT_s) \quad i = 1,2,\ldots,\bar{d}-1 \\
&y(kT_s) = Cx(kT_s)
\end{aligned}
$$

It is important to note that control signal $u(t)$ can be taken outside the integral since $u(t)$ is a piecewise constant during $[kT_s,(k+1)T_s)$ which appears to be the integration interval.

In other words, by incorporating the two random variables (i.e. random delays and packet losses), the above becomes a stochastic nonlinear discrete-time system given by

$$\begin{aligned} x_{k+1} &= Z_{\tau,\gamma}(x_k,u_{k-1},\cdots u_{k-\bar{d}}) + P_{\tau,\gamma}(x_k,u_{k-1},\cdots u_{k-\bar{d}})u_k \\ y_k &= Cx_k \end{aligned} \tag{2}$$

where $x(kT_s)=x_k, y(kT_s)=y_k, \gamma((k-i)T_s)=\gamma_{k-i}$ and $u((k-i)T_s)=u_{k-i}$ $i=0,1,2,...,\bar{d}$ are pervious control inputs, and $Z_{\tau,\gamma}(x_k,u_{k-1}\cdots u_{k-\bar{d}})=x_k+\int_{kT_s}^{(k+1)T_s} f(x(t))dt+\gamma_{k-\bar{d}}$ $\times\left(\int_{kT_s}^{\tau_d-\bar{d}T_s} g(x(t))dt\right)u_{k-\bar{d}-1}+\cdots+\gamma_{k-1}\left(\int_{\tau_2-2T_s}^{\tau_1-T_s} g(x(t))dt\right)u_{k-1}$, and $P_{\tau,\gamma}(x_k,u_{k-1},\cdots u_{k-\bar{d}})$ $=\gamma_k\left(\int_{\tau_0}^{(k+1)T_s} g(x(t))dt\right)$.

Using (2), define a new augment state variable $z_k=\left[x_k^T\ u_{k-1}^T\ \cdots\ u_{k-\bar{d}}^T\right]^T\in\Re^{n+\bar{d}m}$ and a modified state vector as $y_k^o=\left[y_k^T\ u_{k-1}^T\ \cdots\ u_{k-\bar{d}}^T\right]^T\in\Re^{n+\bar{d}m}$, where $u_{k-i}, i=1,...,\bar{d}$ are previous control inputs. Now equation (2) can be represented as

$$\begin{bmatrix} x_{k+1} \\ u_k \\ \vdots \\ u_{k-\bar{d}+1} \end{bmatrix} = \begin{bmatrix} Z_{\tau,\gamma}(x_k,u_{k-1},\cdots u_{k-\bar{d}}) \\ 0 \\ u_{k-1} \\ \vdots \\ u_{k-\bar{d}+1} \end{bmatrix} + \begin{bmatrix} P_{\tau,\gamma}(x_k,u_{k-1},\cdots u_{k-\bar{d}}) \\ I_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} u_k$$

$$\begin{bmatrix} y_k \\ u_{k-1} \\ \vdots \\ u_{k-\bar{d}} \end{bmatrix} = \begin{bmatrix} C & 0 & \cdots & 0 \\ 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_m \end{bmatrix} \begin{bmatrix} x_k \\ u_{k-1} \\ \vdots \\ u_{k-\bar{d}} \end{bmatrix}$$

The above stochastic equation can be written as

$$z_{k+1} = H(z_k) + L(z_k)u_k,\ y_k^o = C_o z_k \tag{3}$$

where $I_m \in \Re^{m\times m}$ is the identity matrix,

$$H(z_k) = \begin{bmatrix} Z_{\tau,\gamma}(x_k, u_{k-1}, \cdots u_{k-\bar{d}}) \\ 0 \\ u_{k-1} \\ \vdots \\ u_{k-\bar{d}+1} \end{bmatrix}, L(z_k) = \begin{bmatrix} P_{\tau,\gamma}(x_k, u_{k-1}, \cdots u_{k-\bar{d}}) \\ I_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and } C_o = \begin{bmatrix} C & 0 & \cdots & 0 \\ 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_m \end{bmatrix}.$$

It is important to note that $H(z_k) \in \Re^{n+\bar{d}m}$ and $L(z_k) \in \Re^{(n+\bar{d}m)\times m}$ are nonlinear matrix functions in terms of newly defined augmented state vector z_k. Since effects of random variables (i.e. random delays and packet losses) are included in (3), the NNCS dynamics (3) still remains as a stochastic affine nonlinear discrete-time system in terms of the augmented state vector. Representing the NNCS in terms of augmented state vector does not change the stochastic properties of the NNCS due to random delays and packet losses. The output matrix C_o will be invertible since C is considered as invertible.

Next, the NNCS can be expressed in the input-output form as

$$y_{k+1}^o = C_o H(C_o^{-1} y_k^o) + C_o L(C_o^{-1} y_k^o)u_k = F(y_k^o) + G(y_k^o)u_k \tag{4}$$

where $F(y_k^o) = C_o H(C_o^{-1} y_k^o)$, $G(y_k^o) = C_o L(C_o^{-1} y_k^o)$, $\|G(y_k^o)\|_F \le G_M$, with $\|\bullet\|_F$ denoting the Frobenius norm [15]. Here due to the effects of random variables (i.e. random delays and packet losses), $F(y_k^o)$ and $G(y_k^o)$ are stochastic real-valued functions and $F(y_k^o), G(y_k^o)$ can be calculated based on equation (2) and (3) provided information on random delays and packet losses are available. In other words, the network imperfections can make the nonlinear dynamics uncertain and stochastic requiring adaptive techniques.

C. Stochastic Value Function for NNCS

Since NNCS dynamics (4) is stochastic, the stochastic optimal adaptive controller is derived to minimize the stochastic value function [20][7] as

$$V_k = \underset{\tau,\gamma}{E}\left[\sum_{i=k}^{\infty}\left(x_i^T Q x_i + u_i^T R u_i\right)\right] \quad k = 0,1,2... \tag{5}$$

where Q and R are symmetric positive semi-definite and symmetric positive definite constant matrices respectively and $\underset{\tau,\gamma}{E}(\bullet)$ is the expectation operator (in this case the mean value) of $\sum_{i=k}^{\infty}\left(x_i^T Q x_i + u_i^T R u_i\right)$ based on the random networked-induced delays and packet losses at different time intervals.

The stochastic value function (5) can be expressed in terms of the augmented state variable z_k as

$$V_k = \underset{\tau,\gamma}{E}\left[\sum_{i=k}^{\infty}\left(z_i^T Q_z z_i + u_i^T R_z u_i\right)\right] \qquad k = 0,1,2,... \tag{6}$$

where $Q_z = \begin{bmatrix} Q & 0 & \cdots & 0 \\ 0 & \frac{R}{\bar{d}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{R}{\bar{d}} \end{bmatrix}$, and $R_z = \frac{1}{\bar{d}} R$.

Using the input-output form of NNCS (4), the stochastic value function (6) can be represented as

$$\begin{aligned} V_k &= \underset{\tau,\gamma}{E}\left[\sum_{i=k}^{\infty}\left(y_i^{oT}\left(C_o^{-1}\right)^T Q_z C_o^{-1} y_i^o + u_i^T R_z u_i\right)\right] \\ &= \underset{\tau,\gamma}{E}\left[\sum_{i=k}^{\infty}\left(y_i^{oT} Q_y y_i^o + u_i^T R_y u_i\right)\right] \qquad k = 0,1,2,... \end{aligned} \tag{7}$$

where $Q_y = (C_o^{-1})^T Q_z C_o^{-1}$, $R_y = R_z$. Note the matrices Q_y and R_y are still symmetric positive semi-definite and symmetric positive definite respectively. Equation (7) can be also expressed as

$$\begin{aligned} V_k &= \underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_k^o + u_k^T R_y u_k\right) + \underset{\tau,\gamma}{E}\left[\sum_{i=k+1}^{\infty}\left(y_i^{oT} Q_y y_i^o + u_i^T R_y u_i\right)\right] \\ &= \underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_i^o + u_k^T R_y u_k\right) + V_{k+1} \end{aligned} \tag{8}$$

Based on the observability condition [20], when $y^o = 0, V = 0$, the stochastic value function V_k serves as a Lyapunov function [16]. According to Bellman principle of optimality [11], the optimal stochastic value function V_k^* satisfies the discrete-time Hamilton-Jacob-Bellman (HJB) equation in the infinite horizon optimization case as

$$V_k^* = \min_{u_k}\left(\underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_k^o + u_k^T R_y u_k\right) + V_{k+1}^*\right) \tag{9}$$

Differentiating (9), the optimal control u_k^* is given by

$$\frac{\partial \underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_k^o + u_k^T R_y u_k\right)}{\partial u_k} + \frac{\partial y_{k+1}^{oT}}{\partial u_k}\frac{\partial V_{k+1}^*}{\partial y_{k+1}^o} = 0 \tag{10}$$

Namely,

$$u^*\left(y_k^o\right) = -\frac{1}{2} R_y^{-1} G^T\left(y_k^o\right)\frac{\partial V_{k+1}^*}{\partial y_{k+1}^o} \tag{11}$$

Substituting (11) into (8), the discrete-time HJB [17] can be represented by using the system inputs and outputs as

$$V_k^* = \underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_k^o + \frac{1}{4}\frac{\partial V_{k+1}^{*T}}{\partial y_{k+1}^o} G\left(y_k^o\right) R^{-1} G^T\left(y_k^o\right)\frac{\partial V_{k+1}^*}{\partial y_{k+1}^o}\right) + V_{k+1}^* \tag{12}$$

where V_k^* is the stochastic value function corresponding to the optimal control input $u^*(y_k^o)$. Note that when $F(\bullet), G(\bullet)$ are linear matrices, discrete-time HJB equation becomes Riccati equation.

III. OPTIMAL REGULATION OF NNCS

In this section, to overcome the drawbacks of HDP-based value iteration algorithm, first a novel online identifier is introduced to relax the partial knowledge of NNCS dynamics. Subsequently, the critic NN is used to estimate stochastic value function. Eventually, novel optimal regulation control of NNCS is proposed by using action NN, identified partial NNCS dynamics and estimated stochastic value function. It is important to note that this work by using online identifier and online critic and action NNs, the optimal control with guaranteed convergence in the mean is obtained without using value and policy iterations in contrast to [17] which offered an offline solution with no proof of convergence.

A. Online NN-Based Indentification of $G(y_k^o)$.

In this part, a novel online NN-based identifier is proposed to generate $G(y_k^o)$. According to [23], NNCS (4) can be expressed by using following approximation representation on a compact set Ω as

$$\begin{aligned} y_k^o &= F(y_{k-1}^o) + G(y_{k-1}^o)u_{k-1} \\ &= W_F^T \theta_F(y_{k-1}^o) + \varepsilon_{Fk-1} + (W_G^T \theta_G(y_{k-1}^o) + \varepsilon_{Gk-1})u_{k-1} \\ &= [W_F^T \; W_G^T][\theta_F^T(y_{k-1}^o) \; \theta_G^T(y_{k-1}^o)]^T [I \; u_{k-1}^T]^T + [\varepsilon_{Fk-1} \; \varepsilon_{Gk-1}][I \; u_{k-1}^T]^T \\ &= W_C^T \psi_C(y_{k-1}^o) U_{k-1} + \bar{\varepsilon}_{Ck-1} \end{aligned} \tag{13}$$

where $W_C = \left[W_F^T \quad W_G^T\right]^T, \psi_C\left(y_{k-1}^o\right) = \left[\theta_F^T\left(y_{k-1}^o\right) \quad \theta_G^T\left(y_{k-1}^o\right)\right]^T, U_{k-1} = [I \ \ u_{k-1}]^T$, $\varepsilon_{Ck-1} = \left[\varepsilon_{Fk-1} \quad \varepsilon_{Gk-1}\right]$, and $\bar{\varepsilon}_{Ck-1} = \varepsilon_{Ck-1} U_{k-1}$, with $\left\|\psi_C\left(y_{k-1}^o\right)\right\| \le \psi_M$ and $\left\|\psi_C\left(y_{k-1}^o\right) U_{k-1}\right\| \le \Psi_M$ are the bounds while the estimation error satisfies $\left\|\bar{\varepsilon}_{Ck-1}\right\| < \varepsilon_{CM}$, $\forall k$. Since the NN activation functions $\theta_F(\bullet), \theta_G(\bullet)$, and $\psi_C(\bullet)$ are known, NNCS dynamics $G\left(y_k^o\right)$ can be identified (i.e. $G\left(y_k^o\right) = W_G^T \theta_G\left(y_k^o\right) + \varepsilon_{Gk}$) when NN-based identifier weights W_C are updated. Hence, in this section, a suitable update law will be proposed to tune the NN weights. Here, in Theorem 1, the inputs are assumed to the bounded for purpose of the identifier stability whereas it is relaxed during controller design and in the proof of Theorem 3.

The output y_k^o can be estimated at time k by using a NN-based identifier as

$$\hat{y}_k^o = \hat{W}_{Ck}^T \psi_C\left(y_{k-1}^o\right) U_{k-1} \tag{14}$$

Using (13) and (14), the identification error is defined as

$$e_{yk} = y_k^o - \hat{y}_k^o = y_k^o - \hat{W}_{Ck}^T \psi_C\left(y_{k-1}^o\right) U_{k-1} \tag{15}$$

The identification error dynamics (15) are expressed as

$$e_{yk+1} = y_{k+1}^o - \hat{y}_{k+1}^o = y_{k+1}^o - \hat{W}_{Ck+1}^T \psi_C\left(y_k^o\right) U_k \tag{16}$$

Based on [16], an auxiliary identification error vector can be written as

$$\Sigma_{yk} = Y_k^o - \hat{W}_{Ck}^T \Delta\psi_{Ck-1} \overline{U}_{k-1} \tag{17}$$

where $Y_k^o = [y_k^o \ y_{k-1}^o \cdots y_{k-l}^o], \Delta\psi_{Ck-1} = [\psi_C\left(y_{k-1}^o\right) \psi_C\left(y_{k-2}^o\right) \cdots \psi_C\left(y_{k-1-l}^o\right)]$ and $\overline{U}_{k-1} = [U_{k-1} \ U_{k-2} \cdots U_{k-1-l}]$, $0 < l < k-1$. Obviously equation (17) represents l previous identification errors which are recalculated by using most recent NN-based weights $\hat{W}_{Ck}$.

Similar to (17), the auxiliary identification error dynamics are revealed to be

$$\Sigma_{yk+1} = Y_{k+1}^{o} - \hat{W}_{Ck+1}^{T} \Delta\psi_{Ck} \overline{U}_k \tag{18}$$

It is desired to tune the NN identifier weights $\hat{W}_{Ck}$ such that the identification error e_{yk} converges to zero asymptotically, i.e. $k \to \infty, e_{yk} \to 0$. Hence, the update law for NN weights can be defined as

$$\hat{W}_{Ck+1} = \overline{U}_k \Delta\psi_{Ck} \left(\Delta\psi_{Ck}^{T} \overline{U}_k^{T} \overline{U}_k \Delta\psi_{Ck} \right)^{-1} \left(Y_{k+1}^{o} - \alpha_C \Sigma_{yk} \right)^{T} \tag{19}$$

where α_C is the tuning parameter of the NN-based identifier satisfying $0 < \alpha_C < 1$. Substituting (19) into (18)

$$\Sigma_{yk+1} = \alpha_C \Sigma_{yk} \tag{20}$$

Remark 1: We can define $\beta_k = \Psi_C\left(y_k^0\right) U_k$, and β_k has to be persistently exiting [15] long enough for the online NN-based identifier to learn the NNCS dynamics $G\left(y_k^o\right)$.

Next, NN-based identifier weight estimation error is defined as $\tilde{W}_{Ck} = W_C - \hat{W}_{Ck}$, and recalling (13), the identification error dynamics can be rewritten as

$$\begin{aligned} e_{yk+1} = y_{k+1}^{o} - \hat{y}_{k+1}^{o} &= W_C^T \psi_C\left(y_k^o\right) U_k + \bar{\varepsilon}_{Ck} - \hat{W}_{Ck+1}^{T} \psi_C\left(y_k^o\right) U_k \\ &= \tilde{W}_{Ck+1}^{T} \psi_C\left(y_k^o\right) U_k + \bar{\varepsilon}_{Ck} \end{aligned} \tag{21}$$

Using $e_{yk+1} = \alpha_C e_{yk}$ from (20), we have

$$\tilde{W}_{Ck+1}^{T} \psi_C\left(y_k^o\right) U_k = \alpha_C \left(\tilde{W}_{Ck}^{T} \psi_C\left(y_{k-1}^o\right) U_{k-1} \right) + \alpha_C \bar{\varepsilon}_{Ck-1} - \bar{\varepsilon}_{Ck} \tag{22}$$

Eventually, the boundedness of the identification error dynamics e_{yk} given by (15) and NN weights estimation error dynamics $\tilde{W}_{Ck}$ given by (22) will be demonstrated. In order to proceed, the following definition is needed.

Definition 1 [15]: An equilibrium point x_e is said to be uniformly ultimately bounded (*UUB*) in the mean if there exists a compact set $S \subset \Re^n$ so that for all $x_0 \in S$ there exists a bound μ and a number $N(\mu, x_0)$ such that $\|x_k - x_e\| \le \mu$ for all $k \ge k_0 + N$.

Theorem 1 *(Boundedness of the identifier)*. Let the proposed NN-based identifier be defined as (14) and NN weights update law be given by (19). Under the assumption that β_k defined in Remark 1 satisfies PE condition, there exists a positive constant α_C satisfying $0 < \alpha_C < \min\{1, \Psi_{\min} / \sqrt{2}\Psi_M\}$ and computable positive constants B_{WC}, B_{ey}, such that the identification error (15) and NN weights estimation errors $\tilde{W}_{Ck}$ (21) are all uniformly ultimately bounded (*UUB*) in the mean with ultimate bounds given by $\|e_{yk}\| \le B_{ey}$ and $\|\tilde{W}_{Ck}\| \le B_{WC}$

Proof: Refer to the Appendix.

Next the optimal regulation control of NNCS is introduced.

B. Approximation of the Optimal Stochastic Value Function and Control Policy Using NN

In [15], by using universal approximation property of NN, the stochastic value function (7) and control policy (11) can be represented with critic and action NN as

$$V\left(y_k^o\right) = W_V^T \vartheta\left(y_k^o\right) + \varepsilon_{Vk} \tag{23}$$

and

$$u^*\left(y_k^o\right) = W_u^T \phi\left(y_k^o\right) + \varepsilon_{uk} \tag{24}$$

where W_V and W_u represent the constant target NN weights, $\varepsilon_{Vk}, \varepsilon_{uk}$ are the reconstruction errors for critic and action NN respectively, and $\vartheta(\bullet)$ and $\phi(\bullet)$ are the vector activation functions for two NNs, respectively. The upper bounds for the two target NN weights are defined as $\|W_V\| \le W_{VM}$ and $\|W_u\| \le W_{uM}$ where W_{VM}, W_{uM} are positive constants [15], and the approximation errors are also considered bounded as $\|\varepsilon_{Vk}\| \le \varepsilon_{VM}$ and $\|\varepsilon_{uk}\| \le \varepsilon_{uM}$ where $\varepsilon_{VM}, \varepsilon_{uM}$ are also positive constants [15] respectively. Additionally, the gradient of approximation error is assumed to be bounded as $\|\partial \varepsilon_{Vk} / \partial y_{k+1}^o\| \le \varepsilon_{VM}^{'}$ with $\varepsilon_{VM}^{'}$ being a positive constant [11,15].

The critic and action NN approximation of (23) and (24) can be expressed as [11,15]

$$\hat{V}\left(y_k^o\right) = \hat{W}_{Vk}^T \vartheta\left(y_k^o\right) \tag{25}$$

and

$$\hat{u}\left(y_k^o\right) = \hat{W}_{uk}^T \phi\left(y_k^o\right) \tag{26}$$

where $\hat{W}_{Vk}$ and $\hat{W}_{uk}$ represent the estimated values of the target weights W_V and W_u, respectively. In this work, the activation functions $\vartheta(\bullet), \phi(\bullet)$ are selected to be a basis function set and linearly independent [15]. Since it is required that $V\left(y_k^o = 0\right) = 0$ and $u\left(y_k^o = 0\right) = 0$, the basis functions $\vartheta(\bullet), \phi(\bullet)$ are chosen such that $\vartheta\left(y_k^o = 0\right) = 0, \phi\left(y_k^o = 0\right) = 0$, respectively.

Substituting (26) into equation (8), it can be rewritten as

$$W_V^T\left(\vartheta\left(y_{k+1}^o\right) - \vartheta\left(y_k^o\right)\right) + \underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_k + u_k^T R_y u_k\right) = \varepsilon_{Vk} - \varepsilon_{Vk+1}$$

In other words,

$$W_V^T \Delta\vartheta\left(y_{k+1}^o\right) + r\left(y_k^o, u_k\right) = \Delta\varepsilon_{Vk} \tag{27}$$

where $r(y_k^o, u_k) = \underset{\tau,\gamma}{E}(y_k^{oT} Q_y y_k^o + u_k^T R_y u_k)$, $\Delta\vartheta(y_k^o) = \vartheta(y_{k+1}^o) - \vartheta(y_k^o)$ and $\Delta\varepsilon_{vk} = \varepsilon_{vk} - \varepsilon_{vk+1}$. However, when implementing the estimated value function (25), equation (27) does not hold. Therefore, using delayed values for convenience, the residual error or cost-to-go error with (27) can be expressed as

$$\begin{aligned} e_{Vk} &= \underset{\tau,\gamma}{E}\left(y_k^{oT} Q_y y_k^o + u_k^T R_y u_k\right) + \hat{V}\left(y_{k+1}^o\right) - \hat{V}\left(y_k^o\right) \\ &= r\left(y_k^o, u_k\right) + \hat{W}_{Vk}^T \Delta\vartheta\left(y_k^o\right) \end{aligned} \tag{28}$$

Based on gradient descent algorithm, the update law of critic NN weights is given by

$$\begin{aligned} \hat{W}_{Vk+1} &= \hat{W}_{Vk} - \alpha_v \frac{\Delta\vartheta\left(y_k^o\right)}{\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1} e_{Vk}^T \\ &= \hat{W}_{Vk} - \alpha_v \frac{\Delta\vartheta\left(y_k^o\right)}{\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1} [r^T\left(y_k^o, u_k\right) + \Delta\vartheta^T\left(y_k^o\right)\hat{W}_{Vk}] \end{aligned} \tag{29}$$

Remark 2: It is important to note that the stochastic value function (8) and critic NN approximation (25) all become zero only when $y_k^o = 0$. Therefore, once the system outputs have converged to zero, the value function approximation is no longer updated. This can be also viewed as a PE requirement for the inputs to the critic NN where the system outputs must be persistently exiting long enough for the approximation so that critic NN learns the optimal stochastic value function. In this paper, the PE condition is met by introducing noise.

As a final step in the critic NN design, define the weight estimation error as $\tilde{W}_{Vk} = W_V - \hat{W}_{Vk}$. Since $r^T\left(y_k^o, u_k\right) = -\Delta\vartheta^T\left(y_{k+1}^o\right)W_V + \Delta\varepsilon_{Vk}^T$ in equation (27), the dynamics of the critic NN weights estimation error can be rewritten as

$$\tilde{W}_{Vk+1} = \tilde{W}_{Vk} - \alpha_V \frac{\Delta \vartheta\left(y_k^o\right)\Delta \vartheta^T\left(y_k^o\right)}{\Delta \vartheta^T\left(y_k^o\right)\Delta \vartheta\left(y_k^o\right)+1}\tilde{W}_{Vk} + \alpha_V \frac{\Delta \vartheta\left(y_k^o\right)\Delta \varepsilon_{Vk}}{\Delta \vartheta^T\left(y_k^o\right)\Delta \vartheta\left(y_k^o\right)+1} \tag{30}$$

Next, the boundedness of the critic NN estimation error dynamics $\tilde{W}_{Vk}$ given by (30) is demonstrated in the following theorem.

Theorem 2: *(Boundness of the Critic NN estimation errors).* Let $u_0\left(y_k^o\right)$ be any admissible control policy for nonlinear NCS (4), and let the critic NN weights update law be given by (29). Then there exists positive constant α_V satisfying $0 < \alpha_V < 1/2$ and computable positive constant B_{Wv}, such that the critic NN weights estimation error (30) is *UUB* in the mean with ultimate bounds given by $\left\|\tilde{W}_{Vk}\right\| \le B_{Wv}$.

Proof: Refer to the Appendix.

Now we need to find the control policy via action NN (26) which minimizes the approximated value function (25). First, the action NN estimation errors are defined to be the difference between the actual optimal control input (26) that is being applied to NNCS (4) and the control input that minimizes the estimated value function (25) with identified NNCS dynamics $\hat{G}\left(y_k^o\right)$. This error can be expressed as

$$e_{uk} = \hat{W}_{uk}^T \phi\left(y_k^o\right) + \frac{1}{2} R_y^{-1} \hat{G}^T\left(y_k^o\right) \frac{\partial \vartheta^T\left(y_{k+1}^o\right)}{\partial y_{k+1}^o} \hat{W}_{Vk} \tag{31}$$

The update law for action NN weights is defined as

$$\hat{W}_{uk+1} = \hat{W}_{uk} - \alpha_u \frac{\phi\left(y_k^o\right)}{\phi^T\left(y_k^o\right)\phi\left(y_k^o\right)+1} e_{uk}^T \tag{32}$$

where $0 < \alpha_u < 1$ is a positive constant. By selecting the control policy u_k to minimize the desired value function (23), it follows that

$$W_u^T\phi\left(y_k^o\right)+\varepsilon_{uk}=-\frac{1}{2}R_y^{-1}\hat{G}^T\left(y_k^o\right)\left(\frac{\partial\vartheta^T\left(y_{k+1}^o\right)}{\partial y_{k+1}^o}W_V+\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}\right)$$

In other words,

$$W_u^T\phi\left(y_k^o\right)+\varepsilon_{uk}+\frac{1}{2}R_y^{-1}\hat{G}^T\left(y_k^o\right)\left(\frac{\partial\vartheta^T\left(y_{k+1}^o\right)}{\partial y_{k+1}^o}W_V+\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}\right)=0 \tag{33}$$

Substituting (33) into (31), the action NN estimation error dynamics can be rewritten as

$$e_{uk}=-\widetilde{W}_{uk}^T\phi\left(y_k^o\right)-\frac{1}{2}R_y^{-1}\hat{G}^T\left(y_k^o\right)\frac{\partial\vartheta^T\left(y_{k+1}^o\right)}{\partial y_{k+1}^o}\widetilde{W}_{Vk}+\frac{1}{2}R_y^{-1}\widetilde{G}^T\left(y_k^o\right)\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}-\varepsilon_{ek} \tag{34}$$

where $\widetilde{G}\left(y_k^o\right)=G\left(y_k^o\right)-\hat{G}\left(y_k^o\right)$, $\varepsilon_{ek}=\varepsilon_{uk}+\frac{1}{2}R_y^{-1}G^T\left(y_k^o\right)\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}$ satisfying $\left\|\varepsilon_{ek}\right\|\le\varepsilon_{eM}$ with ε_{eM} being a positive constant, and $\left\|\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}\right\|\le\varepsilon_{VM}^{'}$.

The action NN weight estimation error dynamics can be represented as

$$\widetilde{W}_{uk+1}=\widetilde{W}_{uk}+\alpha_u\frac{\phi\left(y_k^o\right)}{\phi^T\left(y_k^o\right)\phi\left(y_k^o\right)+1}e_{uk}^T \tag{35}$$

$$=\widetilde{W}_{uk}-\alpha_u\frac{\phi\left(y_k^o\right)}{\phi^T\left(y_k^o\right)\phi\left(y_k^o\right)+1}[\widetilde{W}_{uk}^T\phi\left(y_k^o\right)+\frac{1}{2}R_y^{-1}\hat{G}^T\left(y_k^o\right)\frac{\partial\vartheta^T\left(y_{k+1}^o\right)}{\partial y_{k+1}^o}\widetilde{W}_{Vk}$$

$$-\frac{1}{2}R_y^{-1}\widetilde{G}^T\left(y_k^o\right)\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}+\varepsilon_{ek}]$$

Remark 3: In this work, the proposed NN-based identifier relaxes the need for partial NNCS dynamics $G\left(y_k^o\right)$. Compared to [11], the knowledge of the input transformation matrix $G(y_k^o)$ and internal dynamics $F(y_k^o)$ are considered unknown here.

Next, the stability of NN-based identification error dynamics, NN identifier weight estimation errors, critic NN estimation and action NN estimation error dynamics are considered.

C. Closed-Loop Stability

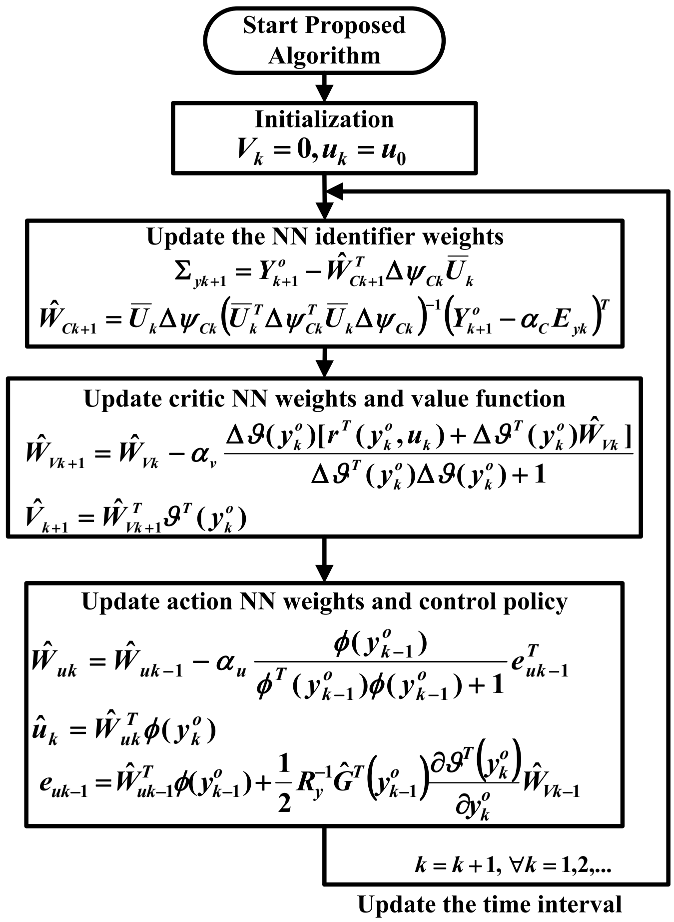


Fig 3. Flowchart of the proposed optimal controller for NNCS

In this section, it will be shown that the closed-loop system is bounded. On the other hand, when the NN approximation errors for the identifier, action and value functions are considered negligible [21], as in the case of standard adaptive control [21], or when the number of hidden-layer neurons is increased significantly, the estimated control policy approaches the optimal control input asymptotically. Before introducing the theorem on system stability, we present the flowchart in Figure 3 of the proposed time-based NDP for NNCS with uncertain system dynamics and unknown network imperfections.

For the closed-loop stability in the mean and convergence proof, the initial system outputs are considered to reside in a compact set $\Omega \in \Re^n$ because of the initial admissible control input $u_0(y_k^o)$. In addition, the critic NN basis function and its gradient as well as the activation function of the action NN are considered bounded with $\|\vartheta(y_k^o)\| \le \vartheta_M$, $\|\partial\vartheta(y_k^o)/\partial y_k^o\| \le \vartheta_M^{'}$, and $\|\phi(y_k^o)\| \le \phi_M$, respectively in Ω. Further, sufficient conditions for the three NN tuning parameters, α_C, α_V and α_u, are derived to guarantee that all future outputs never leave the compact set. In order to proceed, the following lemma is needed.

Lemma 1: There exists admissible control policy be applied to the controllable NNCS (4) such that system dynamics satisfy

$$\|F(y_k^o) + G(y_k^o)u^*(y_k^o)\|^2 \le k^*\|y_k^o\|^2 \tag{36}$$

where $0 < k^* < 1/2$ is a constant.

Theorem 3: *(Convergence of the Optimal Control Signal)*. Let $u_0(y_k^o)$ be any initial stabilizing control policy for the NNCS in (4) which satisfy the bounds in (A.5) and $0 < k^* < 1/2$. Let the NN weight tuning for the identifier, critic and the action NN be provided by (19), (29) and (32), respectively. Then, there exists positive constant $\alpha_C, \alpha_u, \alpha_V$ satisfying $0 < \alpha_C < \min\{1, \frac{\Psi_{\min}}{2\sqrt{2}\Psi_M}\}$, $\frac{1}{4} < \alpha_V < \frac{3+\sqrt{3}}{12}$ and $\frac{1}{6} < \alpha_u < \frac{1}{3}$ respectively, and positive constants b_y, b_V, b_{WC}, b_{ey} and b_u such that the system output vector y_k^o, NN identification error e_{yk}, weight estimation errors $\tilde{W}_{Ck}$, critic and action NN weight estimation errors $\tilde{W}_{Vk}$ and $\tilde{W}_{uk}$, respectively, are all *UUB* in the mean for all $k \geq k_0 + T$ with ultimate bounds given by $\|y_k^o\| \leq b_y, \|e_{yk}\| \leq b_{ey}, \|\tilde{W}_{Ck}\| \leq b_{WC}$, $\|\tilde{W}_{Vk}\| \leq b_V$ and $\|\tilde{W}_{uk}\| \leq b_u$. Further, $\|\hat{u}(y_k^o) - u^*(y_k^o)\| \leq \delta_u$ for a small positive constant δ_u.

Proof: Refer to the Appendix.

Remark 4: It is important to note that Theorems 1 and 2 demonstrated UUB in the mean of NN identifier and Critic NN estimation errors respectively. In Theorem 3, boundness of NN identifier, Critic NN and estimated stochastic optimal control from action NN are all considered simultaneously.

IV. SIMULATION RESULTS

In this section, stochastic optimal control of NNCS with uncertain dynamics in the presence of unknown random delays and packet losses is evaluated. The continuous-time version of original nonlinear affine system is given by

$$\dot{x} = f(x) + g(x)u, \; y = Cx \tag{37}$$

where $f(x) = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2(1 - (\cos(2x_1) + 2)^2) \end{bmatrix}$, $g(x) = \begin{bmatrix} 0 \\ cos(2x_1) + 2 \end{bmatrix}$, and $C = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$

The network parameters of the NNCS are selected as [13,19]:

1. The sampling time: $T_s = 100ms$;

2. The bound of delay is set as two, i.e. $\bar{d} = 2$;

3.The mean random delay values are given by $E(\tau_{sc}) = 80ms \; E(\tau) = 150ms$;

4. Packet losses follow Bernoulli distribution with $p = 0.3$.

The distribution of random delays, includes sensor-to-control delay τ_{sc} with a total delay of τ , is shown in Figure 4 while the packet losses are shown in Figure 5. Incorporating the random delays $\tau(t)$ and packet losses $\gamma(t)$ into the original nonlinear affine system (51), yields the unknown NNCS given by (4).

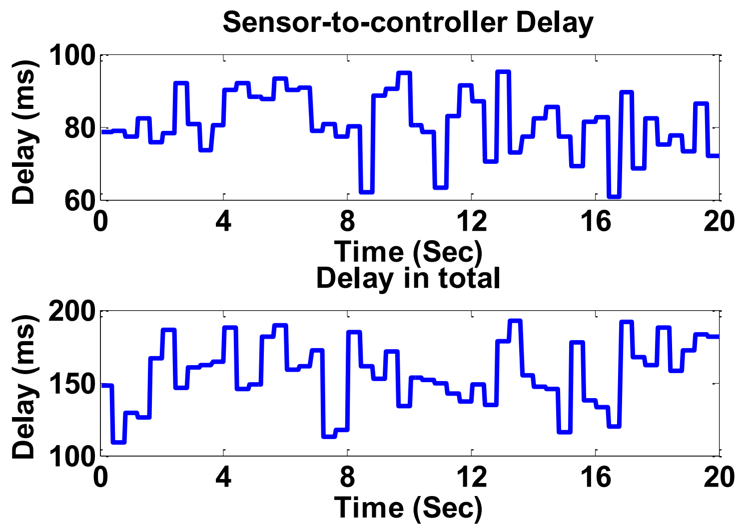


Fig 4. The distribution of random delays in NCS

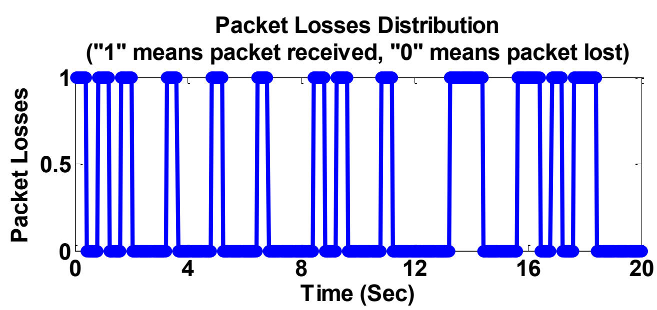


Fig 5. The distribution of packet losses.

First, the effect of random delays and packet losses for NNCS is studied. The initial state is taken as $x_0 = [5 \quad -3]^T$. The initial static control $u_k = [-2 \quad -5]x_k$, which maintains the original nonlinear affine system (51) stable, is shown in Figure 6 (a). By contrast, this controller cannot maintain system stable in the mean for NNCS in presence of random delays and packet losses as Figure 6 (b).

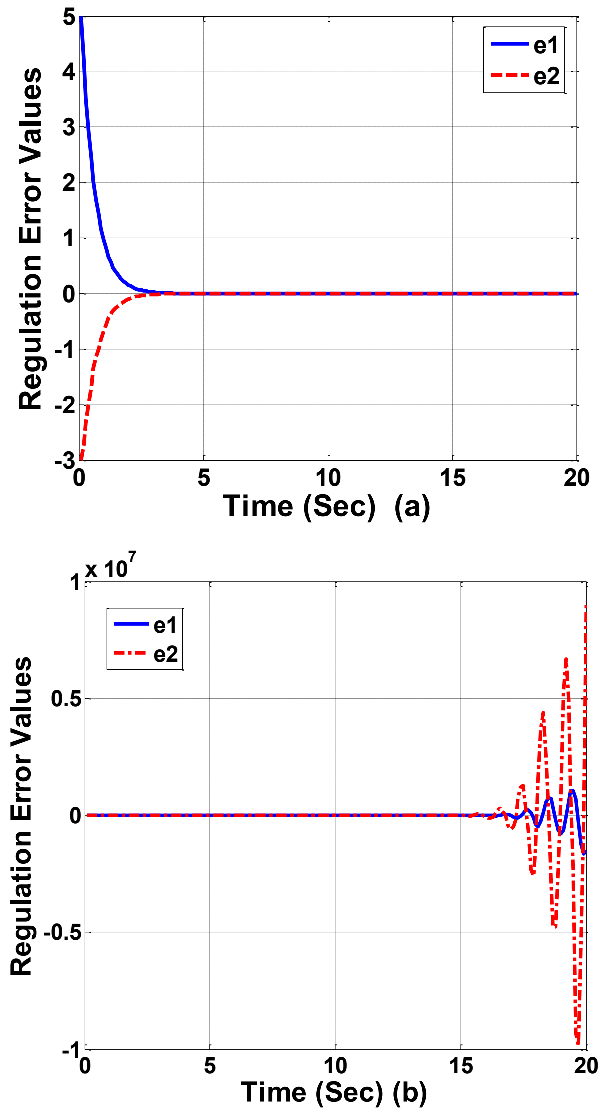


Fig 6. Performance of a static feedback controller: (a) random delays and packet losses are not present; (b) with random delays and packet losses.

Next, the proposed stochastic optimal control is implemented for the NNCS with unknown system dynamics in presence of random delays and packet losses. The augment state y_k^o is generated as $y_k^o = [y_k \quad u_{k-1} \quad u_{k-2}]^T \in \Re^{4\times 1}$, $\forall k$, and the initial stabilizing policy for proposed algorithm was selected as $u_o\left(y_k^o\right)=\left[-2 \quad -5 \quad -1 \quad -1\right]y_k^o$ generated by using standard pole placement method, while the activation functions for NN-based identifier were generated as $\tanh\{\left(y_1^o\right)^2, y_1^o y_2^o, \ldots, \left(y_4^o\right)^2, \left(y_1^o\right)^4, \left(y_1^o\right)^3 y_2^o, \ldots, \left(y_4^o\right)^6\}$, critic NN activation function were selected as sigmoid of sixth order polynomial $\{\left(y_1^o\right)^2, y_1^o y_2^o, \ldots, \left(y_4^o\right)^2, \left(y_1^o\right)^4, \left(y_1^o\right)^3 y_2^o, \ldots, \left(y_4^o\right)^6\}$ and action NN activation function were generated from the gradient of critic NN activation function.

The design parameters for NN-based identifier, critic NN and action NN were selected as $\alpha_C = 0.002, \alpha_V = 10^{-4}$ and $\alpha_u = 0.005$ while the NN-based identifier and critic NN weights are set to zero at the beginning of the simulation. The initial weights of the action NN are chosen to reflect the initial stabilizing control. The simulation was run for 20 seconds (200 time steps), for the first 10 seconds (100 time steps), exploration noise with mean zero and variance 0.06 was added to the system in order to ensure the persistency of excitation (PE) condition (See Remarks 1 and 2).

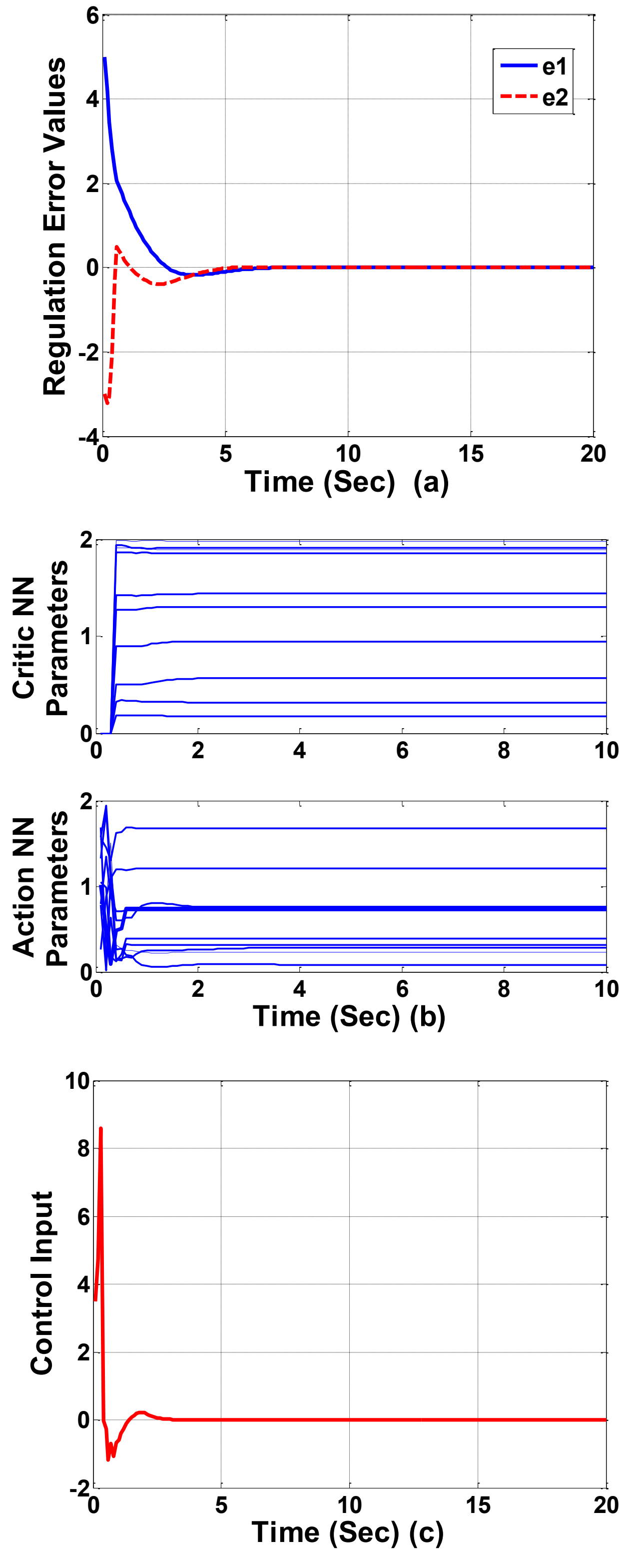


Fig 7. Performance of stochastic optimal controller for NNCS: (a) State regulation errors; (b) Critic NN and Action NN parameters. (c) Control input.

The performance of proposed stochastic optimal controller is evaluated from several aspects: 1) as shown in Figure 7(a), the proposed stochastic optimal controller can make the NNCS state regulation errors converge to zero even when the NNCS dynamics are uncertain which implies that the proposed controller can make the NNCS system stable in the mean; 2) the proposed critic NN and action NN parameters converge to constant values and remain bounded consistent with Theorem 3 as shown in Figure 7(b); 3) The optimal control input for NNCS with uncertain dynamics is shown in Figure 7 (c) which is bounded.

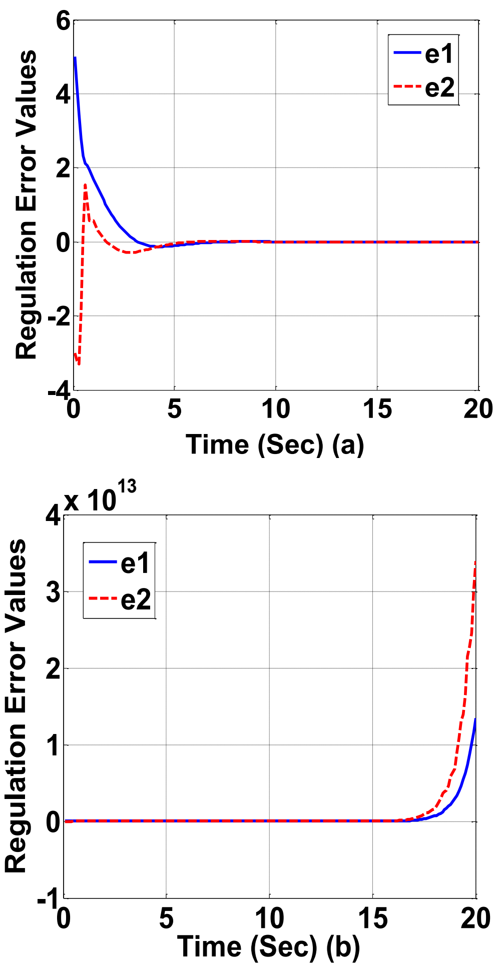


Fig 8. Performance of HDP value iteration for NNCS: (a) Iterations=100 times/sample; (b) Iterations=10 times/ sample.

For comparison, HDP value iteration (VI) [9,17] is also implemented for NNCS with known dynamics $G(\bullet)$ by incorporating the $g(\bullet)$(37) and information of delays and packet losses which are normally not known before hand. The initial admissible control, critic NN and action NN activation function are same as proposed time-based stochastic optimal control.

As shown in Figure 8 (a), the HDP VI method can make the NNCS state regulation errors converge to zero when the number of iterations is 100 times/sample. By contrast, HDP VI cannot maintain NNCS stable in the mean when iterations become 10 times/sample as shown in Figure 8 (b). It implies that HDP VI scheme not only needs partial knowledge of original nonlinear affine system dynamics, $g(\bullet)$, but also information on delays and packet losses. The number of iterations required for a given nonlinear system is unknown. Due to these drawbacks, the HDP VI is not preferred for NNCS implementation in real-time.

Based on the results presented in Figures 4 through 8, the proposed stochastic optimal control scheme with uncertain NNCS dynamics and unknown network imperfections can overcome the drawbacks of HDP-based value iteration method and will render nearly the same performance as that of an optimal controller for NNCS when the system dynamics, random delays and packet losses are known.

V. CONCLUSIONS

In this work, an online approximate dynamic programming technique for NNCS is proposed by using identifier NN, critic NN and action NN to solve the stochastic optimal regulation of NNCS with uncertain dynamics in presence of random delays and packet losses. Compared with other recent NNCS and NDP research works, this paper has developed a NNCS representation with augment states. The NN identifier relaxed the requirement of input gain matrix for NNCS while the information on random delays and packet losses are not needed. Consequently, proposed time-based, forward-in-time scheme can be implemented in practical NNCS. Therefore no value and policy iterations are required since a history of cost to go errors are utilized.

The initial admissible control policy ensured that NNCS is stable in the mean while NN identifier learns the input gain matrix, the critic NN approximates the stochastic value function $V\left(y_k^o\right)$, and the action NN generates the approximate stochastic optimal control. All NN weights were tuned online using proposed update laws and Lyapunov theory demonstrates the asymptotic convergence of the approximated control input to its optimal value over time in the mean.

REFERENCES

[1] J. Nilsson, “Real-time control systems with delays,” Ph.D. dissertation, Dept. Automatic Control, Lund Institute of Technology, Lund, Sweden, 1998.

[2] G. C. Walsh, O. Beldiman, and L. Bushnell. “Error encoding algorithm for networked control system”. Automatica, vol. 38, pp. 261–267, 2002.

[3] I. G. Polushin, P. X. Liu, and C. H. Lung. On the model-based approach to nonlinear networked control systems. Automatica, vol. 44, pp. 2409–2414, 2008.

[4] J. B. Rawlings, E. S. Meadows, and K. Muske, “Nonlinear model predictive control: a tutorial and survey.” in Proc. IFAC ADCHEM, pp. 203-214, 1994.

[5] G. P. Liu, Y. Xia, D. Rees, and W. S. Hu, “Design and stability criteria of networked predictive control systems with random network delay in the feedback channel,” IEEE Trans. Syst., Man Cybern., vol. 37, pp. 173–184, 2007.

[6] P. J. Werbos, “A menu of designs for reinforcement learning over time”. J. Neural Networks Contr. pp. 67–95. MA: MIT Press, 1991.

[7] D. S. Bernstein, and W. M. Haddad, “LQG control with an H_∞ performance bound: a Riccati equation approach,” IEEE Trans. Automat. Contr., vol. 34, pp. 293-305, 1989.

[8] D. P. Bertsekas and J. Tsitsiklis, Neuro-dynamics programming, Athena Scientic, 1996.

[9] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control", IEEE Circ. and Syst. Magaz., vol. 9, pp. 32-50, 2009.

[10] C. Zheng and S. Jagannathan, "Generalized Hamilton–Jacobi–Bellman formulation-based neural network control of affine nonlinear discrete-time systems," IEEE Trans. Neural Networks, vol. 19, no. 1, pp. 90–106, 2008.

[11] T. Dierks, and S. Jagannathan, "Optimal control of affine nonlinear discrete-time systems with unknown internal dynamics", in Proc. of the IEEE Conf. on Decision and Contr., pp. 6750-6755, 2009.

[12] M. G. Safonov and T. C. Tsao, "The unfalsified control concept and learning," IEEE Trans. Autom. Control, vol. 42, pp. 843–847, 1997.

[13] L. W. Liou, and A. Ray, "A stochastic regulator for integrated communication and control systems: part I—formulation of control law". ASME J. Dynamic Syst., Measure. Contr., vol. 4, pp. 604–611, 1991.

[14] X. Liu and A. Goldsmith, "Wireless medium access control in networked control systems", in Proc. IEEE Amer. Contr. Conf., pp. 688-694, 2004.

[15] S. Jagannathan, Neural network control of nonlinear discrete-time systems, CRC Press, 2006.

[16] M. Abu-Khalaf, F. L. Lewis, and J. Huang, "Neurodynamic programming and zero-sum games for constrained control systems," IEEE Trans. Neural Networks, vol. 19, pp. 1243-1252, 2008.

[17] A. Al-Tamimi and F. L. Lewis, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof," IEEE Trans. Syst., Man., Cybern. B, vol. 38, pp. 943–949, 2008.

[18] B. A. Finlayson, The Method of Weighted Residuals and VariationalPrinciples. New York: Academic, 1972.

[19] H. Shousong and Z. Qixin, "Stochastic optimal control and analysis of stability of networked control systems with long delay," Automatica, vol. 39, pp. 1877–1884, 2003.

[20] H.F. Chen and L. Guo, Identification and Stochastic Adaptive Control. Cambridge, MA, 1991.

[21] W. Stallings. Wireless Communications and Networks. Prentice Hall, 1st edition, 2002.

[22] D. V. Prokhorov, and D. C. Wunsch II, "Adaptive critic designs," IEEE Trans. Neur. Netw., vol. 8, pp. 997-1007, 1997.

[23] J. Dankert, L. Yang, and J. Si, "A performance gradient perspective on approximate dynamic programming and its application to partially observable markov decision process," in Proc. Int. Symp. Intell. Contr., pp. 458-463, 2006.

[24] R. Enns, and J. Si, "Helicopter trimming and tracking control using direct neural dynamic programming," IEEE Trans. Neur. Netw., vol. 14, pp. 929-939, 2003.

[25] H. Zhang, Y. Luo, and D. Liu, "Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints," IEEE Trans. Neur. Netw., vol. 20, pp. 1490-1503, 2009.

[26] H. Zhang, Q. Wei, and Y. Luo, "A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP

iteration algorithm," IEEE Trans. Syst., Man, Cybernet., B, vol. 28, pp. 937-942, 2008.

[27] D. P. Bertsekas, Dynamic programming and optimal control. Athena Scientific, 3rd edition, 2007.

APPENDIX

Proof of Theorem 1: Consider Lyapunov function candidate

$$L_C = tr\left\{e_{yk}^T e_{yk}\right\} + \Psi_{\min}^2 tr\left\{\widetilde{W}_{Ck}^T \widetilde{W}_{Ck}\right\} \tag{A.1}$$

where $0 < \Psi_{\min}^2 \le \left\|\left(\psi_C\left(y_{k-1}^o\right)U_{k-1}\right)^T\right\|^2$ is ensured by the PE condition described in Remark 1 [19,23]. The first difference of (A.1) is given by

$$\Delta L_C = tr\left\{e_{yk+1}^T e_{yk+1}\right\} - tr\left\{e_{yk}^T e_{yk}\right\} + \Psi_{\min}^2 [tr\left\{\widetilde{W}_{Ck+1}^T \widetilde{W}_{Ck+1}\right\} - tr\left\{\widetilde{W}_{Ck}^T \widetilde{W}_{Ck}\right\}]$$

$$\le tr\left\{e_{yk+1}^T e_{yk+1}\right\} - tr\left\{e_{yk}^T e_{yk}\right\} + tr\left\{\left(\widetilde{W}_{Ck+1}^T \psi_C\left(y_k^o\right)U_k\right)^T \left(\widetilde{W}_{Ck+1}^T \psi_C\left(y_k^o\right)U_k\right)\right\} - \Psi_{\min}^2 tr\left\{\widetilde{W}_{Ck}^T \widetilde{W}_{Ck}\right\}$$

and using (15) and (22) yields

$$\begin{aligned}
\Delta L_C &\le \alpha_C^2\left\|e_{yk}\right\|^2 - \left\|e_{yk}\right\|^2 + \left\|\alpha_C\left(\widetilde{W}_{Ck}^T \psi_C\left(y_{k-1}^o\right)U_{k-1}\right) + \alpha_C \bar{\varepsilon}_{Ck-1} - \bar{\varepsilon}_{Ck}\right\|^2 - \Psi_{\min}^2\left\|\widetilde{W}_{Ck}\right\|^2 \\
&\le -\left(1-\alpha_C^2\right)\left\|e_{yk}\right\|^2 + 2\alpha_C^2\left\|\widetilde{W}_{Ck}^T \psi_C\left(y_{k-1}^o\right)U_{k-1}\right\|^2 + 2\left\|\alpha_C \bar{\varepsilon}_{Ck-1} - \bar{\varepsilon}_{Ck}\right\|^2 - \Psi_{\min}^2\left\|\widetilde{W}_{Ck}\right\|^2 \\
&\le -\left(1-\alpha_C^2\right)\left\|e_{yk}\right\|^2 - \left(\Psi_{\min}^2 - 2\alpha_C^2\Psi_M^2\right)\left\|\widetilde{W}_{Ck}^T\right\|^2 + 2\left\|\alpha_C \bar{\varepsilon}_{Ck-1} - \bar{\varepsilon}_{Ck}\right\|^2 \qquad \text{(A.2)} \\
&\le -\left(1-\alpha_C^2\right)\left\|e_{yk}\right\|^2 - \left(\Psi_{\min}^2 - 2\alpha_C^2\Psi_M^2\right)\left\|\widetilde{W}_{Ck}\right\|^2 + 2\Delta\bar{\varepsilon}_{CM}^2
\end{aligned}$$

where $\left\|\alpha_C \bar{\varepsilon}_{Ck-1} - \bar{\varepsilon}_{Ck}\right\|^2 \le \Delta\bar{\varepsilon}_{CM}^2$, with $\Delta\bar{\varepsilon}_{CM}^2$ is a constant which is guaranteed by the boundedness of $\bar{\varepsilon}_{Ck}$ in Section III-A. When tuning parameter α_C satisfies $0 < \alpha_C < \min\{1, \dfrac{\Psi_{\min}}{\sqrt{2}\Psi_M}\}$ the first difference of (A.1) is less than zero provided the following inequalities hold

$$\left\|\widetilde{W}_{Ck}\right\| > \frac{\Delta\bar{\varepsilon}_{CM}^2}{\left(\Psi_{\min}^2 - 2\alpha_C^2\Psi_M^2\right)} \equiv B_{WC} \quad \text{or} \quad \left\|e_{yk}\right\| > \frac{\Delta\bar{\varepsilon}_{CM}^2}{\left(1-\alpha_C^2\right)} \equiv B_{ey} \tag{A.3}$$

Using the standard Lyapunov extension [23], the identification errors and NN weights estimation errors are *UUB* in the mean.

Proof of Theorem 2: Consider the Lyapunov function candidate

$$L_V\left(\widetilde{W}_{Vk}\right) = tr\left\{\widetilde{W}_{Vk}^T \widetilde{W}_{Vk}\right\} \tag{A.4}$$

The first difference of (A.4) is given by $\Delta L_V\left(\widetilde{W}_{Vk}\right) = tr\left\{\widetilde{W}_{Vk+1}^T \widetilde{W}_{Vk+1}\right\} - tr\left\{\widetilde{W}_{Vk}^T \widetilde{W}_{Vk}\right\}$, and using (30) yields

$$\Delta L_V\left(\widetilde{W}_{Vk}\right) = tr\left\{\widetilde{W}_{Vk}^T \widetilde{W}_{Vk}\right\} - \frac{2\alpha_V}{\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1} tr\left\{\widetilde{W}_{Vk}^T \Delta\vartheta\left(y_k^o\right)\Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\}$$

$$+\frac{2\alpha_V}{\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1} tr\left\{\Delta\varepsilon_{Vk}^T \Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\}$$

$$+\frac{\alpha_V^2 tr\left\{\widetilde{W}_{Vk}^T \Delta\vartheta\left(y_k^o\right)\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)\Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\}}{\left(\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1\right)^2}$$

$$+\frac{2\alpha_V^2 tr\left\{\Delta\varepsilon_{Vk}^T \Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)\Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\}}{\left(\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1\right)^2}$$

$$+\frac{\alpha_V^2 tr\left\{\Delta\varepsilon_{Vk}^T \Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)\Delta\varepsilon_{Vk}\right\}}{\left(\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1\right)^2} - tr\left\{\widetilde{W}_{Vk}^T \widetilde{W}_{Vk}\right\}$$

$$\leq -\frac{\alpha_V}{\left(\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1\right)}[(2-a_V)\left\|\Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\|^2$$

$$-(2+2a_V)\left\|\Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\|\left\|\Delta\varepsilon_{Vk}\right\| - a_V\left\|\Delta\varepsilon_{Vk}\right\|^2]$$

$$\leq -\frac{\alpha_V}{\left(\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1\right)}[(1-2a_V)\left\|\Delta\vartheta^T\left(y_k^o\right)\widetilde{W}_{Vk}\right\|^2 - (2a_V+1)\left\|\Delta\varepsilon_{Vk}\right\|^2]$$

$$\leq -\frac{\alpha_V}{(2-a_V)\left(\Delta\vartheta^T\left(y_k^o\right)\Delta\vartheta\left(y_k^o\right)+1\right)}[(1-2a_V)\Delta\vartheta_{\min}^2\left\|\widetilde{W}_{Vk}\right\|^2 - (2a_V+1)\Delta\varepsilon_{VM}^2]$$

$$\leq -\frac{\alpha_V(1-2a_V)\Delta\vartheta_{\min}^2\left\|\widetilde{W}_{Vk}\right\|^2}{(2-a_V)(\Delta\vartheta_M^2+1)}+\frac{\alpha_V(2a_V+1)\Delta\varepsilon_{VM}^2}{(2-a_V)(\Delta\vartheta_M^2+1)} \tag{A.5}$$

where $0<\Delta\vartheta_{\min}<\left\|\Delta\vartheta(y_k^o)\right\|$ is ensured by the PE condition described in Remark 2 and $\left\|\Delta\varepsilon_{Vk}\right\|\leq\Delta\varepsilon_{VM}$ for a constant $\Delta\varepsilon_{VM}$ is ensured by the boundness of ε_{Vk} . Therefore, $\Delta L_V\left(\widetilde{W}_{Vk}\right)<0$ if

$$\left\|\widetilde{W}_{Vk}\right\|>\sqrt{\frac{2a_V+1}{(1-2a_V)\Delta\vartheta_{\min}^2}\Delta\varepsilon_{VM}^2}\equiv B_{Wv} \tag{A.6}$$

Using standard Lyapunov theory [15], it can be concluded that $\Delta L_V\left(\widetilde{W}_{Vk}\right)$ is less than zero outside of a compact set rendering the critic NN weights estimation errors to be *UUB* in the mean.

Proof of Theorem 3: Consider the Lyapunov function candidate

$$L=L_{DN}+L_{uN}+L_{VN}+L_{CN}+L_{AN}+L_{BN} \tag{A.7}$$

where $L_{DN}=\left(y_k^o\right)^T y_k^o$, $L_{uN},L_{VN},L_{CN},L_{AN}$ and L_{BN} are defined as

$$L_{uN}=tr\left\{\widetilde{W}_{uk}^T\Omega\widetilde{W}_{uk}\right\};\ L_{VN}=tr\left\{\widetilde{W}_{Vk}^T\Lambda\widetilde{W}_{Vk}\right\} \tag{A.8}$$

$$L_{CN}=tr\left\{e_{yk}^T e_{yk}\right\}+tr\left\{\widetilde{W}_{Ck}^T O\widetilde{W}_{Ck}\right\}$$

$$L_{AN}=\left(tr\left\{\widetilde{W}_{Vk}^T\Gamma\widetilde{W}_{Vk}\right\}\right)^2;\ L_{BN}=\left(tr\left\{\widetilde{W}_{Ck}^T\Theta\widetilde{W}_{Ck}\right\}\right)^2$$

with $\Omega=\frac{24G_M^2\phi_M^2\left(\phi_M^2+1\right)}{\phi_{\min}^2}\mathbf{I}$, $\Lambda=\frac{288\phi_M^2\Xi^2\left(\Delta\vartheta_M^2+1\right)}{\phi_{\min}^2\Delta\vartheta_{\min}^2}\times\left(G_M^2+12\frac{\psi_M^2\Delta\varepsilon_{VM}^2}{\Delta\vartheta_{\min}^2}\right)\mathbf{I}$,

$O=2[\Psi_M^2+\frac{9\left(\varepsilon_{VM}^{'}\psi_M\right)^2\phi_M^2\Xi^2}{2\phi_{\min}^2}+6\left(\Xi\psi_M\right)^2\phi_M^2\quad\frac{\phi_M^2\Delta\bar{\varepsilon}_{CM}^2}{\phi_{\min}^2\Psi_{\min}^2}]\mathbf{I}$, $\Gamma=\left(85\sqrt{\frac{\left(\Xi\psi_M\phi_M\right)^2\left(\Delta\vartheta_M^2+1\right)^2}{\phi_{\min}^2\Delta\vartheta_{\min}^4}}\right)\mathbf{I}$

and $\Theta = \left(\sqrt{\frac{24(\Xi \psi_M \phi_M)^2}{\phi_{\min}^2}} \right)\mathbf{I}$ are positive definite matrices, $\mathbf{I}$ is identity matrix, Ξ is defined as $\lambda_{\max}(R^{-1})\vartheta_M' G_M$, and $\lambda_{\max}(R^{-1})$ is the maximum singular value of R. The first difference of (A.7) is given by $\Delta L = \Delta L_{DN} + \Delta L_{uN} + \Delta L_{VN} + \Delta L_{CN} + \Delta L_{AN} + \Delta L_{BN}$.

Considering first difference $\Delta L_{DN} = (y_{k+1}^o)^T y_{k+1}^o - (y_k^o)^T y_k^o$, using the NNCS dynamics (4), and applying the Cauchy-

Schwartz inequality reveals that the first difference becomes

$$
\begin{aligned}
\Delta L_{DN} &\le \left\| \begin{matrix} F(y_k^o) + G(y_k^o)u^*(y_k^o) - G(y_k^o)u^*(y_k^o) \\ + G(y_k^o)\hat{u}(y_k^o) - G(y_k^o)\hat{u}(y_k^o) + \hat{G}(y_k^o)\hat{u}(y_k^o) \end{matrix} \right\|^2 - (y_k^o)^T y_k^o \\
&\le \left\| \begin{matrix} F(y_k^o) + G(y_k^o)u^*(y_k^o) - \tilde{G}(y_k^o)\hat{u}(y_k^o) \\ - G(y_k^o)\tilde{W}_{uk}^T\varphi(y_k^o) - G(y_k^o)\varepsilon_{uk} \end{matrix} \right\|^2 - (y_k^o)^T y_k^o \\
&\le 2\left\| F(y_k^o) + G(y_k^o)u^*(y_k^o) \right\|^2 + 4\left\| \tilde{G}(y_k^o)\hat{u}(y_k^o) \right\|^2 \\
&+ 8\left\| G(y_k^o)\tilde{W}_{uk}^T\phi(y_k^o) \right\|^2 + 8\left\| G(y_k^o)\varepsilon_{uk} \right\|^2 - (y_k^o)^T y_k^o \\
&\le -(1-2k^*)\left\| y_k^o \right\|^2 + 4\left\| \tilde{G}(y_k^o)\hat{u}(y_k^o) \right\|^2 + 8G_M^2\left\| \tilde{W}_{uk}^T\phi(y_k^o) \right\|^2 + 8G_M^2\varepsilon_{uM}^2 \\
&\le -(1-2k^*)\left\| y_k^o \right\|^2 + 4\Psi_M^2\left\| \tilde{W}_{Ck}^T \right\|^2 + 8G_M^2\phi_M^2\left\| \tilde{W}_{uk} \right\|^2 + 8G_M^2\varepsilon_{uM}^2
\end{aligned}
\tag{A.9}
$$

Next, first different L_u can be expressed as

$$
\begin{aligned}
\Delta L_{uN} &= tr\left\{ \tilde{W}_{uk+1}^T \Omega \tilde{W}_{uk+1} \right\} - tr\left\{ \tilde{W}_{uk}^T \Omega \tilde{W}_{uk} \right\} \\
&= tr\left\{ \begin{matrix} \left(\tilde{W}_{uk} + \alpha_u \frac{\phi(y_k^o)}{\phi^T(y_k^o)\phi(y_k^o)+1} e_{uk}^T \right)^T \Omega \\ \times \left(\tilde{W}_{uk} + \alpha_u \frac{\phi(y_k^o)}{\phi^T(y_k^o)\phi(y_k^o)+1} e_{uk}^T \right) \end{matrix} \right\} - tr\left\{ \tilde{W}_{uk}^T \Omega \tilde{W}_{uk} \right\}
\end{aligned}
$$

$$= \frac{\alpha_u}{\phi^T(y_k^o)\phi(y_k^o)+1} tr\{e_{uk}\Omega\phi^T(y_k^o)\widetilde{W}_{uk} + \widetilde{W}_{uk}^T\phi(y_k^o)\Omega e_{uk}^T\}$$

$$+ \frac{\alpha_u^2\phi^T(y_k^o)\phi(y_k^o)}{(\phi^T(y_k^o)\phi(y_k^o)+1)^2} tr\{e_{uk}\Omega e_{uk}^T\} \qquad \text{(A.10)}$$

Substituting (34) into (A.10) we get

$$\Delta L_u = \frac{2\alpha_u}{\phi^T(y_k^o)\phi(y_k^o)+1} tr\{(-\widetilde{W}_{uk}^T\phi(y_k^o) - \frac{1}{2}R_y^{-1}G^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk}$$

$$+ \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk} + \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\varepsilon_{Vk}}{\partial y_{k+1}^o} - \varepsilon_{ek})\Omega\phi^T(y_k^o)\widetilde{W}_{uk}\}$$

$$+ \frac{\alpha_u^2\phi^T(y_k^o)\phi(y_k^o)}{(\phi^T(y_k^o)\phi(y_k^o)+1)^2} tr\{\Omega(-\widetilde{W}_{uk}^T\phi(y_k^o) - \frac{1}{2}R_y^{-1}G^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk}$$

$$+ \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk} + \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\varepsilon_{Vk}}{\partial y_{k+1}^o} - \varepsilon_{ek})^T$$

$$\times(-\widetilde{W}_{uk}^T\phi(y_k^o) - \frac{1}{2}R_y^{-1}G^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk}$$

$$+ \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk} + \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\varepsilon_{Vk}}{\partial y_{k+1}^o} - \varepsilon_{ek})\}$$

$$\leq -\frac{(3\alpha_u - 4\alpha_u^2)\|\Omega\|}{\phi^T(y_k^o)\phi(y_k^o)+1}\left\|\widetilde{W}_{uk}^T\phi(y_k^o)\right\|^2 + \frac{(a_u - 2\alpha_u^2)\Xi\|\Omega\|}{\phi^T(y_k^o)\phi(y_k^o)+1}\left\|\widetilde{W}_{uk}^T\phi(y_k^o)\right\|\left\|\widetilde{W}_{Vk}\right\|$$

$$\frac{3\alpha_u^2\Xi^2\left\|\widetilde{W}_{Vk}\right\|^2\|\Omega\|}{2(\phi^T(y_k^o)\phi(y_k^o)+1)} + \Delta\varepsilon_{eM}^2 + \frac{\alpha_u\|\Omega\|}{\phi^T(y_k^o)\phi(y_k^o)+1}\left\|\phi^T(y_k^o)\widetilde{W}_{uk}\right\|^2$$

$$+ \frac{\alpha_u\|\Omega\|}{\phi^T(y_k^o)\phi(y_k^o)+1}\left\|\frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk} + \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}\right\|^2$$

$$+ \frac{2\alpha_u^2\|\Omega\|}{\phi^T(y_k^o)\phi(y_k^o)+1}\left\|\frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\vartheta^T(y_{k+1}^o)}{\partial y_{k+1}^o}\widetilde{W}_{Vk} + \frac{1}{2}R_y^{-1}\widetilde{G}^T(y_k^o)\frac{\partial\varepsilon_{Vk}^T}{\partial y_{k+1}^o}\right\|^2$$

$$\leq -\frac{(3\alpha_u - 6\alpha_u^2)\phi_{\min}^2 \|\Omega\|}{\phi_M^2 + 1}\left\|\widetilde{W}_{uk}\right\|^2 + \frac{(\alpha_u^2 + \alpha_u)\Xi^2\|\Omega\|}{2(\phi^T(y_k^o)\phi(y_k^o)+1)}\left\|\widetilde{W}_{Vk}\right\|^2 + \|\Omega\|\Delta\varepsilon_{eM}^2$$

$$+\frac{(2\alpha_u^2 + \alpha_u)(\Xi\psi_M)^2\|\Omega\|}{2(\phi_M^2+1)G_M^2}\left\|\widetilde{W}_{Ck}\right\|^2\left\|\widetilde{W}_{Vk}\right\|^2 + \frac{(2\alpha_u^2 + \alpha_u)(\Xi\varepsilon_{VM}^{'}\psi_M)^2\|\Omega\|}{2(\phi_M^2+1)G_M^2}\left\|\widetilde{W}_{Ck}\right\|^2$$

$$\leq -\frac{(3\alpha_u - 6\alpha_u^2)\phi_{\min}^2 \|\Omega\|}{\phi_M^2 + 1}\left\|\widetilde{W}_{uk}\right\|^2 + \frac{(\alpha_u^2 + \alpha_u)\Xi^2\|\Omega\|}{2(\phi^T(y_k^o)\phi(y_k^o)+1)}\left\|\widetilde{W}_{Vk}\right\|^2$$

$$+\frac{(2\alpha_u^2 + \alpha_u)(\Xi\psi_M)^2\|\Omega\|}{4(\phi_M^2+1)G_M^2}\left\|\widetilde{W}_{Ck}\right\|^4 + \frac{(2\alpha_u^2 + \alpha_u)(\Xi\psi_M)^2\|\Omega\|}{4(\phi_M^2+1)G_M^2}\left\|\widetilde{W}_{Vk}\right\|^4$$

$$+\frac{(2\alpha_u^2 + \alpha_u)(\Xi\varepsilon_{VM}^{'}\psi_M)^2\|\Omega\|}{2(\phi_M^2+1)G_M^2}\left\|\widetilde{W}_{Ck}\right\|^2 + \|\Omega\|\Delta\varepsilon_{eM}^2 \qquad \text{(A.11)}$$

where $0 < \phi_{\min} < \left\|\phi(y_k^o)\right\|$ is ensured by the PE condition described in Remarks 1&2, $\left\|\Delta\varepsilon_{euk}\right\|^2 = \frac{\alpha_u \left\|\varepsilon_{ek}\right\|^2}{(\phi^T(y_k^o)\phi(y_k^o)+1)} \leq \Delta\varepsilon_{eM}^2$, which is a bounded positive constant.

Next, first difference L_{AN} can be expressed as

$$\Delta L_{AN} \leq \left(tr\left\{\widetilde{W}_{Vk+1}^T\Gamma\widetilde{W}_{Vk+1}\right\}\right)^2 - \left(tr\left\{\widetilde{W}_{Vk}^T\Gamma\widetilde{W}_{Vk}\right\}\right)^2$$

$$\leq \|\Gamma\|^2\left[\left(1 - \frac{\alpha_V(1-2a_V)\Delta\vartheta_{\min}^2}{(2-a_V)(\Delta\vartheta_M^2+1)}\right)\left\|\widetilde{W}_{Vk}\right\|^2 + \frac{\alpha_V(2a_V+1)\Delta\varepsilon_{VM}^2}{(2-a_V)(\Delta\vartheta_M^2+1)}\right]^2 - \|\Gamma\|^2\left\|\widetilde{W}_{Vk}\right\|^4$$

$$\leq \|\Gamma\|^2\left(1 - \frac{\alpha_V(1-2a_V)\Delta\vartheta_{\min}^2}{(\Delta\vartheta_M^2+1)} + \frac{\Delta\vartheta_{\min}^4}{12(\Delta\vartheta_M^2+1)^2}\right)\left\|\widetilde{W}_{Vk}\right\|^4$$

$$+\|\Gamma\|^2\frac{4\Delta\varepsilon_{VM}^2}{3(\Delta\vartheta_M^2+1)}\left\|\widetilde{W}_{Vk}\right\|^2 + \frac{4\Delta\varepsilon_{VM}^4}{9(\Delta\vartheta_M^2+1)^2} - \|\Gamma\|^2\left\|\widetilde{W}_{Vk}\right\|^4$$

$$\leq -\frac{\left(\alpha_V - 2a_V^2 - \frac{1}{12}\right)\Delta\vartheta_{\min}^2\|\Gamma\|^2}{(\Delta\vartheta_M^2+1)}\left\|\widetilde{W}_{Vk}\right\|^4 + \|\Gamma\|^2\frac{4\Delta\varepsilon_{VM}^2}{3(\Delta\vartheta_M^2+1)}\left\|\widetilde{W}_{Vk}\right\|^2$$

$$+\frac{4\|\Gamma\|^2 \Delta\varepsilon_{VM}^4}{9\left(\Delta\vartheta_M^2+1\right)^2} \tag{A.12}$$

Next, the first difference of L_{BN} can be expressed as

$$\begin{aligned}
\Delta L_{BN} &= \left(tr\left\{\widetilde{W}_{Ck+1}^T \Theta \widetilde{W}_{Ck+1}\right\}\right)^2 - \left(tr\left\{\widetilde{W}_{Ck}^T \Theta \widetilde{W}_{Ck}\right\}\right)^2 \\
&\le \|\Theta\|^2 \left(2\alpha_C^2 \frac{\Psi_M^2}{\Psi_{\min}^2}\left\|\widetilde{W}_{Ck}\right\|^2 + 2\Delta\bar{\varepsilon}_{CM}^2\right)^2 - \|\Theta\|^2\left\|\widetilde{W}_{Ck}\right\|^4 \\
&\le -\|\Theta\|^2\left(1-4\alpha_C^4 \frac{\Psi_M^4}{\Psi_{\min}^4}\right)\left\|\widetilde{W}_{Ck}\right\|^4 + \frac{8\alpha_C^2\Psi_M^2\|\Theta\|^2}{\Psi_{\min}^2}\left\|\widetilde{W}_{Ck}\right\|^2 \Delta\bar{\varepsilon}_{CM}^2 + \frac{4\|\Theta\|^2}{\Psi_{\min}^4}\Delta\bar{\varepsilon}_{CM}^4
\end{aligned} \tag{A.13}$$

Next, using (A.2), (A.5), (A.9), (A.11),(A.12)and (A.13) to form ΔL as:

$$\begin{aligned}
\Delta L \le& -\left(1-2k^*\right)\left\|y_k^o\right\|^2 + 4\Psi_M^2\left\|\widetilde{W}_{Ck}^T\right\|^2 + 8G_M^2\phi_M^2\left\|\widetilde{W}_{uk}\right\|^2 + 8G_M^2\varepsilon_{uM}^2 \\
& - 24G_M^2\phi_M^2\left(3\alpha_u - 6\alpha_u^2\right)\left\|\widetilde{W}_{uk}\right\|^2 + \frac{12G_M^2\phi_M^2\Xi^2\left(\alpha_u^2+\alpha_u\right)}{\phi_{\min}^2}\left\|\widetilde{W}_{Vk}\right\|^2 \\
& + \frac{6\left(2\alpha_u^2+\alpha_u\right)\left(\Xi\psi_M\phi_M\right)^2}{\phi_{\min}^2}\left\|\widetilde{W}_{Ck}\right\|^4 + \frac{6\left(2\alpha_u^2+\alpha_u\right)\left(\Xi\psi_M\phi_M\right)^2}{\phi_{\min}^2}\left\|\widetilde{W}_{Vk}\right\|^4 \\
& + \frac{12\left(2\alpha_u^2+\alpha_u\right)\left(\Xi\varepsilon'_{VM}\psi_M\right)^2}{\phi_{\min}^2}\left\|\widetilde{W}_{Ck}\right\|^2 + \frac{24\left(\phi_M^2+1\right)G_M^2\phi_M^2}{\phi_{\min}^2}\Delta\varepsilon_{eM}^2 \\
& - \frac{288\alpha_V\left(1-2a_V\right)\phi_M^2\Xi^2\left(G_M^2 + 12\frac{\psi_M^2\Delta\varepsilon_{VM}^2}{\Delta\vartheta_{\min}^2}\right)\left\|\widetilde{W}_{Vk}\right\|^2}{\phi_{\min}^2} + \frac{576G_M^2\phi_M^2\Xi^2}{\phi_{\min}^2\Delta\vartheta_{\min}^2}\Delta\varepsilon_{VM}^2 \\
& - \left(1-\alpha_C^2\right)\left\|e_{yk}\right\|^2 + 16\frac{\Psi_M^2}{\Psi_{\min}^2}\Delta\bar{\varepsilon}_{CM}^2 - 4\Psi_M^2\left(2-4\alpha_C^2\frac{\Psi_M^2}{\Psi_{\min}^2}\right)\left\|\widetilde{W}_{Ck}\right\|^2 \\
& - \frac{36\left(\varepsilon'_{VM}\psi_M\right)^2\phi_M^2\Xi^2}{\phi_{\min}^2}\left(1-2\alpha_C^2\frac{\Psi_M^2}{\Psi_{\min}^2}\right)\left\|\widetilde{W}_{Ck}\right\|^2 + \frac{72\left(\varepsilon'_{VM}\psi_M\right)^2\phi_M^2\Xi^2}{\phi_{\min}^2}\Delta\bar{\varepsilon}_{CM}^2
\end{aligned}$$

$$-\frac{216\left(\alpha_V-2a_V^2-\frac{1}{12}\right)(\Xi\psi_M\phi_M)^2}{\phi_{\min}^2}\left\|\tilde{W}_{Vk}\right\|^4+\frac{288(\Xi\psi_M\phi_M)^2\Delta\varepsilon_{VM}^2}{\phi_{\min}^2\Delta\vartheta_{\min}^2}\left\|\tilde{W}_{Vk}\right\|^2$$

$$+\frac{96(\Xi\psi_M\phi_M)^2\Delta\varepsilon_{VM}^4}{\phi_{\min}^2\Delta\vartheta_{\min}^2\left(\Delta\vartheta_M^2+1\right)}-\left(1-4\alpha_C^4\frac{\Psi_M^4}{\Psi_{\min}^4}\right)\frac{12(\Xi\psi_M\phi_M)^2}{\phi_{\min}^2}\left\|\tilde{W}_{Ck}\right\|^4$$

$$+\frac{96\alpha_C^2\Psi_M^2(\Xi\psi_M\phi_M)^2}{\phi_{\min}^2\Psi_{\min}^2}\left\|\tilde{W}_{Ck}\right\|^2\Delta\bar{\varepsilon}_{CM}^2+\frac{48(\Xi\psi_M\phi_M)^2}{\phi_{\min}^2\Psi_{\min}^4}\Delta\bar{\varepsilon}_{CM}^4$$

$$+\frac{4608(\Xi\psi_M\phi_M)^2\Delta\varepsilon_{VM}^4}{\phi_{\min}^2\Delta\vartheta_{\min}^4}+\frac{96(\Xi\psi_M)^2\phi_M^2\Delta\bar{\varepsilon}_{CM}^4}{\phi_{\min}^2\Psi_{\min}^4}$$

$$-48\left(1-2\alpha_C^2\frac{\Psi_M^2}{\Psi_{\min}^2}\right)\frac{(\Xi\psi_M)^2\phi_M^2\Delta\bar{\varepsilon}_{CM}^2}{\phi_{\min}^2\Psi_{\min}^2}\left\|\tilde{W}_{Ck}\right\|^2$$

$$\leq-\left(1-2k^*\right)\left\|y_k^o\right\|^2-\frac{288\left(\alpha_V-2a_V^2-\frac{1}{12}\right)\phi_M^2\Xi^2}{\phi_{\min}^2}$$

$$\times\left(G_M^2+12\frac{\psi_M^2\Delta\varepsilon_{VM}^2}{\Delta\vartheta_{\min}^2}\right)\left\|\tilde{W}_{Vk}\right\|^2-\left(1-\alpha_C^2\right)\left\|e_{yk}\right\|^2$$

$$-24G_M^2\phi_M^2\left(3\alpha_u-6\alpha_u^2-\frac{1}{3}\right)\left\|\tilde{W}_{uk}\right\|^2-4\Psi_M^2\left(1-4\alpha_C^2\frac{\Psi_M^2}{\Psi_{\min}^2}\right)\left\|\tilde{W}_{Ck}\right\|^2$$

$$-\frac{18\left(\varepsilon_{VM}^{'}\psi_M\right)^2\phi_M^2\Xi^2}{\phi_{\min}^2}\left(1-4\alpha_C^2\frac{\Psi_M^2}{\Psi_{\min}^2}\right)\left\|\tilde{W}_{Ck}\right\|^2$$

$$-24\left(1-4\alpha_C^2\frac{\Psi_M^2}{\Psi_{\min}^2}\right)\frac{(\Xi\psi_M)^2\phi_M^2\Delta\bar{\varepsilon}_{CM}^2}{\phi_{\min}^2\Psi_{\min}^2}\left\|\tilde{W}_{Ck}\right\|^2-\frac{216\left(\alpha_V-2a_V^2-\frac{1}{9}\right)(\Xi\psi_M\phi_M)^2}{\phi_{\min}^2}\left\|\tilde{W}_{Vk}\right\|^4$$

$$-\left(1-8\alpha_C^4\frac{\Psi_M^4}{\Psi_{\min}^4}\right)\frac{6(\Xi\psi_M\phi_M)^2}{\phi_{\min}^2}\left\|\tilde{W}_{Ck}\right\|^4+8G_M^2\varepsilon_{uM}^2+16\frac{\Psi_M^2}{\Psi_{\min}^2}\Delta\bar{\varepsilon}_{CM}^2$$

$$+\frac{24\left(\phi_M^2+1\right)G_M^2\phi_M^2}{\phi_{\min}^2}\Delta\varepsilon_{eM}^2+\frac{576G_M^2\phi_M^2\Xi^2}{\phi_{\min}^2\Delta\vartheta_{\min}^2}\Delta\varepsilon_{VM}^2$$

$$+\frac{72\left(\varepsilon_{VM}^{'}\psi_{M}\right)^{2}\phi_{M}^{2}\Xi^{2}}{\phi_{\min}^{2}}\Delta\bar{\varepsilon}_{CM}^{2}+\frac{96\left(\Xi\psi_{M}\phi_{M}\right)^{2}\Delta\varepsilon_{VM}^{4}}{\phi_{\min}^{2}\Delta\vartheta_{\min}^{2}\left(\Delta\vartheta_{M}^{2}+1\right)}$$

$$+\frac{48\left(\Xi\psi_{M}\phi_{M}\right)^{2}}{\phi_{\min}^{2}\Psi_{\min}^{4}}\Delta\bar{\varepsilon}_{CM}^{4}+\frac{4608\left(\Xi\psi_{M}\phi_{M}\right)^{2}\Delta\varepsilon_{VM}^{4}}{\phi_{\min}^{2}\Delta\vartheta_{\min}^{4}}+\frac{96\left(\Xi\psi_{M}\right)^{2}\phi_{M}^{2}\Delta\bar{\varepsilon}_{CM}^{4}}{\phi_{\min}^{2}\Psi_{\min}^{4}}$$

$$\leq-\left(1-2k^{*}\right)\left\|y_{k}^{o}\right\|^{2}-288\mathrm{A}\left(\alpha_{V}-2a_{V}^{2}-\frac{1}{12}\right)\left\|\widetilde{W}_{Vk}\right\|^{2}-\left(1-\alpha_{C}^{2}\right)\left\|e_{yk}\right\|^{2}$$

$$-24G_{M}^{2}\phi_{M}^{2}\left(3\alpha_{u}-6\alpha_{u}^{2}-\frac{1}{3}\right)\left\|\widetilde{W}_{uk}\right\|^{2}-4\mathrm{M}\left(1-4\alpha_{C}^{2}\frac{\Psi_{M}^{2}}{\Psi_{\min}^{2}}\right)\left\|\widetilde{W}_{Ck}\right\|^{2}$$

$$-\frac{216\left(\alpha_{V}-2a_{V}^{2}-\frac{1}{9}\right)\left(\Xi\psi_{M}\phi_{M}\right)^{2}}{\phi_{\min}^{2}}\left\|\widetilde{W}_{Vk}\right\|^{4}$$

$$-\left(1-8\alpha_{C}^{4}\frac{\Psi_{M}^{4}}{\Psi_{\min}^{4}}\right)\frac{6\left(\Xi\psi_{M}\phi_{M}\right)^{2}}{\phi_{\min}^{2}}\left\|\widetilde{W}_{Ck}\right\|^{4}+\varepsilon_{TM} \qquad \text{(A.14)}$$

where $\Xi=\left(\lambda_{\max}\left(R_{y}^{-1}\right)\vartheta_{M}^{'}G_{M}\right)$, $\eta=[\Psi_{M}^{2}+\frac{9\left(\varepsilon_{VM}^{'}\psi_{M}\right)^{2}\phi_{M}^{2}\Xi^{2}}{2\phi_{\min}^{2}}+\frac{6\left(\Xi\psi_{M}\right)^{2}\phi_{M}^{2}\Delta\bar{\varepsilon}_{CM}^{2}}{\phi_{\min}^{2}\Psi_{\min}^{2}}]$ and

$\rho=\frac{\phi_{M}^{2}\Xi^{2}}{\phi_{\min}^{2}}\left(G_{M}^{2}+12\frac{\psi_{M}^{2}\Delta\varepsilon_{VM}^{2}}{\Delta\vartheta_{\min}^{2}}\right)$ are positive constant and ε_{TM} is

$$\varepsilon_{TM}=8G_{M}^{2}\varepsilon_{uM}^{2}+16\frac{\Psi_{M}^{2}}{\Psi_{\min}^{2}}\Delta\bar{\varepsilon}_{CM}^{2}+\frac{24\left(\phi_{M}^{2}+1\right)G_{M}^{2}\phi_{M}^{2}}{\phi_{\min}^{2}}\Delta\varepsilon_{eM}^{2}$$

$$+\frac{\left(23G_{M}\phi_{M}\Xi\right)^{2}}{\phi_{\min}^{2}\Delta\vartheta_{\min}^{2}}\Delta\varepsilon_{VM}^{2}+\frac{72\left(\varepsilon_{VM}^{'}\psi_{M}\right)^{2}\phi_{M}^{2}\Xi^{2}}{\phi_{\min}^{2}}\Delta\bar{\varepsilon}_{CM}^{2}$$

$$+\frac{48\left(\Xi\psi_{M}\phi_{M}\right)^{2}}{\phi_{\min}^{2}\Psi_{\min}^{4}}\Delta\bar{\varepsilon}_{CM}^{4}+\frac{96\left(\Xi\psi_{M}\phi_{M}\right)^{2}\Delta\varepsilon_{VM}^{4}}{\phi_{\min}^{2}\Delta\vartheta_{\min}^{2}\left(\Delta\vartheta_{M}^{2}+1\right)}$$

$$+\frac{\left(68\Xi\psi_{M}\phi_{M}\right)^{2}\Delta\varepsilon_{VM}^{4}}{\phi_{\min}^{2}\Delta\vartheta_{\min}^{4}}+\frac{96\left(\Xi\psi_{M}\right)^{2}\phi_{M}^{2}\Delta\bar{\varepsilon}_{CM}^{4}}{\phi_{\min}^{2}\Psi_{\min}^{4}}$$

Therefore, ΔL is less than zero when the following inequalities hold

$$\left\|\tilde{e}_{yk}\right\| > \sqrt{\frac{\varepsilon_{TM}}{\left(1-\alpha_C^2\right)}} \equiv b_{ey} \text{ or}$$

$$\left\|\tilde{W}_{Ck}\right\| > \max\{\sqrt{\frac{\varepsilon_{TM}\Psi_{\min}^2}{4\left(\Psi_{\min}^2 - 4\alpha_C^2\Psi_M^2\right)\mathrm{M}}}, \sqrt[4]{\frac{\varepsilon_{TM}\Psi_{\min}^4\phi_{\min}^2}{6\left(\Psi_{\min}^4 - 8\alpha_C^4\Psi_M^4\right)\left(\Xi\psi_M\phi_M\right)^2}}\} \equiv b_{WC} \quad \text{or} \tag{A.15}$$

$$\left\|\tilde{W}_{Vk}\right\| > \max\{\sqrt{\frac{\varepsilon_{TM}}{288\mathrm{A}\left(\alpha_V - 2a_V^2 - \frac{1}{12}\right)}}, \sqrt[4]{\frac{\phi_{\min}^2}{216\left(\alpha_V - 2a_V^2 - \frac{1}{9}\right)\left(\Xi\psi_M\phi_M\right)^2}}\} \equiv b_{WV} \quad \text{or}$$

$$\left\|\tilde{W}_{uk}\right\| > \sqrt{\frac{\varepsilon_{TM}}{8G_M^2\phi_M^2\left(9\alpha_u - 18\alpha_u^2 - 1\right)}} \equiv b_{Wu} \quad \text{or}$$

$$\left\|y_k^o\right\| > \sqrt{\frac{\varepsilon_{TM}}{\left(1-2k^*\right)}} \equiv b_y$$

provided the tuning gains are selected according to (19) (29) and (32) for the NNCS (4). Using the standard Lyapunov extension [15], the system outputs, NN identifier and weight estimation errors, critic and action NN estimation errors are *UUB* in the mean while the system outputs never leave the compact set.

Next using (24) and (26), we have $\hat{u}\left(y_k^o\right) - u^*\left(y_k^o\right) = -\tilde{W}_u^T\phi\left(y_k^o\right) - \varepsilon_{uk}$. When $k \to \infty$, the upper bound of $\hat{u}\left(y_k^o\right) - u^*\left(y_k^o\right)$ can be represented as

$$\left\|\hat{u}\left(y_k^o\right) - u^*\left(y_k^o\right)\right\| \le \left\|\tilde{W}_u^T\phi\left(y_k^o\right)\right\| + \left\|\varepsilon_{uk}\right\| \le \left\|\tilde{W}_u^T\phi\left(y_k^o\right)\right\| + \varepsilon_{uM}$$

$$\le b_{Wu} + \varepsilon_{uM} \equiv \varepsilon_{bu} \tag{A.16}$$

Now, if the NN identifier, critic and action NN approximation errors ε_e, ε_u and ε_V are neglected as in [24,29] and when $k \to \infty$, ε_{TM} in (A.15) and ε_{bu} in (A.16) will become zero. In this case, it can be shown that the NN-based identification, action NN and critic NN estimation errors converge to zero asymptotically, i.e. $\hat{u}\left(y_k^o\right) \to u^*\left(y_k^o\right)$.

PAPER IV

STOCHASTIC OPTIMAL DESIGN FOR UNKNOWN NETWORKED CONTROL SYSTEM WITH COMMUNICATION NETWORK PROTOCOLS

H. Xu and S. Jagannathan

Abstract—In this paper, stochastic optimal control and estimation problems have been considered for linear discrete-time systems with wireless imperfections referred to as linear networked control system (NCS). The network imperfections include packet losses and random delays. For evaluating the impact of network reliability on controller performance, Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) are considered with NCS. First a novel observer is introduced to estimate the state vector in the presence of unknown system dynamics due to network imperfections and the communication protocols. Next, a novel stochastic optimal adaptive output-feedback controller by using adaptive dynamics programming (ADP) is utilized to solve the infinite horizon optimal regulation of NCS under the TCP and UDP protocol respectively by estimating the value function. Update laws for tuning the unknown parameters of proposed novel observer and value function estimator are derived. Stable regions of proposed observer for linear NCS under TCP and UDP with and without known system dynamics are given respectively. Lyapunov stability analysis indicate that for NCS under TCP, all signals are asymptotically stable (AS) in the mean and the estimated control and observed state signals converge to optimal control inputs and actual states of NCS in the mean respectively, and for NCS under UDP all signals are uniformly ultimately bounded (UUB) in the mean while the approximated control input converges close to its optimal

value with time in the mean. Simulation results are included to show the effectiveness of the proposed scheme.

I. INTRODUCTION

Networked Control Systems (NCS) [1] are feedback control systems wherein the control loop is closed through a real-time communication network. Although NCS brings many advantages (e.g. saves installation cost, etc.), insertion of a communication network into the feedback loop causes many challenging issues due to network imperfections such as network-induced delays and packet losses that occur while exchanging data among devices. In fact, the performance of the control system degrades significantly due to these network imperfections.

Therefore, recently, the authors in [1] analyzed the stability of NCS with network-induced delays, whereas the work in [2] proposed a stability region for NCS with network-induced delays and packet losses. The optimal controller design is derived for NCS with random delays in [3]. On the other hand, the authors in [4] introduced stochastic optimal control of NCS with network imperfections [5]. These optimal control designs [3-4] are obtained backward-in-time by assuming that the NCS system dynamics and information of network imperfection such as network-induced delays and packet losses, which cannot be obtained beforehand, are assumed to be known accurately. In addition, current NCS designs [1-4] did not include the impact of network protocols (e.g. TCP, UDP etc.) that cause these network imperfections until recently in [12].

On the contrary, adaptive dynamic programming (ADP) techniques, proposed by Werbos [6], intend to solve optimal control design for unknown nonlinear system in a forward-in-time manner instead of traditional optimal control scheme [8] where backward-in-time approach is utilized with known system dynamics. In ADP, the dynamic programming is utilized via value and/or policy iterations [6][8][27] to generate optimal control input. However, the value and policy iteration-based optimal control design [6][8][27] needs a significant number of iterations within a sample interval for convergence which can be an issue for closed-loop stability and hardware implementation. Less iteration within the sampling interval can lead to instability.

Therefore, Dierk and Jagannathan [9] utilized the Hamilton-Jacob-Bellman (HJB) equation in forward-in-time manner for the optimal control of a class of general unknown nonlinear affine discrete-time systems by using state feedback. Here, value and policy iterations are not utilized; instead the dynamic programming based optimal control over time utilizes past history of system states and cost errors thus making the technique suitable for real-time control. However, the ADP-approaches from [6-9][27] are not suitable for NCS since effects of network imperfections caused by practical network protocol is not considered. Addition of network protocol will require output feedback which is more involved than state feedback.

In our previous paper [18], stochastic optimal design of state-feedback NCS is undertaken in the presence of uncertain dynamics due to unknown network imperfections by assuming the states are measurable. However, the impact of network protocol such as the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) are not studied. For TCP or UDP protocols, an observer is required [12] in the controller design which can complicate the optimal controller design [18] and stability analysis.

Therefore, optimal adaptive output feedback control technique is undertaken in this paper to obtain stochastic optimal regulation of linear NCS in discrete-time under

TCP or UDP protocol with uncertain system dynamics and unknown network imperfections. The network imperfections considered in this paper include network-induced delays and packet losses. First, for implementing the output feedback under standard TCP or UDP protocol, a novel observer is introduced to estimate the system states when the dynamics are unknown.

Next, by using the observed system states and an initial stabilizing control, the value function is estimated [9] and its parameter vector is tuned online and forward-in-time by using Bellman Equation [7]. Eventually, stochastic control inputs which optimize the value function can be calculated based on parameters provided by the value function estimator. Compared with traditional optimal control theory which requires the knowledge of system dynamics to solve the Stochastic Riccati Equation (SRE), the proposed novel observer and value function estimator relax the need for system states and dynamics, and information on network-induced delays and packet losses respectively for NCS under TCP or UDP, and yields optimal control without using value or policy iterations.

This paper is organized as follows. In Section II, the background of NCS under TCP or UDP and traditional optimal control for linear discrete-time system is given first. Next, the stochastic optimal control of NCS under TCP without known system dynamics is derived and stability of proposed stochastic optimal scheme is verified by using Lyapunov theory in Section III. Section IV proposes stochastic optimal control of NCS under UDP with unknown system dynamics and analyzed stability of the proposed scheme based on Lyapunov theory. Then the effectiveness of proposed schemes is illustrated via numerical simulations in Section V, and Section VI provides concluding remarks.

II. BACKGROUND

A. NCS Under TCP or UDP

In Figure 1, the basic structure of NCS is shown where the feedback loop is closed over a communication network by using either TCP or UDP network protocol. Due to the presence of a communication network, two types of network-induced delays and two types of packet losses are observed: (1) $\tau_{sc}(t)$: sensor-to-controller delay, (2) $\tau_{ca}(t)$: controller-to-actuator delay, (3) $\gamma(t)$: indicator of packet lost at controller and (4) $\upsilon(t)$: indicator of packet lost at actuator.

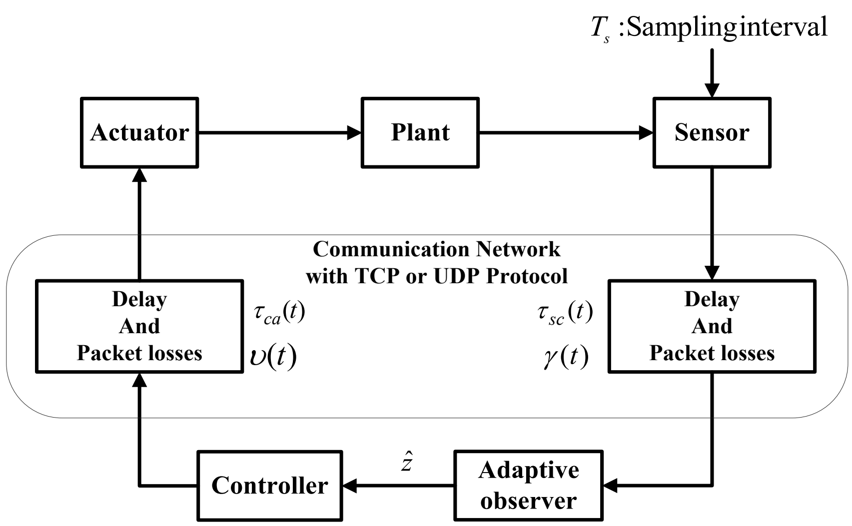


Fig 1. Networked Control System under TCP or UDP.

Based on standard TCP and UDP protocols [12] and other recent NCS results, the following assumption is needed for NCS under TCP or UDP [13-14]:

Assumption 1:

a) Sensor is time-driven; controller and actuator are event-driven [13].

b) Communication network is wide area network such that two types of networked-induced delays are independent, ergodic and unknown whereas their probability distribution functions are considered known. The sensor-to-controller delay is kept less than one sampling interval [14].

c) The sum of the two delays is bounded while initial state of the system is deterministic [14].

After incorporating the network-induced delays and packet losses, the original time-invariant plant $\dot{x}(t) = Ax(t) + Bu(t)$, $y(t) = Cx(t)$ can be expressed as

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu^{a}(t) \\ u^{a}(t) &= \upsilon(t)u^{c}(t - \tau_{ca}(t)) \end{aligned} \tag{1}$$

$$y(t) = \gamma(t)Cx(t - \tau_{sc}(t))$$

where $\gamma(t) = \begin{cases} \mathbf{I}^{n\times n} & \text{if controller received } x(t) \text{ from NCS plant at time } t \\ \mathbf{0}^{n\times n} & \text{if controller lost } x(t) \text{ at time } t \end{cases}$;

$\upsilon(t) = \begin{cases} \mathbf{I}^{m\times m} & \text{if actuator received control from controller at } t \\ \mathbf{0}^{m\times m} & \text{if controller lost control input at } t \end{cases}$ with $x(t) \in \mathbb{R}^{n\times n}$, $u^{c}(t), u^{a}(t) \in \mathbb{R}^{m\times m}$ and $y(t) \in \mathbb{R}^{n\times n}$ represent the system state, control inputs computed at the controller and control inputs received at the actuator, and output of NCS plant respectively and $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ denote the system matrices. According to Assumption 1, sum of network-induced delays is considered to be bounded. (i.e. $\tau_{sc}(t) + \tau_{ca}(t) < bT_s$ where b denotes the delay bound while T_s being the sampling interval.)

Since actuator is event-driven, control input received by actuator $u^{a}(t)$ to the plant is a piecewise constant. According to NCS under TCP or UDP protocols, at most b number of current and previous control inputs can be received at the actuator at the same time, and only the latest control input is allowed to be applied on the plant during any sampling interval (i.e. $[kT_s,(k+1)T_s)$, $\forall k$), and other previous control inputs are ignored. It is important to note that since controller and actuator are event-driven, the plant can implement control inputs at the time instant $kT_s + t_i^k, i = 0,1,2,...,\bar{d}$ and $t_i^k < t_{i-1}^k$ where $t_i^k = \tau_i^k - iT_s$ as shown in Figure 2 [13-14].

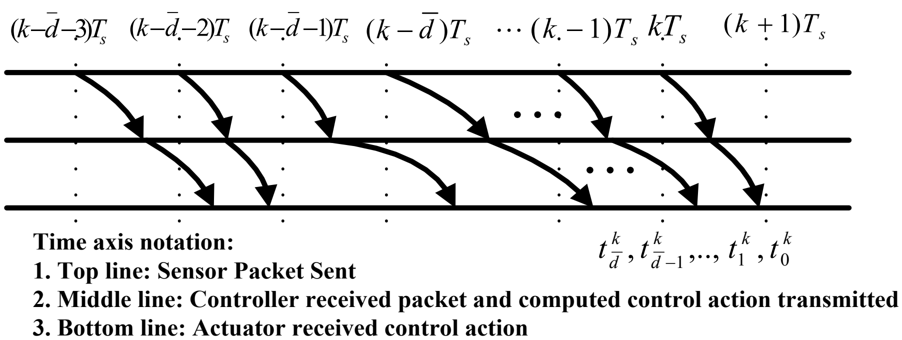


Fig 2. Timing diagram of signal transmission under TCP and UDP

Since controller is event-driven, the integration of (1) over a sampling interval $[kT_s,(k+1)T_s)$ yields

$$
\begin{aligned}
x_{k+1} &= A_s x_k + B_k^1 u_{k-1}^a + B_k^2 u_{k-2}^a + \cdots + B_k^b u_{k-b}^a + B_k^0 u_k^a \\
u_{k-i}^a &= \upsilon_{k-i} u_{k-i}^c \quad \forall i = 0,1,...,b, \forall k = 0,1,... \\
y_k &= \gamma_k C x_k \quad \forall k = 0,1,2,...
\end{aligned}
\tag{2}
$$

where $x_k = x(kT_s)$, $A_s = e^{AT_s}$, $B_k^0 = \int_{\tau_0^k}^{T_s} e^{A(T_s-s)} ds B \delta(T_s - \tau_0^k)$,

$B_k^i = \int_{\tau_i^k - iT_s}^{\tau_{i-1}^k-(i-1)T_s} e^{A(T_s-s)} ds B \bullet \delta(T_s + \tau_{i-1}^k - \tau_i^k) \bullet \delta(\tau_i^k - iT_s) \qquad \forall i = 1,2,...,b$, and

$\delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$. u_k^a is the control input received at the actuator and at time kT_s while u_k^c is the control input computed at the controller and at time kT_s, and γ_k, υ_k are the packet loss indicators at the controller and actuator respectively, which are also independent and identically distributed Bernoulli random variables with $P(\gamma_k = 1) = \bar{\gamma}$ and $P(\upsilon_k = 1) = \bar{\upsilon}$.

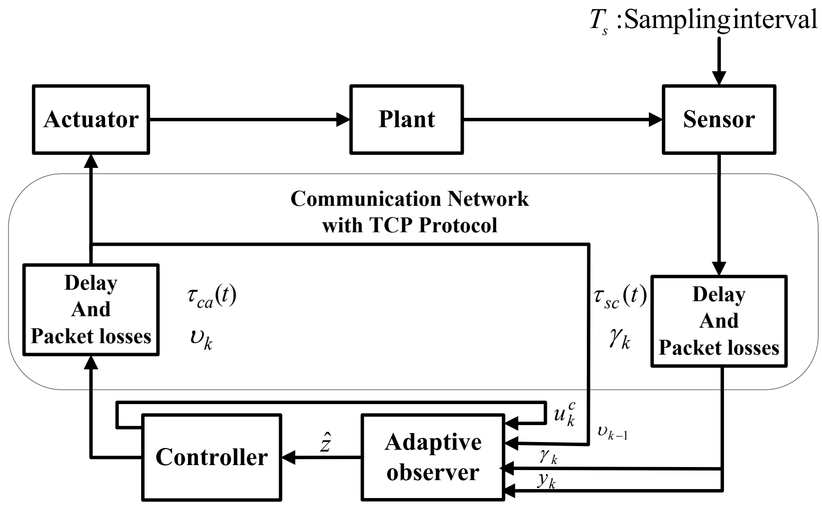


Fig 3. Block diagram of NCS under TCP

For simplifying the NCS representation (2), a new augment state variable consisting of current state and previous control inputs (i.e. $z_k = [x_k^T \ u_{k-1}^{cT} \ u_{k-2}^{cT} \cdots u_{k-b}^{cT}]^T \in \ u_{k-b}^{cT}]^T \in \mathbb{R}^{l=n+bm}$) is introduced. Equation (2) can be rewritten as

$$z_{k+1} = A_{zk} z_k + B_{zk} u_k^c, \ y_k = \Gamma_k z_k \quad \forall k = 0,1,2,... \tag{3}$$

where time-varying system matrices are given by

$$A_{zk} = \begin{bmatrix} A_s & \upsilon_{k-1} B_k^1 & \cdots & \cdots & \upsilon_{k-b+1} B_k^{b-1} & \upsilon_{k-b} B_k^b \\ 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & I_m & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & I_m & 0 \end{bmatrix}, B_{zk} = \begin{bmatrix} \upsilon_k B_k^0 \\ I_m \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \ \Gamma_k = [C\gamma_k \ 0 \ 0 \cdots 0] .$$

Here the system matrices are uncertain due to the presence of network imperfections caused by the communication protocol TCP or UDP with output vector alone is measurable. The block diagrams of NCS under TCP or UDP are shown in Figure 3 and 4.

Compared with UDP, the TCP [23][25] uses acknowledgments to indicate the reception of a packet (i.e. $\boldsymbol{\upsilon}_{k-1}$). Therefore, similar to [12], the following network information set ψ_k, ζ_k can be defined for NCS under TCP or UDP respectively as

$$\begin{aligned} \psi_k &= \{\mathbf{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\tau}_{k-1}, \boldsymbol{\upsilon}_{k-1}\} \\ \zeta_k &= \{\mathbf{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\tau}_{k-1}\} \end{aligned} \tag{4}$$

where $\mathbf{y}_k = \{y_k, y_{k-1}, ..., y_1\}, \boldsymbol{\gamma}_k = \{\gamma_k, \gamma_{k-1}, ..., \gamma_1\}$, $\boldsymbol{\tau}_k = \{\tau_k \ \ , \tau_{k-1}, ..., \tau_1\}$ and $\boldsymbol{\upsilon}_k = \{\upsilon_k, \upsilon_{k-1}, ..., \upsilon_1\}$ represent current and previous outputs, packet loss indicators at the controller and the actuator, and network-induced delays respectively.

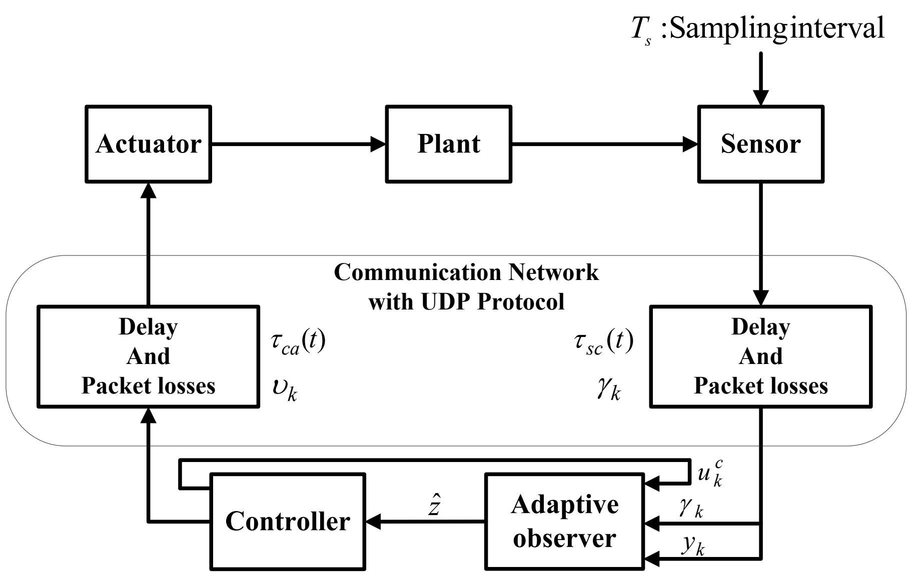


Fig 4. Block diagram of NCS under UDP

In this paper, based on the representation of NCS under unreliable communication network (TCP or UDP) (3), the stochastic optimal control of NCS under TCP or UDP are derived respectively by minimizing the related value function

$$
\begin{aligned}
V^*(z_k) &= \min_{u_m} E[\sum_{m=k}^{\infty}(z_m^T O z_m + u_m^T R u_m)|\psi_k] \quad \text{-- NCS under TCP} \\
V^*(z_k) &= \min_{u_m} E[\sum_{m=k}^{\infty}(z_m^T O z_m + u_m^T R u_m)|\zeta_k] \quad \text{-- NCS under UDP}
\end{aligned} \tag{5}
$$

where the value or cost function, $V^*(z_k)$, is defined in next section, O and R are symmetric positive semi-definite and symmetric positive definite constant matrices respectively and $E[(\bullet)|\psi_k], E[(\bullet)|\zeta_k]$ are the expected operators (i.e. mean value) of $\sum_{m=k}^{\infty}(z_m^T O z_m + u_m^T R u_m)$ based on the TCP information set ψ_k or UDP information set ζ_k defined in (4). Next a brief introduction of the optimal control of linear discrete-time system by using dynamic programming is given.

B. Trational Optimal Control of Discrete-Time Systems

Consider a linear discrete-time system given by

$$
x_{k+1} = A_k x_k + B_k u_k \tag{6}
$$

where $x_k \in \mathbb{R}^n$ is the system state vector, $u_k \in \mathbb{R}^m$ is the control input vector and $A_k \in \mathbb{R}^{n\times n}$, $B_k \in \mathbb{R}^{n\times m}$ are system matrices. According to [7], the infinite-horizon optimal value function can be defined as

$$
V^*(x_k) = \min_{u_l} \sum_{l=k}^{\infty} r(x_l, u_l) = \min_{u_l} \sum_{l=k}^{\infty}(x_l^T Q x_l + u_l^T R u_l) \tag{7}
$$

with $r(x_k, u_k) = x_k^T Q x_k + u_k^T R u_k$, and Q, R are symmetric positive semi-definite and definite matrices respectively.

Using dynamic programming, the optimization problem for linear discrete-time system (6) and (7) can be derived as

$$V^*(x_k) = \min_{u_k}(r(x_k, u_k) + V^*(x_{k+1})) \tag{8}$$

where $V^*(x_k)$ is the infinite-horizon optimal value function. Then Bellman equation in discrete-time can be represented as

$$0 = \min_{u_k}(r(x_k, u_k) + V^*(x_{k+1}) - V^*(x_k)) \tag{9}$$

Assuming that minimum on the right hand side of (3) exists and is unique, and then optimal control policy can be derived as [7]

$$u_k^* = -\frac{1}{2} R^{-1} B_k \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}} \tag{10}$$

Substituting optimal control policy (10) into Bellman equation (9), discrete-time (DT) HJB equation with optimal control u_k^* can be expressed as

$$0 = x_k^T Q x_k + \frac{1}{4} \frac{\partial V^{*T}(x_{k+1})}{\partial x_{k+1}} B_k^T R^{-1} B_k \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}} + V^*(x_{k+1}) - V^*(x_k) \tag{11}$$

For linear systems, value function (7) is taken as a quadratic function of state vector [7] given by

$$V^*(x_k) = x_k^T P_k x_k \tag{12}$$

where P_k is a positive-definite kernel matrix. Substituting (12) into (11), DT HJB equation becomes Riccati Equation (RE) given by

$$0 = A_k^T [P_{k+1} - P_{k+1} B_k (B_k^T P_{k+1} B_k + R)^{-1} B_k^T P_{k+1}] A_k + Q - P_k \tag{13}$$

with P_k becoming the solution of the Riccati Equation. Meanwhile, optimal control input can be expressed in terms of P_k and system matrices as

$$u_k^* = -(B_k^T P_{k+1} B_k + R)^{-1} B_k^T P_{k+1} A_k x_k \tag{14}$$

Traditionally RE (13) is solved backward-in-time and optimal control input (14) is obtained by using P_k and system dynamics A_k , B_k [7]. For linear systems with uncertain system dynamics, solving (13) and (14) is a challenge. Instead, in ADP, policy and/or value iteration algorithms [6-7] have been implemented to approximate the value function (12) and subsequently to obtain the control inputs based on estimated value function using (10) so that system matrices are not needed. However, with policy and value iteration-based schemes, it is not clear how to select the number of iterations required within a sampling interval for convergence and stability while meeting the hardware constraints. Inadequate number of policy and value iteration can lead to instability [15].

Therefore, to mitigate this drawback with policy and value iteration-based ADP, a time based value and policy update scheme will be proposed to solve stochastic optimal control of NCS under TCP or UDP. This proposed scheme works in forward-in-time manner and does not use an iterative methodology and known system dynamics as will be discussed in the next section.

C. Representing System States in Terms of Measured Output and Input Sequence Data

Similar to [26], NCS states can be expressed by using available measured data i.e. current and historical input and output sequences. Consider the NCS dynamics (3) as $z_{k+1} = A_{zk} z_k + B_{zk} u_k^c, y_k = \Gamma_k z_k$ where (A_{zk}, B_{zk}) is controllable and (A_{zk}, Γ_k) is observable. According to observability property of (A_{zk}, Γ_k) , the full system states z_k can be reconstructed by using observations of NCS output y_k over a long-time horizon. For current time k, NCS dynamics on the time horizon $[k-N,k]$ can be written as

$$z_k = \prod_{i=k-N}^{k-1} A_{zi} z_{k-N} + \left[B_{zk-1} \;\; A_{zk-1}B_{zk-2} \;\cdots\; \left(\prod_{i=k-N+1}^{k-1} A_{zi} \right) B_{zk-N} \right] \begin{bmatrix} u_{k-1}^c \\ u_{k-2}^c \\ \vdots \\ u_{k-N}^c \end{bmatrix} \tag{15}$$

$$\begin{bmatrix} y_{k-1} \\ y_{k-2} \\ \vdots \\ y_{k-N} \end{bmatrix} = \begin{bmatrix} \Gamma_{k-1} \prod_{i=k-N}^{k-2} A_{zi} \\ \Gamma_{k-2} \prod_{i=k-N}^{k-3} A_{zi} \\ \vdots \\ \Gamma_{k-N} \end{bmatrix} z_{k-N} + \begin{bmatrix} 0 & \Gamma_{k-1}B_{zk-2} & \Gamma_{k-1}A_{zk-2}B_{zk-3} & \cdots & \Gamma_{k-1}\left(\prod_{i=k-N+1}^{k-2} A_{zi} \right) B_{zk-N} \\ 0 & 0 & \Gamma_{k-2}B_{zk-2} & \cdots & \Gamma_{k-2}\left(\prod_{i=k-N+1}^{k-3} A_{zi} \right) B_{zk-N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Gamma_{k-N}B_{zk-N} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} u_{k-1}^c \\ u_{k-2}^c \\ \vdots \\ u_{k-N}^c \end{bmatrix}$$

Defining controllability and observability matrices of NCS as

$$\begin{aligned} F_{Nk}^o &= \left[B_{zk-1} \;\; A_{zk-1}B_{zk-2} \;\cdots\; \left(\prod_{i=k-N+1}^{k-1} A_{zi} \right) B_{zk-N} \right] \\ H_{Nk}^o &= \left[\left(\Gamma_{k-1} \prod_{i=k-N}^{k-2} A_{zi} \right)^T \;\; \left(\Gamma_{k-2} \prod_{i=k-N}^{k-3} A_{zi} \right)^T \;\cdots\; \Gamma_{k-N}^T \right]^T \end{aligned} \tag{16}$$

Meanwhile, Toeplitz matrix of Markov parameter and the available measured data (i.e. input and output sequences) over time horizon $[k-1,k-N]$ can be defined as

$$G_{kN}^o = \begin{bmatrix} 0 & \Gamma_{k-1}B_{zk-2} & \Gamma_{k-1}A_{zk-2}B_{zk-3} & \cdots & \Gamma_{k-1}\left(\prod_{i=k-N+1}^{k-2} A_{zi} \right) B_{zk-N} \\ 0 & 0 & \Gamma_{k-2}B_{zk-2} & \cdots & \Gamma_{k-2}\left(\prod_{i=k-N+1}^{k-3} A_{zi} \right) B_{zk-N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Gamma_{k-N}B_{zk-N} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbf{y}_{k-1}^{o} = \begin{bmatrix} y_{k-1} \\ y_{k-2} \\ \vdots \\ y_{k-N} \end{bmatrix}, \text{ and } \mathrm{u}_{k-1} = \begin{bmatrix} u_{k-1}^{c} \\ u_{k-2}^{c} \\ \vdots \\ u_{k-N}^{c} \end{bmatrix} \tag{17}$$

Using (16) and (17), (15) can be represented as

$$z_k = \prod_{i=k-N}^{k-1} A_{zi} z_{k-N} + F_{Nk}^{o}\mathbf{u}_{k-1} \tag{18a}$$

$$\mathbf{y}_{k-1}^{o} = H_{Nk}^{o} z_{k-N} + G_{Nk}^{o}\mathbf{u}_{k-1} \tag{18b}$$

Since (A_{zk}, Γ_k) is observable, there exists an observability index l such that H_{Nk}^{o} is full rank when $N \geq l$. Therefore, let $N \geq l$, then left inverse of H_{Nk}^{o} is given as

$$(H_{Nk}^{o})^{+} = (H_{Nk}^{oT} H_{Nk}^{o})^{-1} H_{Nk}^{oT} \tag{19}$$

Multiplying left inverse of H_{Nk}^{o} on both side of (18b), then $z_{k-N} = (H_{Nk}^{o})^{+}\mathbf{y}_{k-1}^{o} - (H_{Nk}^{o})^{+} G_{Nk}^{o}\mathbf{u}_{k-1}$. Substituting z_{k-N} into (18a), NCS states z_k can be expressed as

$$\begin{aligned} z_k &= \prod_{i=k-N}^{k-1} A_{zi} [(H_{Nk}^{o})^{+}\mathbf{y}_{k-1}^{o} - (H_{Nk}^{o})^{+} G_{Nk}^{o}\mathbf{u}_{k-1}] + F_{Nk}^{o}\mathbf{u}_{k-1} \\ &= \left(\prod_{i=k-N}^{k-1} A_{zi}\right)(H_{Nk}^{o})^{+}\mathbf{y}_{k-1}^{o} + \left[F_{Nk}^{o} - \left(\prod_{i=k-N}^{k-1} A_{zi}\right)(H_{Nk}^{o})^{+} G_{Nk}^{o}\right]\mathbf{u}_{k-1} \\ &= D_y^{o}\mathbf{y}_{k-1}^{o} + D_u^{o}\mathbf{u}_{k-1} \\ &= \begin{bmatrix} D_y^{o} & D_u^{o} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{k-1}^{o} \\ \mathbf{u}_{k-1} \end{bmatrix} \end{aligned} \tag{20}$$

where $D_y^{o} = \left(\prod_{i=k-N}^{k-1} A_{zi}\right)(H_{Nk}^{o})^{+}$ and $D_u^{o} = F_{Nk}^{o} - \left(\prod_{i=k-N}^{k-1} A_{zi}\right) \times (H_{Nk}^{o})^{+} G_{Nk}^{o}$. Also since (A_{zk}, Γ_k) is observable and $N \geq l$, D_y^{o} is full column rank and left inverse of D_y^{o} can be expressed as

$$(D_y^o)^+ = (D_y^{oT} D_y^o)^{-1} D_y^{oT} \tag{21}$$

On the other hand, it is important to note that $\|D_y^o\| \le D_M$ and D_M is known since D_y^o is composed by NCS dynamics A_{zk}, B_{zk}, Γ_k which are bounded as $\|A_{zk}\| \le A_M, \|B_{zk}\| \le B_M$, $\|\Gamma_k\| \le \Gamma_M$ and A_M, B_M, Γ_M are assumed known.

III. OPTIMAL CONTROL DESIGN FOR NCS UNDER TCP

In this section, observers [12] and ADP [17] are used to derive stochastic optimal control of NCS under network protocol such as TCP with uncertain system dynamics due to unknown network imperfections. First, a novel observer is designed online to estimate the augment system state vector at the controller. Second, we estimate the unknown value function for NCS with network imperfections under TCP. Third, a model-free online tuning of the parameters of observer and value function estimator by using ADP method incorporating the observed augment system states is proposed. Eventually, the convergence proof is given.

A. Novel Observer Design for NCS Under TCP

An observer or estimator is normally utilized when certain states are unavailable for measurement. However, the observer design for NCS requires the knowledge of system dynamics [12], which is unknown due to the presence of unknown network imperfections such as packet losses and random delays. Therefore, in this section, a novel observer is proposed to estimate the system states online for NCS under TCP by relaxing the need for system dynamics and network imperfections.

The observer design, similar to a [12][24], can be separated into two steps: 1) Innovation step where the system states are predicted based on current and previous system information (e.g. system outputs and control inputs); 2) Correction step where the estimated system states obtained from the innovation step in the previous time interval are adjusted based on current measured system output. Next, the details of novel observer design are given.

In the presence of TCP, the system states z_k can be estimated at the correction step of time instant, kT_s, as

$$\hat{z}_{k|k} = E[z_k|\psi_k] = E[\vartheta_k^T M_{k-1}\hat{s}_{k-1}|n_{k-1}] + \kappa_o \tilde{\mathbf{y}}_{k-2}^o \tag{22}$$

where $\tilde{\mathbf{y}}_k^o = \begin{bmatrix} \tilde{y}_k \\ \tilde{y}_{k-1} \\ \vdots \\ \tilde{y}_{k-N+1} \end{bmatrix}$, $M_k = \begin{bmatrix} I^n & 0 & \cdots & 0 & 0 \\ 0 & \upsilon_{k-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \upsilon_{k-b} & 0 \\ 0 & 0 & \cdots & 0 & \upsilon_k \end{bmatrix}$ and I^n is a $n \times n$ identity matrix, $\tilde{y}_k = y_k - \hat{y}_k$, $\tilde{z}_{k-1|k-1} = z_{k-1} - \hat{z}_{k-1|k-1}$ is the correction step estimation error at $(k-1)T_s$, $\hat{s}_{k-1} = a_{k-1} + n_{k-1}$, $a_{k-1} = E[[\hat{z}_{k-1|k-1}^T \; u_{k-1}^T]^T|\psi_{k-1}]$, with $\|a_{k-1}\| < \Psi_M$, and $s_{k-1} = E[[z_{k-1}^T \; u_{k-1}^T]^T|\psi_{k-1}] + n_{k-1}$ since $E[\hat{z}_{k-1|k-1}^T|\psi_{k-1}]$, M_k are known at time kT_s, $\hat{\vartheta}_k$ is the estimated parameters for the observer and n_k is independent and identically distributed white Gaussian noise (i.e. $n_k \sim N(0,\sigma_0^2)$ where σ_0 is the variance of white Gaussian noise with $\sigma_0 \neq 0$), and. The observer design is detailed using the two mentioned steps.

a) Innovation step at time kT_s:

In this step of the observer design, future system states $\hat{z}_{k+1|k}$ can be predicted as

$$\hat{z}_{k+1|k} = E[z_{k+1}|\psi_k] = E[\hat{\vartheta}_k^T M_k \hat{s}_k|n_k] + \kappa_o \tilde{\mathbf{y}}_{k-1}^o \tag{23}$$

According to (20), $\tilde{z}_{k|k}$ can be express as $\tilde{z}_{k|k} = D_y^o \tilde{\mathbf{y}}_{k-1}^o + E[\hat{\vartheta}_k^T M_k \hat{s}_k | n_k]$. Since desired system state vector is given by $z_{k+1} = \vartheta^T M_k s_k$, the prediction error $e_{k+1|k}$ in this step can be derived as

$$\begin{aligned}\tilde{z}_{k+1|k} &= z_{k+1} - \hat{z}_{k+1|k} \\ &= [U_z - \kappa_o (D_y^o)^+]\tilde{z}_{k|k} + [1 - \kappa_o (D_y^o)^+] E[\tilde{\vartheta}_k^T M_k \hat{s}_k | n_k]\end{aligned} \tag{24}$$

where $U_z = diag\{A_z, I_m, \cdots I_m\} \in \mathbb{R}^{(n+(\bar{d}-1)m)\times(n+(\bar{d}-1)m)}$, $\tilde{\vartheta}_k$ is the observer parameter estimation error $\tilde{\vartheta}_k = \vartheta - \hat{\vartheta}_k$. Since U_z is composed by A_z and identity matrix, and $\|A_z\| \le A_M$ (A_M is known), $\|U_z\| \le U_M$ and U_M is also known. Therefore, κ_o is designed as $\kappa_o = (U_M - \eta) D_M$ where $\|U_z\| \le U_M$ and $0 < \eta < \dfrac{1}{\sqrt{2}}$. Meanwhile $\|U_z - \kappa_o (D_y^o)^+\| =$ $\|U_z - (U_M - \eta) D_M (D_y^o)^+\| < \eta < 1/\sqrt{2}$.

b) Correction step at time $(k+1)T_s$:

Now define the update law for the parameter vector $\hat{\vartheta}_k$ of the observer as

$$\hat{\vartheta}_{k+1} = \hat{\vartheta}_k + \alpha_o E\left[\frac{M_k \hat{s}_k (y_{k+1} - \Gamma_{k+1} \hat{z}_{k+1|k})^T}{(M_k \hat{s}_k \Gamma_{k+1})^T M_k \hat{s}_k \Gamma_{k+1} + 1} \middle| n_k \right] \tag{25}$$

Substitute (3) in (25), $\hat{\vartheta}_{k+1}$ can be expressed as

$$\hat{\vartheta}_{k+1} = \hat{\vartheta}_k + \alpha_o \gamma_{k+1} E\left[\frac{M_k \hat{s}_k \Gamma_{k+1} \tilde{z}_{k+1|k}^T}{(M_k \hat{s}_k \Gamma_{k+1})^T M_k \hat{s}_k \Gamma_{k+1} + 1} \middle| n_k \right] \tag{26}$$

where α_o is the tuning parameter satisfying $0 < \alpha_o < 1$. Meanwhile, observer parameter estimation error dynamics $\tilde{\vartheta}_k$ can be represented as

$$\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_k - \alpha_o \gamma_{k+1} E\left[\frac{M_k \hat{s}_k \Gamma_{k+1} \tilde{z}^T_{k+1|k}}{(M_k \hat{s}_k \Gamma_{k+1})^T M_k \hat{s}_k \Gamma_{k+1} + 1} \middle| n_k \right] \tag{27}$$

Eventually, at time $(k+1)T_s$ in the correction step, the observed state $\hat{z}_{k+1|k+1}$ and the estimation error dynamics $\tilde{z}_{k+1|k+1}$ in this step can be expressed as

$$\hat{z}_{k+1|k+1} = E[z_{k+1} | \psi_{k+1}] = E[\hat{\vartheta}^T_{k+1} M_k \hat{s}_k | n_k] + \kappa_o \tilde{\mathbf{y}}^o_{k-1} \tag{28}$$

$$\begin{aligned} \tilde{z}_{k+1|k+1} &= z_{k+1} - \hat{z}_{k+1|k+1} \\ &= [U_z - \kappa_o (D^o_y)^+] \tilde{z}_{k|k} + [1 - \kappa_o (D^o_y)^+] E[\tilde{\vartheta}^T_{k+1} M_k \hat{s}_k | n_k] \end{aligned}$$

Next the convergence of observer parameter estimation error vector $\tilde{\vartheta}_k$ and observation errors $\tilde{z}_{k|k}$ is demonstrated. Before convergence proof, the following assumption is needed.

Assumption 2 [12]: *(Observability)* In order to meet the observability criterion, critical arrival probability of packets between the sensor and the controller need to be in the region [12] defined by $P(\gamma_k = 1) > \dfrac{1}{N_0}$ where $P(\gamma_k = 1)$ is the arrival probability at the controller and N_0 is a finite positive constant.

Remark 2: This assumption implies that a broken communication link is not present between the controller and the system which in turn ensures that there exists at least one packet that traverses through the network so as to observe the system states.

Theorem 1: *(Convergence of observer parameter estimation and errors dynamics):* Let the proposed novel observer, estimation errors and parameter vector update be defined by (22), (24) and (25) respectively. Under Assumption 2 and TCP, there exists positive constant α_o and η satisfying $\frac{1}{2}\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\sqrt{\frac{\chi_{\min}^2-3}{2(\chi_{\min}^2+1)}}\right)<\alpha_o<\frac{1}{2}\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\right.$ $\left.+\sqrt{\frac{\chi_{\min}^2-3}{2(\chi_{\min}^2+1)}}\right)$ and $0<\eta<\frac{1}{\sqrt{(U_M-1)^2(2+16\alpha_o+32\alpha_o^2)}}$ such that the estimation errors $\tilde{z}_{k|k}$ and parameter estimation errors $\tilde{\vartheta}_k$ (27) converge to zero asymptotically, where $\chi_{\min}^2 \le \left\|\Gamma_{k+1}M_k\hat{s}_k\right\|^2$.

Proof: Consider the Lyapunov candidate function $L_o = L_{\tilde{z}} + L_{\vartheta}$ where $L_{\tilde{z}} = tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i|i}^T\Xi\tilde{z}_{i|i}\}$ and $L_{\vartheta}=14tr\{\sum_{i=k-N_0}^{k}\tilde{\vartheta}_i^T\tilde{\vartheta}_i\}$ with $\Xi=\frac{1}{(\Psi_M^2+1)(U_M-1)^2}\mathbf{I}$ is a positive definite matrix with $\mathbf{I}$ being an identity matrix. Then the first difference can be expressed as $\Delta L_o = \Delta L_{\tilde{z}} + \Delta L_{\vartheta}$. Now take first term $L_{\tilde{z}}$ (i.e. $\Delta L_{\tilde{z}} = tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i+1|i+1}^T\Xi$ $\tilde{z}_{i+1|i+1}\} - tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i|i}^T\Xi\tilde{z}_{i|i}\}$). Using (27) and (28), and

Applying the Cauchy-Schwartz inequality reveals

$$
\begin{aligned}
\Delta L_{\tilde{z}} &= tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i+1|i+1}^T\Xi\tilde{z}_{i+1|i+1}\} - tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i|i}^T\Xi\tilde{z}_{i|i}\} \\
&= tr\{\sum_{i=k-N_0}^{k}\{[U_z-\kappa_o(D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o(D_y^o)^+]E[\tilde{\vartheta}_{i+1}^T M_i\hat{s}_i|n_i]\}^T \\
&\quad \times \Xi\{[U_z-\kappa_o(D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o(D_y^o)^+]\times E[\tilde{\vartheta}_{i+1}^T M_i\hat{s}_i|n_i]\}\} - tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i|i}^T\Xi\tilde{z}_{i|i}\} \\
&\le 2\Xi tr\{\sum_{i=k-N_0}^{k}[U_z-\kappa_o(D_y^o)^+]^T\tilde{z}_{i|i}^T\tilde{z}_{i|i}[U_z-\kappa_o(D_y^o)^+]\} \\
&\quad + 2tr\{\sum_{i=k-N_0}^{k}E[\hat{s}_iM_i\tilde{\vartheta}_{i+1}^T[1-\kappa_o(D_y^o)^+]^T \\
&\quad \times[1-\kappa_o(D_y^o)^+]\tilde{\vartheta}_{i+1}M_i^Ts_i^T|n_i]\} - \Xi tr\{\sum_{i=k-N_0}^{k}\tilde{z}_{i|i}^T\tilde{z}_{i|i}\}
\end{aligned}
\qquad (29)
$$

$$\leq \sum_{i=k-N_0}^{k} (2\Xi \left\| U_z - \kappa_o (D_y^o)^+ \right\|^2 \left\| \widetilde{z}_{i|i} \right\|^2 - \Xi \left\| \widetilde{z}_{i|i} \right\|^2)$$

$$+ 2 \sum_{i=k-N_0}^{k} (\left\| 1 - \kappa_o (D_y^o)^+ \right\|^2 tr\{ E[\hat{s}_i M_i M_i^T s_i^T | n_i \} tr\{ \widetilde{\vartheta}_{i+1}^T \widetilde{\vartheta}_{i+1} \})$$

$$\leq -(1-2\eta^2)\Xi \sum_{i=k-N_0}^{k} \left\| \widetilde{z}_{i|i} \right\|^2 + 2 \sum_{i=k-N_0}^{k} (\left\| 1 - \kappa_o (D_y^o)^+ \right\|^2 \left\| M_i \hat{s}_i \right\|^2 \Xi tr\{ \widetilde{\vartheta}_{i+1}^T \widetilde{\vartheta}_{i+1} \})$$

$$\leq -(1-2\eta^2)\Xi \sum_{i=k-N_0}^{k} \left\| \widetilde{z}_{i|i} \right\|^2 + 2 \sum_{i=k-N_0}^{k} ((U_M - \eta - 1)^2 \left\| M_i \hat{s}_i \right\|^2 \Xi tr\{ \widetilde{\vartheta}_{i+1}^T \widetilde{\vartheta}_{i+1} \})$$

Next, according to parameters estimation error dynamics (27), the term ΔL_ϑ can be derived as

$$\Delta L_\vartheta = 14 tr\{ \sum_{i=k-N_0}^{k} \widetilde{\vartheta}_{i+1}^T \widetilde{\vartheta}_{i+1} \} - 14 tr\{ \sum_{i=k-N_0}^{k} \widetilde{\vartheta}_i^T \widetilde{\vartheta}_i \} \tag{30}$$

$$= 14 \sum_{i=k-N_0}^{k} \{ tr\{ (\widetilde{\vartheta}_i - \alpha_o \gamma_{i+1} E\left[\frac{M_i s_i \Gamma_{i+1} \widetilde{z}_{i+1|i}^T}{(M_i \hat{s}_i \Gamma_{i+1})^T M_i \hat{s}_i \Gamma_{i+1} + 1} \middle| n_i \right])^T (\widetilde{\vartheta}_i$$

$$- \alpha_o \gamma_{i+1} E\left[\frac{M_i s_i \Gamma_{i+1} \widetilde{z}_{i+1|i}^T}{(M_i \hat{s}_i \Gamma_{i+1})^T M_i \hat{s}_i \Gamma_{i+1} + 1} \middle| n_i \right]) \} - 14 tr\{ \widetilde{\vartheta}_i^T \widetilde{\vartheta}_i \} \}$$

$$= 14 \sum_{i=k-N_0}^{k} \{ tr\{ \widetilde{\vartheta}_i^T \widetilde{\vartheta}_i \} - 28 tr\{ \sum_{i=k-N_0}^{k} (\alpha_o \gamma_{i+1}$$

$$\times E\left[\frac{M_i s_i \Gamma_{i+1} \widetilde{z}_{i+1|i}^T}{(M_i \hat{s}_i \Gamma_{i+1})^T M_i \hat{s}_i \Gamma_{i+1} + 1} \middle| n_i \right])^T \widetilde{\vartheta}_i \} + 14 tr\{ \sum_{i=k-N_0}^{k} (\alpha_o \gamma_{i+1}$$

$$\times E\left[\frac{M_i s_i \Gamma_{i+1} \widetilde{z}_{i+1|i}^T}{(M_i \hat{s}_i \Gamma_{i+1})^T M_i \hat{s}_i \Gamma_{i+1} + 1} n_i \right])^T (\alpha_o \gamma_{i+1}$$

$$\times E\left[\frac{M_i s_i \Gamma_{i+1} \widetilde{z}_{i+1|i}^T}{(M_i \hat{s}_i \Gamma_{i+1})^T M_i \hat{s}_i \Gamma_{i+1} + 1} \middle| n_i \right]) \} - 14 tr\{ \sum_{i=k-N_0}^{k} \widetilde{\vartheta}_i^T \widetilde{\vartheta}_i \}$$

Substituting (24) into (30), ΔL_ϑ can be expressed as

$$\Delta L_{\vartheta} \le 14 \sum_{i=k-N_0}^{k} \{ -2\alpha_o \gamma_{i+1} tr\{E\left[\left. \frac{\widetilde{\vartheta}_i^T M_i s_i \Gamma_{i+1} (M_i s_i \Gamma_{i+1})^T \widetilde{\vartheta}_i}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}$$

$$-2 \sum_{i=k-N_0}^{k} tr\{\alpha_o \gamma_{i+1} E\left[\left. \frac{\widetilde{\vartheta}_i^T M_i s_i \Gamma_{i+1} [U_z - \kappa_o (D_y^o)^+]^T \widetilde{z}_{i|i}^T}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}$$

$$+2 \sum_{i=k-N_0}^{k} tr\{\alpha_o^2 \gamma_{i+1}^2 E\left[\left. \frac{[U_z - \kappa_o (D_y^o)^+]^T \widetilde{z}_{i|i}^T \widetilde{z}_{i|i} [U_z - \kappa_o (D_y^o)^+]}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}$$

$$+2 \sum_{i=k-N_0}^{k} tr\{\alpha_o^2 \gamma_{i+1}^2 E\left[\left. \frac{\widetilde{\vartheta}_i^T M_i s_i \Gamma_{i+1} (M_i s_i \Gamma_{i+1})^T \widetilde{\vartheta}}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}$$

$$\le 14 \sum_{i=k-N_0}^{k} \{ -\alpha_o \gamma_{i+1} tr\{\widetilde{\vartheta}_i^T \widetilde{\vartheta}_i\} + 2\alpha_o^2 \gamma_{i+1}^2 tr\{\widetilde{\vartheta}_i^T \widetilde{\vartheta}_i\}$$

$$+2\alpha_o \gamma_{i+1} tr\{E\left[\left. \frac{\widetilde{\vartheta}_i^T \widetilde{\vartheta}_i}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}\} + \sum_{i=k-N_0}^{L} \alpha_o \gamma_{i+1} (1 + 2\alpha_o \gamma_{i+1}) \tag{31}$$

$$\times tr\{E\left[\left. \frac{[U_z - \kappa_o (D_y^o)^+]^T \widetilde{z}_{i|i}^T \widetilde{z}_{i|i} [U_z - \kappa_o (D_y^o)^+]}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}$$

$$\le -14\alpha_o \sum_{i=k-N_0}^{k} \{\gamma_{i+1} (1 - \alpha_o \gamma_{i+1} - \frac{2}{\sigma_0^2 + 1}) tr\{\widetilde{\vartheta}_i^T \widetilde{\vartheta}_i\}\} + 14\alpha_o \sum_{i=k-N_0}^{k} \gamma_{i+1} (1 + 2\alpha_o \gamma_{i+1})$$

$$\times tr\{E\left[\left. \frac{[U_z - \kappa_o (D_y^o)^+]^T \widetilde{z}_{i|i}^T \widetilde{z}_{i|i} [U_z - \kappa_o (D_y^o)^+]}{(M_i s_i \Gamma_{i+1})^T M_i s_i \Gamma_{i+1} + 1} \right| n_i \right]\}$$

$$\le -14\alpha_o \sum_{i=k-N_0}^{k} \gamma_{i+1} (1 - \alpha_o \gamma_{i+1} - \frac{2}{\chi_{\min}^2 + 1}) \left\| \widetilde{\vartheta}_i \right\|^2 + 14\alpha_o \sum_{i=k-N_0}^{k} \gamma_{i+1} (1 + 2\alpha_o \gamma_{i+1}) \frac{\eta^2}{\chi_{\min}^2 + 1} \left\| \widetilde{z}_{i|i} \right\|^2$$

Finally consider the overall first difference and using (29) and (31), ΔL_o can be expressed as

$$\Delta L_o = \Delta L_{\tilde{z}} + \Delta L_{\vartheta}$$

$$= tr\{ \sum_{i=k-N_0}^{k} \widetilde{z}_{i+1|i+1}^T \Xi \widetilde{z}_{i+1|i+1} \} - tr\{ \sum_{i=k-N_0}^{k} \widetilde{z}_{i|i}^T \Xi \widetilde{z}_{i|i} \} + 14 tr\{ \sum_{i=k-N_0}^{k} \widetilde{\vartheta}_{i+1}^T \widetilde{\vartheta}_{i+1} \} - 14 tr\{ \sum_{i=k-N_0}^{k} \widetilde{\vartheta}_i^T \widetilde{\vartheta}_i \}$$

$$\le -(1 - 2\eta^2) \Xi \sum_{i=k-N_0}^{k} \left\| \widetilde{z}_{i|i} \right\|^2 + 2 \sum_{i=k-N_0}^{k} ((U_M - \eta - 1)^2 \left\| M_i \hat{s}_i \right\|^2 \Xi tr\{\widetilde{\vartheta}_{i+1}^T \widetilde{\vartheta}_{i+1}\}) \tag{32}$$

$$+14\sum_{i=k-N_0}^{k}\{-\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1}-\frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_i\right\|^2+\alpha_o\gamma_{i+1}(1+2\alpha_o\gamma_{i+1})\frac{\eta^2}{\chi_{\min}^2+1}\left\|\tilde{z}_{i|i}\right\|^2\}$$

Since packet loss indicator γ_{k+1} can be equal to 0 or 1, ΔL_o (32) needs to be separated into two different cases for further consideration as given below.

Case 1: $\gamma_{k+1}=1$ (No packet losses).

Substituting γ_{k+1} value into (32), ΔL_o can be derived as

$$\Delta L_o=\Delta L_{\tilde{z}}+\Delta L_{\vartheta} \tag{33}$$

$$\leq-(1-2\eta^2)\Xi\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2+2\sum_{i=k-N_0}^{k}((U_M-\eta-1)^2\left\|M_i\hat{s}_i\right\|^2\Xi tr\{\tilde{\vartheta}_{i+1}^T\tilde{\vartheta}_{i+1}\})$$

$$+14\sum_{i=k-N_0}^{k}\{-\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1}-\frac{2}{\sigma_0^2+1})\left\|\tilde{\vartheta}_i\right\|^2+\alpha_o\gamma_{i+1}(1+2\alpha_o\gamma_{i+1})\frac{\eta^2}{\chi_{\min}^2+1}\left\|\tilde{z}_{i|i}\right\|^2\}$$

$$\leq-(1-2\eta^2)\Xi\left\|\tilde{z}_{k|k}\right\|^2+2(U_M-\eta-1)^2\left\|M_k\hat{s}_k\right\|^2\Xi tr\{\tilde{\vartheta}_{k+1}^T\tilde{\vartheta}_{k+1}\}$$

$$-14\alpha_o(1-\alpha_o-\frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_k\right\|^2+14\alpha_o(1+2\alpha_o)\frac{\eta^2}{\chi_{\min}^2+1}\left\|\tilde{z}_{k|k}\right\|^2$$

$$\leq-\frac{(1-2\eta^2)}{(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2+2(1-\alpha_o(1-\alpha_o-\frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_k\right\|^2$$

$$+\alpha_o(1+2\alpha_o)\frac{2\eta^2}{\chi_{\min}^2+1}\left\|\tilde{z}_{k|k}\right\|^2-14\alpha_o(1-\alpha_o-\frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_k\right\|^2+14\alpha_o(1+2\alpha_o)\frac{\eta^2}{\chi_{\min}^2+1}\left\|\tilde{z}_{k|k}\right\|^2$$

$$\leq-\frac{(1-\eta^2(U_M-1)^2(2+16\alpha_o+32\alpha_o^2))}{(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2-(16\alpha_o(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\alpha_o)-2)\left\|\tilde{\vartheta}_k\right\|^2$$

Case 2: $\gamma_{k+1}=0$. (with packet losses)

Based on Assumption 2, packets lost probability between the sensor and controller has to satisfy the observer stability region (i.e. $P(\gamma_k = 1) > \frac{1}{N_0}$). Therefore, if $\gamma_{k+1} = 0$, then there exists $j \in [k - N_0, k]$ such that $\gamma_{k+1} = \gamma_k = \cdots = \gamma_{k-j+1} = 0$ and $\gamma_{k-j} = 1$. Therefore, using update law for observer estimated parameters (25), observer parameters estimation errors $\tilde{\vartheta}_{k+1}$ can be expressed as $\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_{k-j}$. Then, substituting $\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_{k-j}$ into (32), ΔL_o can be derived as

$$
\begin{aligned}
\Delta L_o &= \Delta L_{\tilde{z}} + \Delta L_{\vartheta} \\
&\le -(1-2\eta^2)\Xi \sum_{i=k-N_0}^{k} \left\|\tilde{z}_{i|i}\right\|^2 + 2\sum_{i=k-N_0}^{k} ((U_M - \eta - 1)^2 \left\|M_i \hat{s}_i\right\|^2 \Xi tr\{\tilde{\vartheta}_{i+1}^T \tilde{\vartheta}_{i+1}\}) \\
&+14 \sum_{i=k-N_0}^{k} \{-\alpha_o \gamma_{i+1}(1-\alpha_o \gamma_{i+1} - \frac{2}{\chi_{\min}^2 + 1})\left\|\tilde{\vartheta}_i\right\|^2 + \alpha_o \gamma_{i+1}(1+2\alpha_o \gamma_{i+1})\frac{\eta^2}{\chi_{\min}^2+1}\left\|\tilde{z}_{i|i}\right\|^2\} \qquad (34) \\
&\le -(1-2\eta^2)\Xi \sum_{i=k-N_0}^{k} \left\|\tilde{z}_{i|i}\right\|^2 + 2(U_M - \eta - 1)^2 \left\|M_{k-j}\hat{s}_{k-j}\right\|^2 \Xi tr\{\tilde{\vartheta}_{k-j}^T \tilde{\vartheta}_{k-j}\}) \\
&-14\alpha_o(1-\alpha_o - \frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_{k-j-1}\right\|^2 + 14\alpha_o(1+2\alpha_o)\frac{\eta^2}{\chi_{\min}^2+1}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2 \\
&\le -\frac{(1-2\eta^2)}{(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2 - 14\alpha_o(1-\alpha_o - \frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_{k-j-1}\right\|^2 + 2(U_M - \eta - 1)^2 \\
&* \left\|M_{k-j}\hat{s}_{k-j}\right\|^2 \Xi tr\{\tilde{\vartheta}_{k-j}^T \tilde{\vartheta}_{k-j}\} + 14\alpha_o(1+2\alpha_o)\frac{\eta^2}{\chi_{\min}^2+1}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2 \\
&\le -\frac{(1-2\eta^2)}{(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2 + \frac{2\eta^2\alpha_o(1+2\alpha_o)}{\chi_{\min}^2+1}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2 \\
&+2(1-\alpha_o(1-\alpha_o - \frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_{k-j-1}\right\|^2 \; -14\alpha_o(1-\alpha_o - \frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_{k-j-1}\right\|^2
\end{aligned}
$$

$$+14\alpha_o(1+2\alpha_o)\frac{\eta^2}{\chi_{\min}^2+1}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2$$

$$\leq -\frac{(1-\eta^2(U_M-1)^2(2+16\alpha_o+32\alpha_o^2))}{(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2-(16\alpha_o(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\alpha_o)-2)\left\|\tilde{\vartheta}_{k-j-1}\right\|^2$$

Since Ξ is defined as $\Xi=\frac{1}{\Psi_M^2+1}\mathbf{I}$ and α_o,η are positive constants which satisfy

$$\frac{1}{2}\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\sqrt{\frac{\chi_{\min}^2-3}{2(\chi_{\min}^2+1)}}\right)<\alpha_o<\frac{1}{2}\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}+\sqrt{\frac{\chi_{\min}^2-3}{2(\chi_{\min}^2+1)}}\right)$$

and

$0<\eta<\frac{1}{\sqrt{(U_M-1)^2(2+16\alpha_o+32\alpha_o^2)}}$, then according to (33) and (34), ΔL_o is negative definite and L_o is positive definite. Therefore, observer error dynamics $\tilde{z}_{k|k}$ and its parameter estimation errors $\tilde{\vartheta}_k$ for NCS under TCP are asymptotically stable in the mean for both cases. In other words, as $k\to\infty$, $\tilde{z}_{k|k}\to 0$ and $\tilde{\vartheta}_k\to 0$.

Next, stochastic optimal control for NCS under TCP with estimated system states $\hat{z}_{k|k}$ is proposed by using the value function estimator.

B. Value Function Definition for NCS Under TCP

Consider NCS under TCP with network imperfections represented by equation (3) where $\left\|A_{zk}\right\|_F\leq A_M$, $\left\|B_{zk}\right\|_F\leq B_M$ and $\left\|\bullet\right\|_F$ denote the Frobenius norm. Given NCS under TCP with a unique equilibrium point, $z=0$, on a set $\mathbf{\Omega}$, minimizing the stochastic value function $V_k(z)$ (5) renders the stochastic optimal control input as $u_k^{c*}=-K_k z_k$ where K_k being the optimal gain. According to [7], the stochastic value function can be rewritten as

$$V^*(z_k)=E[z_k^T P_k z_k|\psi_k] \tag{35}$$

with $P_k \geq 0$ is the solution of the SRE [5]. Then we can define the optimal action dependent value function in terms of expected value as

$$\begin{aligned} V^*(z_k) &= E\{[r(z_k,u_k)+V^*(z_{k+1})]|\psi_k\} \\ &= E\{[z_k^T \ u_k^T]\Theta_k[z_k^T \ u_k^T]^T|\psi_k\} \end{aligned} \tag{36}$$

with cost to go defined as $r(z_k,u_k)=z_k^T O z_k + u_k^T R u_k$. Then, similar to [18], using Bellman equation and stochastic value function, substituting value function into the Bellman equation results in

$$\begin{bmatrix} z_k \\ u_k \end{bmatrix}^T E(\Theta_k|\psi_k)\begin{bmatrix} z_k \\ u_k \end{bmatrix} = E\{[r(z_k,u_k)+V^*(z_{k+1})]|\psi_k\}$$

$$= \begin{bmatrix} z_k \\ u_k \end{bmatrix}^T \begin{bmatrix} O+E(A_{zk}^T P_{k+1} A_{zk}|\psi_k) & E(A_{zk}^T P_{k+1} B_{zk}|\psi_k) \\ E(B_{zk}^T P_{k+1} A_{zk}|\psi_k) & R+E(B_{zk}^T P_{k+1} B_{zk}|\psi_k) \end{bmatrix} \begin{bmatrix} z_k \\ u_k \end{bmatrix} \tag{37}$$

Therefore, $E(\Theta_k|\psi_k)$ can be expressed as

$$\overline{\Theta}_k = E(\Theta_k|\psi_k) = \begin{bmatrix} \overline{\Theta}_k^{zz} & \overline{\Theta}_k^{zu} \\ \overline{\Theta}_k^{uz} & \overline{\Theta}_k^{uu} \end{bmatrix} = \begin{bmatrix} O+E(A_{zk}^T P_{k+1} A_{zk}|\psi_k) & E(A_{zk}^T P_{k+1} B_{zk}|\psi_k) \\ E(B_{zk}^T P_{k+1} A_{zk}|\psi_k) & R+E(B_{zk}^T P_{k+1} B_{zk}|\psi_k) \end{bmatrix}$$

(38)

Then, according to [5] and (38), the optimal control gain for NCS under TCP can be represented in terms of value function parameters, $\overline{\Theta}_k$, as

$$K_k = [R+E(B_{zk}^T P_{k+1} B_{zk}|\psi_k)]^{-1} E(B_{zk}^T P_{k+1} A_{zk}|\psi_k) = (\overline{\Theta}_k^{uu})^{-1}\overline{\Theta}_k^{uz} \tag{39}$$

According to optimality [7], since K_k is the optimal control gain, $\|K_k\|$ is bounded for all k (i.e. $\|K_k\| \leq K_M$).

It is important to note that even if the kernel matrix P_k is known; solving time-varying optimal control gain still requires slowly time-varying system matrices.

However, if the parameter vector $\overline{\Theta}_k$ which is slowly-varying can be estimated online, then system dynamics are not needed to calculate the optimal control gain.

C. Model-Free Online Tuning Value Function Estimator

In this section, when the value function (35) is estimated online, matrix $\overline{\Theta}_k$ is obtained which in turn is used to derive stochastic optimal control inputs via (39) without the system matrices. By assuming the value function, $V^*(z_k)$, can be represented as the linear in the unknown parameters (LIP) and according to [5] and (35), the value function is given as

$$V^*(z_k) = \varphi_k^T \overline{\Theta}_k \varphi_k = \overline{\theta}_k^T \overline{\varphi}_k \tag{40}$$

with $\overline{\theta}_k = vec(\overline{\Theta}_k)$, $\varphi_k = [z_k^T \; u(z_k^T)]^T \in \mathbb{R}^{n+bm=l}$ and $\overline{\varphi}_k = ($ $\varphi_{k1}^2,...,\varphi_{k1}\varphi_{kl},\varphi_{k2}^2,...,\varphi_{kl-1}\varphi_{kl},\varphi_{kl}^2)$ is Kronecker product quadratic polynomial stochastic independent basis vector [17] consisting of current state and past control inputs, $vec(\bullet)$ function is constructed by stacking the columns of matrix into one column vector with off-diagonal elements [17]. Since matrix $\overline{\Theta}_k$ can be considered as slowly time varying, value function can be represented as a function of target unknown parameter vector and regression function $\overline{\varphi}_k$.

Meanwhile, value function can also be represented in terms of $\overline{\Theta}_k$ as

$$V^*(z_k) = \varphi_k^T \overline{\Theta}_k \varphi_k = \overline{\theta}_k^T \overline{\varphi}_k \tag{41}$$

Since the observed system states $\hat{z}_{k|k}$ are only available at the controller, the value function with observed system states can be expressed as

$$V_k^*(\hat{z}) = \varphi_k^{eT} \overline{\Theta}_k \varphi_k^e = \overline{\theta}_k^T \overline{\varphi}_k^e \tag{42}$$

where $\varphi_k^e = [\hat{z}_{k|k}^T \; u_k^T]^T$ and $\overline{\varphi}_k^e$ is Kronecker product quadratic polynomial stochastic independent basis vector of φ_k^e.

Next, we will derive the residual errors by using the Bellman equation. Normally, the Bellman equation can be rewritten as $V_{k+1}^*(z) - V_k^*(z) + r(z_k, u_k) = 0$. However, this relationship does not hold when we apply the observed states $\hat{z}_{k|k}$. Hence, substituting observed system states into the Bellman equation, the residual errors $V^*(\hat{z}_{k|k}) - V^*(\hat{z}_{k-1|k-1})$ $+ r(\hat{z}_{k-1|k-1}, u_{k-1}) = e_k^o$. It is important to note e_k^o is caused by observer error dynamics $e_{k|k}$ and can be derived as

$$
\begin{aligned}
e_k^o &= -\tilde{z}_{k-1|k-1}^T P_{k-1} \tilde{z}_{k-1|k-1} - 2\tilde{z}_{k-1|k-1}^T P_{k-1} z_{k-1} + 2e_{k|k}^T P_k z_k - \tilde{z}_{k|k}^T P_k \tilde{z}_{k|k} + 2\tilde{z}_{k-1|k-1}^T Q z_{k-1} - \tilde{z}_{k-1|k-1}^T Q \tilde{z}_{k-1|k-1} \\
&+ 2(A_{zk}\tilde{z}_{k-1|k-1} - \tilde{z}_{k|k})^T P_k (A_{zk} - B_{zk} K_k)\hat{z}_{k-1} + (A_{zk}\tilde{z}_{k-1|k-1} - \tilde{z}_{k|k})^T P_k (A_{zk}\tilde{z}_{k-1|k-1} - \tilde{z}_{k|k}) \\
&= -\tilde{z}_{k-1|k-1}^T P_{k-1} \tilde{z}_{k-1|k-1} - 2\tilde{z}_{k-1|k-1}^T P_{k-1} \hat{z}_{k-1} + 2\tilde{z}_{k|k}^T P_k z_k - \tilde{z}_{k|k}^T P_k \tilde{z}_{k|k} + 2\tilde{z}_{k-1|k-1}^T Q \hat{z}_{k-1} + \tilde{z}_{k-1|k-1}^T Q \tilde{z}_{k-1|k-1} \\
&+ 2\tilde{z}_{k-1|k-1}^T A_{zk}^T P_k (A_{zk} - B_{zk} K_k)\hat{z}_{k-1} - 2\tilde{z}_{k|k}^T P_k z_k + 2\tilde{z}_{k|k}^T P A_{zk} \tilde{z}_{k-1|k-1} + \tilde{z}_{k-1|k-1}^T A_{zk}^T P_k A_{zk} \tilde{z}_{k-1|k-1} \\
&- 2\tilde{z}_{k|k}^T P A_{zk} \tilde{z}_{k-1|k-1} + \tilde{z}_{k|k}^T P_k \tilde{z}_{k|k} \\
&= -2\tilde{z}_{k-1|k-1}^T (P_{k-1} - Q - A_{zk}^T P_k (A_{zk} - B_{zk} K_k))\hat{z}_{k-1} + \tilde{z}_{k-1|k-1}^T (Q - P_{k-1} + A_{zk}^T P_k A_{zk})\tilde{z}_{k-1|k-1} \\
&= -2\tilde{z}_{k-1|k-1}^T (P_{k-1} - Q - A_{zk}^T P_k A_{zk} + A_{zk}^T P_k B_{zk} (B_{zk}^T P_k B_{zk} + R)^{-1} B_{zk}^T P_k A_{zk}))\hat{z}_{k-1} \\
&+ \tilde{z}_{k-1|k-1}^T (Q - P_{k-1} + A_{zk}^T P_k A_{zk})\tilde{z}_{k-1|k-1} \\
&= -2\tilde{z}_{k-1|k-1}^T (P_{k-1} - P_{k-1})\hat{z}_{k-1} + \tilde{z}_{k-1|k-1}^T (Q - P_{k-1} + A_{zk}^T P_k A_{zk})\tilde{z}_{k-1|k-1} \\
&= \tilde{z}_{k-1|k-1}^T (Q - P_{k-1} + A_{zk}^T P_k A_{zk})\tilde{z}_{k-1|k-1}
\end{aligned} \tag{43}
$$

In other words,

$$
(e_k^o)^2 \le (\tilde{z}_{k-1|k-1}^T O \tilde{z}_{k-1|k-1})^2 \tag{44}
$$

Since P_k is the solution to the SRE, $\|P\|_k \le P_M$ and $O = \|Q\| + P_M A_M^2$ which will be used in the proof of Theorem 2.

Next, the value function estimation with observed system states will be considered. First, value function with observed system states $V_k^*(\hat{z})$ can be expressed in terms of estimated parameters $\hat{\bar{\theta}}_k$ as

$$\hat{V}(\hat{z}_{k|k}) = \varphi_k^{eT} \hat{\bar{\Theta}}_k \varphi_k^e = \hat{\bar{\theta}}_k^T \bar{\varphi}_k^e \tag{45}$$

where $\hat{\bar{\theta}}_k$ is the estimated value of the target parameter vector $\bar{\theta}_k$ defined before.

Substituting (45) and estimated system states into Bellman equation, the equation is not guaranteed to hold. By using delayed values for convenience, the residual error associated with (45) can be expressed as

$$e_k^a + e_k^o = \hat{V}(\hat{z}_{k|k}) - \hat{V}(\hat{z}_{k-1|k-1}) + r(\hat{z}_{k-1|k-1}, u_{k-1}) = r(\hat{z}_{k-1|k-1}, u_{k-1}) + \hat{\bar{\theta}}_k^T \Delta\bar{\varphi}_{k-1}^e \tag{46}$$

where $\Delta\bar{\varphi}_k^e = \bar{\varphi}_k^e - \bar{\varphi}_{k-1}^e$ is the first difference of regression function and e_k^a represent estimation errors.

The dynamics of (46) can be expressed as $e_{k+1}^a + e_{k+1}^o = r(\hat{z}_{k|k}, u_k) + \hat{\bar{\theta}}_{k+1}^T \Delta\bar{\varphi}_k^e$. Next define the update law for the parameter vector $\hat{\bar{\theta}}_k$ as

$$\hat{\bar{\theta}}_{k+1} = \hat{\bar{\theta}}_k + \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^o + e_k^a)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1} \tag{47}$$

where $0 < \alpha_h < 1$ is the tuning parameter for value function estimation.

Remark 3: It is observed that the value function $V^*(z_k)$ and its estimation $\hat{V}(\hat{z}_{k|k})$ (45) will become zero only when $z_k = 0$ and $\hat{z}_{k|k} = 0$. Hence when system states have converged to zero, the estimated system states $\hat{z}_{k|k}$ will also converge to zero according to Theorem 2 and the value function estimation is no longer updated. It can be seen as a persistency of excitation (PE) requirement for the inputs to value function estimator wherein the system states must be persistently existing long enough for the estimator to learn the optimal value function. Therefore, exploration noise is added to NCS in order to satisfy the PE condition [18].

Next dynamics of parameter estimation errors of the value function can be expressed as

$$\tilde{\tilde{\theta}}_{k+1} = \tilde{\tilde{\theta}}_k - \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^o + e_k^a)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1} \tag{48}$$

Then, the convergence of the value function estimation errors with parameter error dynamics $\tilde{\tilde{\theta}}_k$ given by (48) is demonstrated for an initial admissible control [10] policy. It is important to note that slowly time-varying [21] linear NCS is asymptotically stable in the mean if an initial admissible control can be implemented provided system matrices are known. However, proposed estimated value function with observed system states results in estimation errors for the value function V_k, whose stability needs to be studied. Therefore, Theorem 2 will prove the value function estimation errors converge while the overall closed-loop system stability is shown in Theorem 3 with an initial admissible control policy.

Theorem 2: *(Asymptotic stability of the value function estimation errors)*. Given the initial parameter vector $\hat{\bar{\theta}}_0$ for the value function estimator to be bounded in the set $\mathbf{\Omega}$, let u_{0k} be an initial admissible control policy for the linear NCS under TCP (3). Let the observer parameter update law be given by (47). Then there exists positive constants α_h, α_o and η satisfying $0 < \alpha_h < \dfrac{2\Delta\phi_{\min}^2}{3(\Delta\phi_{\min}^2 + 1)}$ with $0 < \Delta\phi_{\min} < \left\|\Delta\bar{\varphi}_k^e\right\|$,and $\Delta\phi_{\min}$ is the lower bound of $\left\|\Delta\bar{\varphi}_k^e\right\|$, $0<\eta<\sqrt[4]{\dfrac{1}{2(4+276\alpha_o(1+6\alpha_o+9\alpha_o^2))}}$, $\dfrac{\chi_{\min}^2+1}{72}<\alpha_o<\dfrac{\chi_{\min}^2+1}{36}$ and such that the value function estimation errors for NCS under TCP converge to zero asymptotically, where $\chi_{\min}^2<\left\|\Gamma_{k+1}M_k\hat{s}_k\right\|^2$.

Proof: Consider the positive definite Lyapunov candidate

$$L_J = L_\theta + L_{ao} \tag{49}$$

where $L_\theta = tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\}$ and $L_{ao} = 276tr\{\sum_{i=k-D_0}^{k}\vartheta_{i-1}^T\vartheta_{i-1}\vartheta_{i-1}^T\vartheta_{i-1}\} + tr\{\sum_{i=k-D_0}^{k}\tilde{z}_{i-1|i-1}\tilde{z}_{i-1|i-1}^T\Lambda\tilde{z}_{i-1|i-1}\tilde{z}_{i-1|i-1}^T\}$ with $\Lambda = \dfrac{1}{(U_M-1)^4(\Psi_M^2+1)^2}\mathbf{I}$ and $\Pi = \dfrac{(\Delta\phi_{\min}^2+1)}{4\alpha_h^2 O^2(\chi_{\min}^2+1)^2(U_M-1)^4}\mathbf{I}$ and are positive definite matrices, $\mathbf{I}$ is identity matrix and $\forall k=1,2,\ldots$. The first difference of (49) is given by $\Delta L_J = \Delta L_\theta + \Delta L_{ao}$. Since Lyapunov candidate function (49) includes observer parameters, we have to separate the proof into two cases similar to Theorem 1.

Case 1: $\gamma_{k+1}=1$. (No packet losses)

Using (27), (28) and (48), ΔL_J can be derived as

$$\begin{aligned}
\Delta L_J &= \Delta L_\theta + \Delta L_{ao} \\
&= tr\{(\tilde{\tilde{\theta}}_k - \alpha_h\frac{\Delta\bar{\varphi}_k^e(e_k^a+e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e+1})^T \Pi(\tilde{\tilde{\theta}}_k - \alpha_h\frac{\Delta\bar{\varphi}_k^e(e_k^a+e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e+1})\} - tr\{\tilde{\tilde{\theta}}_k^T\Pi\tilde{\tilde{\theta}}_k\} + \Lambda\left\|\tilde{z}_{k|k}\right\|^4 \\
&\quad - \Lambda\left\|\tilde{z}_{k-1|k-1}\right\|^4 + 276\left\|\tilde{\vartheta}_k\right\|^4 - 276\left\|\tilde{\vartheta}_{k-1}\right\|^4
\end{aligned} \tag{50}$$

$$\le tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} - \alpha_h tr\{(\frac{(e_k^a + e_k^o)\Delta\bar{\varphi}_k^{eT}}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\Pi\tilde{\tilde{\theta}}_k\} - \alpha_h tr\{\tilde{\tilde{h}}_k^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\}$$

$$+ \alpha_h^2 tr\{(\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\} - tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + 4\Lambda\left\|U_z - \kappa_o (D_y^o)^+\right\|^4 \left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$+ 4\Lambda\left\|\tilde{\vartheta}_{k-1}\right\|^4 \left\|M_{k-1}\hat{s}_{k-1}\right\|^4 - \Lambda\left\|\tilde{z}_{k-1|k-1}\right\|^4 - 276\alpha_o(1 - 12\alpha_o - \frac{6\alpha_o}{\chi_{\min}^2 + 1} - 3\alpha_o^2)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$+ 276\frac{\alpha_o(1 + 6\alpha_o + 9\alpha_o^2)}{(\chi_{\min}^2 + 1)^2}\left\|U_z - \kappa_o (D_y^o)^+\right\|^4 \left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$\le -2\alpha_h tr\{\tilde{\tilde{\theta}}_k^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\} + \alpha_h^2 tr\{(\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\}$$

$$- \frac{1}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^4}(1 - 4\eta^4 - 276(U_M - 1)^4 - 4(70\alpha_o(1 - 12\alpha_o - \frac{6\alpha_o}{\chi_{\min}^2 + 1} - 3\alpha_o^2) - 1)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$\times \alpha_o(1 + 6\alpha_o + 9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$\le -2\alpha_h tr\{\frac{(\tilde{\tilde{\theta}}_k^T \Delta\bar{\varphi}_k^e + e_k^o)\Pi\Delta\bar{\varphi}_k^{eT}\tilde{\tilde{\theta}}_k}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\} + \alpha_h^2 tr\{\frac{(\tilde{\tilde{\theta}}_k^T \Delta\bar{\varphi}_k^e + e_k^o)\Delta\bar{\varphi}_k^{eT}\Pi\Delta\bar{\varphi}_k^e (e_k^{oT} + \Delta\bar{\varphi}_k^{eT}\tilde{\tilde{\theta}}_k)}{(\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1)^2}\}$$

$$- 4(70\alpha_o(1 - 12\alpha_o - \frac{6\alpha_o}{\sigma_0^2 + 1} - 3\alpha_o^2) - 1)\left\|\tilde{\vartheta}_{k-1}\right\|^4 - \frac{1}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^4}$$

$$\times (1 - 4\eta^4 - 276(U_M - 1)^4 \alpha_o(1 + 6\alpha_o + 9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$+ 2\alpha_k^2 \Pi tr\{\frac{(e_k^o)^2}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\} \le -2\alpha_h tr\{\tilde{\tilde{\theta}}_k^T \Pi\tilde{\tilde{\theta}}_k\} - 2\alpha_h tr\{\frac{e_k^o \Pi\Delta\bar{\varphi}_k^{eT}\tilde{\tilde{\theta}}_k}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\} + 2\alpha_h^2 tr\{\tilde{\tilde{\theta}}_k^T \Pi\tilde{\tilde{\theta}}_k\}$$

$$+ 2\alpha_h tr\{\frac{\tilde{\tilde{\theta}}_k^T \Pi\tilde{\tilde{\theta}}_k}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\} \quad -4(70\alpha_o(1 - 12\alpha_o - \frac{6\alpha_o}{\chi_{\min}^2 + 1} - 3\alpha_o^2) - 1)\left\|\tilde{\vartheta}_{k-1}\right\|^4 - \frac{1}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^2}$$

$$\times (1 - 4\eta^4 - 276(U_M - 1)^2 \alpha_o(1 + 6\alpha_o + 9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$\le -2\alpha_h tr\{\tilde{\tilde{\theta}}_k^T \Pi\tilde{\tilde{\theta}}_k\} - 2\alpha_h tr\{\frac{e_k^o \Pi\Delta\varphi_k^{eT}\tilde{\tilde{\theta}}_k}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\} + 2\alpha_h^2 tr\{\tilde{\tilde{\theta}}_k^T \Pi\tilde{\tilde{\theta}}_k\} + 2\alpha_h^2 tr\{\frac{(e_k^o)^2}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\}$$

$$+ 2\alpha_h tr\{\frac{\tilde{\tilde{\theta}}_k^T \Pi\tilde{\tilde{\theta}}_k}{\Delta\phi_{\min}^2 + 1}\} - 4(70\alpha_o(1 - 12\alpha_o - \frac{6\alpha_o}{\sigma_0^2 + 1} - 3\alpha_o^2) - 1)\left\|\tilde{\vartheta}_{k-1}\right\|^4 - \frac{1}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^2}$$

$$\times (1 - 4\eta^4 - 276(U_M - 1)^2 \alpha_o(1 + 6\alpha_o + 9\alpha_o^2)\eta^4)\left\|e_{k-1|k-1}\right\|^4$$

According to Cauchy-Schwartz inequality, the term $-2\alpha_h tr\{\frac{e_k^o \Pi\Delta\bar{\varphi}_k^{eT}\tilde{\tilde{\theta}}_k}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\}$ can be derived as

$$-2\alpha_h tr\{\frac{e_k^o \Pi \Delta\overline{\varphi}_k^{eT} \tilde{\tilde{\theta}}_k}{\Delta\overline{\varphi}_k^{eT}\Delta\overline{\varphi}_k^e+1}\} \le \alpha_h \Pi tr\{\frac{(e_k^o)^2}{\Delta\overline{\varphi}_k^{eT}\Delta\overline{\varphi}_k^e+1}\} + \alpha_h tr\{\frac{\tilde{\tilde{\theta}}_k^T \Delta\overline{\varphi}_k^e \Pi \Delta\overline{\varphi}_k^{eT} \tilde{\tilde{\theta}}_k}{\Delta\overline{\varphi}_k^{eT}\Delta\overline{\varphi}_k^e+1}\}$$

$$\le \alpha_h \Pi tr\{\frac{(e_k^o)^2}{\Delta\phi_{\min}^2+1}\} + \alpha_h tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} \tag{51}$$

Substituting (51) into (50), ΔL_J can be rewritten as

$$\begin{aligned}
\Delta L_J &= \Delta L_\theta + \Delta L_{ao} \\
&\le -2\alpha_h tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + \frac{1}{\Delta\phi_{\min}^2+1}\Pi (e_k^o)^2 + \alpha_h^2 tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + 2\alpha_h^2 tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\}
\end{aligned} \tag{52}$$

$$+2\alpha_h^2 \Pi tr\{\frac{(e_k^o)^2}{\Delta\phi_{\min}^2+1}\} + 2\alpha_h tr\{\frac{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k}{\Delta\phi_{\min}^2+1}\} - 4(70\alpha_o(1-12\alpha_o - \frac{6\alpha_o}{\chi_{\min}^2+1} - 3\alpha_o^2)-1)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$-\frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^2}(1-4\eta^4-276(U_M-1)^2\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$\le -\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + \frac{2\alpha_h^2 O^2 \Pi}{\Delta\phi_{\min}^2+1}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$-4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\chi_{\min}^2+1}-3\alpha_o^2)-1)\left\|\tilde{\vartheta}_{k-1}\right\|^4 - \frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^2}$$
$$\times(1-4\eta^4-276(U_M-1)^2\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$\le -\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\left\|\tilde{\tilde{\theta}}_k\right\|^2 + \frac{1}{2(\chi_{\min}^2+1)^2(U_M-1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4 - 4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\chi_{\min}^2+1}$$

$$-3\alpha_o^2)-1)\left\|\tilde{\vartheta}_{k-1}\right\|^4 - \frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4}(1-4\eta^4-276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$\le -\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\left\|\tilde{\tilde{\theta}}_k\right\|^2 - 4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\chi_{\min}^2+1}-3\alpha_o^2)-1)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$-\frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4}(\frac{1}{2}-4\eta^4-276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

Case 2: $\gamma_{k+1}=0$. (With packet lost)

Based on Assumption 2, packets lost probability between sensor and controller has to satisfy the observer stability region (i.e. $P(\gamma_k = 1) > \frac{1}{N_0}$). Therefore, if $\gamma_{k+1} = 0$, then there exists $j \in [k - N_0, k]$ such that $\gamma_{k+1} = \gamma_k = \cdots = \gamma_{k-j+1} = 0$ and $\gamma_{k-j} = 1$. Therefore, using update law for the estimated parameters of observer (25), the observer parameter estimation errors $\tilde{\vartheta}_{k+1}$ can be expressed as $\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_{k-j}$. Then substituting $\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_{k-j}$ into ΔL_J, we have

$$\begin{aligned}
\Delta L_J &= \Delta L_\theta + \Delta L_{ao} \\
&= tr\{(\tilde{\tilde{\theta}}_k - \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})^T \Pi (\tilde{\tilde{\theta}}_k - \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})\} - tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + \Lambda \sum_{i=k-N_0}^{k} \left\| \tilde{z}_{i|i} \right\|^4 \\
&- \Lambda \sum_{i=k-N_0}^{k} \left\| \tilde{z}_{i-1|i-1} \right\|^4 + 276 \sum_{i=k-N_0}^{k} \left\| \tilde{\vartheta}_i \right\|^4 - 276 \sum_{i=k-N_0}^{k} \left\| \tilde{\vartheta}_{i-1} \right\|^4 \\
&\le -\alpha_h (2 - 3\alpha_h - \frac{2}{\Delta\phi_{\min}^2 + 1}) \Pi \left\| \tilde{\tilde{\theta}}_k \right\|^2 - 4(70\alpha_o (1 - 12\alpha_o - \frac{6\alpha_o}{\chi_{\min}^2 + 1} - 3\alpha_o^2) - 1) \left\| \tilde{\vartheta}_{k-j-1} \right\|^4 \\
&- \frac{1}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^4} (\frac{1}{2} - 4\eta^4 - 276(U_M - 1)^4 \alpha_o (1 + 6\alpha_o + 9\alpha_o^2)\eta^4) \sum_{i=k-N_0}^{k} \left\| \tilde{z}_{i-1|i-1} \right\|^4
\end{aligned} \tag{53}$$

Since $0 < \alpha_h < \frac{2\Delta\phi_{\min}^2}{3(\Delta\phi_{\min}^2 + 1)}$, $\frac{\chi_{\min}^2 + 1}{72} < \alpha_o < \frac{\chi_{\min}^2 + 1}{36}$ and $0 < \eta < \sqrt[4]{\frac{1}{2(4 + 276(U_M - 1)^4 \alpha_o (1 + 6\alpha_o + 9\alpha_o^2))}}$, ΔL_J is negative definite and L_J is positive definite. Therefore, for the two cases, the observed system states $\hat{z}_{k|k}$, value function parameter estimation errors $\tilde{\tilde{\theta}}_k$ and value function estimation errors for NCS under TCP are all asymptotically stable in the mean. In other words, as $k \to \infty$, $\tilde{\tilde{\theta}}_k \to 0$, $\tilde{z}_{k|k} \to 0$, and $\hat{V}(\hat{z}_{k|k}) \to V^*(z_k)$.

Next, it is shown estimated control input based on estimated parameter vector $\hat{\overline{\Theta}}_k$ will indeed converge to the optimal control input.

D. Stochastic Optimal Control Signal Estimation for NCS under TCP

Similar to [18], there are two ways to estimate stochastic optimal control signal for linear NCS under TCP. One is based on slowly time-varying matrix $\overline{\Theta}_k$, and another is based on standard optimal control theory by minimizing the stochastic value function. The main differences are that the later method needs the system dynamics and solves the optimal controller backward-in-time. However, it is shown here that ultimately both are equivalent and therefore are used in the proofs.

Method I: Slowly time-varying matrix $\overline{\Theta}_k$ can be approximated by using proposed value function estimation. Using (39), the estimated optimal control signal for NCS can be expressed in terms of estimated $\overline{\Theta}_k$ as

$$\hat{u}_{1k} = -\hat{K}_k \hat{z}_{k|k} = -(\hat{\overline{\Theta}}_k^{uu})^{-1} \hat{\overline{\Theta}}_k^{uz} \hat{z}_{k|k} \tag{54}$$

Method II: Alternatively, estimated optimal control signal that minimizes the estimated stochastic value function (41) with actual parameters $\hat{\overline{\Theta}}_k$ is given by

$$\hat{u}_{2k} = -\frac{1}{2} R^{-1} E\left(B_{zk}^T \frac{\partial \hat{V}(\hat{z}_{k+1|k+1})}{\partial \hat{z}_{k+1|k+1}} \middle| \psi_k \right) \tag{55}$$

where $\hat{V}(\hat{z}_{k|k}) = E(\varphi_k^{eT} \overline{\Theta}_k \varphi_k^e | \psi_k)$.

Next, it will be shown that optimal control input obtained by method I and II are equivalent.

Lemma 1: The estimated optimal control obtained with the value function estimation of $V^*(z_k)$ is equivalent to the optimal control calculated by minimizing the stochastic value function V_k^*, i.e. $\hat{u}_{1k} = \hat{u}_{2k}$.

Proof: Using the Bellman equation and estimated value function with matrix $\overline{\Theta}_k$, we have

$$\hat{V}_k(\hat{z}) + e_k^o + e_k^a = r(\hat{z}_{k|k}, u(\hat{z}_{k|k})) + \hat{V}(\hat{z}_{k+1|k+1}) \tag{56}$$

Now consider (56) and

1) The left part of (56) can be expressed as

$$\hat{V}_k(\hat{z}) + e_k^o + e_k^a = \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}^T \hat{\overline{\Theta}}_k \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix} = \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}^T \begin{bmatrix} \hat{\overline{\Theta}}_k^{z} & \hat{\overline{\Theta}}_k^{zu} \\ \hat{\overline{\Theta}}_k^{uz} & \hat{\overline{\Theta}}_k^{uu} \end{bmatrix} \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix} \tag{57}$$

2) The right side of (56) can be shown as

$$\begin{aligned} r(\hat{z}_{k|k}, u(\hat{z}_{k|k})) + \hat{V}(\hat{z}_{k+1|k+1}) &= r(\hat{z}_{k|k}, u(\hat{z}_{k|k})) + E(\varphi_{k+1}^{eT} \hat{\Theta}_{k+1} \varphi_{k+1}^{e} | \psi_k) \\ &= (\hat{z}_{k|k}, u(\hat{z}_{k|k})) + E((A_{zk}\hat{z}_{k|k} + B_{zk}u_k)^T \hat{P}_{k+1}(A_{zk}\hat{z}_{k|k} + B_{zk}u_k)|\psi_k) \end{aligned} \tag{58}$$

$$= \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}^T \begin{bmatrix} O & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}$$

$$+ \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}^T \begin{bmatrix} E(A_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k) & E(A_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k) \\ E(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k) & E(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k) \end{bmatrix} \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}$$

$$= \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}^T \begin{bmatrix} E(A_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k) + O & E(A_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k) \\ E(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k) & E(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k) + R \end{bmatrix} \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}$$

According to (57) and (58), (56) can be derived as

$$\begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix}^T \begin{bmatrix} \hat{\overline{\Theta}}_k^{zz} & \hat{\overline{\Theta}}_k^{zu} \\ \hat{\overline{\Theta}}_k^{uz} & \hat{\overline{\Theta}}_k^{uu} \end{bmatrix} \begin{bmatrix} \hat{z}_{k|k} \\ u(\hat{z}_{k|k}) \end{bmatrix} + e_k^o + e_k^a$$

$$=\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}^T\begin{bmatrix}E(A_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)+O & E(A_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)\\ E(B_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k) & E(B_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)+R\end{bmatrix}\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}$$

Hence,

$$\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}^T\begin{bmatrix}\hat{\overline{\Theta}}_k^{zz} & \hat{\overline{\Theta}}_k^{zu}\\ \hat{\overline{\Theta}}_k^{uz} & \hat{\overline{\Theta}}_k^{uu}\end{bmatrix}\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}=\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}^T$$

$$\times\begin{bmatrix}E(A_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)+O & E(A_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)\\ E(B_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k) & E(B_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)+R\end{bmatrix}\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}-e_k^o-e_k^a$$

$$=\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}^T\begin{bmatrix}E(A_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)+O & E(A_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)\\ E(B_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k) & E(B_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)+R\end{bmatrix}\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}$$

$$-\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}^T\begin{bmatrix}\dfrac{e_k^o+e_k^a}{tr\{\hat{z}_{k|k}^T\hat{z}_{k|k}\}} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}$$

$$=\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}^T\begin{bmatrix}E(A_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)+O-\dfrac{e_k^o+e_k^a}{tr\{\hat{z}_{k|k}^T\hat{z}_{k|k}\}} & E(A_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)\\ E(B_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k) & E(B_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)+R\end{bmatrix}\begin{bmatrix}\hat{z}_{k|k}\\ u(\hat{z}_{k|k})\end{bmatrix}$$

And

$$\begin{bmatrix}\hat{\overline{\Theta}}_k^{z} & \hat{\overline{\Theta}}_k^{zu}\\ \hat{\overline{\Theta}}_k^{uz} & \hat{\overline{\Theta}}_k^{uu}\end{bmatrix}=\begin{bmatrix}E(A_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)+O-\dfrac{e_k^o+e_k^a}{tr\{\hat{z}_{k|k}^T\hat{z}_{k|k}\}} & E(A_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)\\ E(B_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k) & E(B_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)+R\end{bmatrix}\tag{59}$$

where

$$\begin{cases}\hat{\overline{\Theta}}_k^{z}=E(A_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)+O-\dfrac{e_k^o+e_k^a}{tr\{\hat{z}_{k|k}^T\hat{z}_{k|k}\}}\\ \hat{\overline{\Theta}}_k^{zu}=E(A_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)\\ \hat{\overline{\Theta}}_k^{uz}=E(B_{zk}^T\hat{P}_{k+1}A_{zk}|\psi_k)\psi_k\\ \hat{\overline{\Theta}}_k^{uu}=E(B_{zk}^T\hat{P}_{k+1}B_{zk}|\psi_k)+R\end{cases}\tag{60}$$

According to the estimated optimal control law (39), we have $\hat{u}_{1k} = -(\hat{\overline{\Theta}}_k^{uu})^{-1}\hat{\overline{\Theta}}_k^{uz}\hat{z}_{k|k}$, which is expressed by using (47) as

$$\hat{u}_{1k} = -[E(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k) + R]^{-1} E(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k)\hat{z}_{k|k} \tag{61}$$

At the same time, according to the optimal control theory [7] we define $\hat{V}_{k+1}(\hat{z}) = E((A_{zk}\hat{z}_{k|k} + B_{zk}u_k)^T \hat{P}_{k+1} (A_{zk}\hat{z}_{k|k} + B_{zk}u_k) | \psi_k)$. Therefore, we can minimize the stochastic value function to get the optimal control as

$$\begin{aligned}
\hat{u}_{2k} &= -\frac{1}{2} R^{-1} E\left(B_{zk}^T \frac{\partial \hat{V}(\hat{z}_{k+1|k+1})}{\partial \hat{z}_{k+1|k+1}} \middle| \psi_k \right) \\
&= -R^{-1} E\left[B_{zk}^T \hat{P}_{k+1} \left(A_{zk}\hat{z}_{k|k} + B_{zk}\hat{u}_{2k} \right) \middle| \psi_k \right] \\
&= -R^{-1} E\left(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k \right)\hat{z}_{k|k} - R^{-1} E\left(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k \right)\hat{u}_{2k}
\end{aligned} \tag{62}$$

The term $\hat{u}_{2k}$ can be solved by (62) as

$$\left(I + R^{-1} E\left(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k \right) \right)\hat{u}_{2k} = -R^{-1} E\left(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k \right)\hat{z}_{k|k}$$

$$\left(R + E\left(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k \right) \right)\hat{u}_{2k} = -E\left(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k \right)\hat{z}_{k|k}$$

$$\begin{aligned}
\hat{u}_{2k} &= -\left(R + E\left(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k \right) \right)^{-1} E\left(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k \right)\hat{z}_{k|k} \\
&= -\left[R + E\left(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k \right) \right]^{-1} E\left(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k \right)\hat{z}_{k|k}
\end{aligned} \tag{63}$$

According to (61) and (63), we have

$$\begin{aligned}
\hat{u}_{1k} &= -(\hat{\overline{\Theta}}_k^{uu})^{-1}\hat{\overline{\Theta}}_k^{uz}\hat{z}_{k|k} \\
&= -\left[R + E\left(B_{zk}^T \hat{P}_{k+1} B_{zk} | \psi_k \right) \right]^{-1} E\left(B_{zk}^T \hat{P}_{k+1} A_{zk} | \psi_k \right)\hat{z}_{k|k} = \hat{u}_{2k}
\end{aligned} \tag{64}$$

Therefore, $\tilde{u}_k = \hat{u}_{1k} - \hat{u}_{2k} = 0$ since $\hat{u}_{1k} = \hat{u}_{2k}$.

Since the equality in this lemma is both ways and noting the drawback of second method, we use method I to solve the optimal control inputs for NCS under TCP. However, we will use the Lemma 1 to complete the convergence proof because they are equivalent. Next, stability of estimated stochastic value function and control input, observation error and estimation error dynamics are considered.

E. Closed-Loop System Stability

In this section, we will show that observer errors, slowly time-varying parameter and value function estimation error dynamics are asymptotically stable in the mean. Moreover, the observed system states and estimated control inputs for NCS under TCP will converge asymptotically to actual system states and optimal control signals respectively. Next, the flowchart of proposed stochastic optimal regulator of linear NCS under TCP is shown in Figure 5.

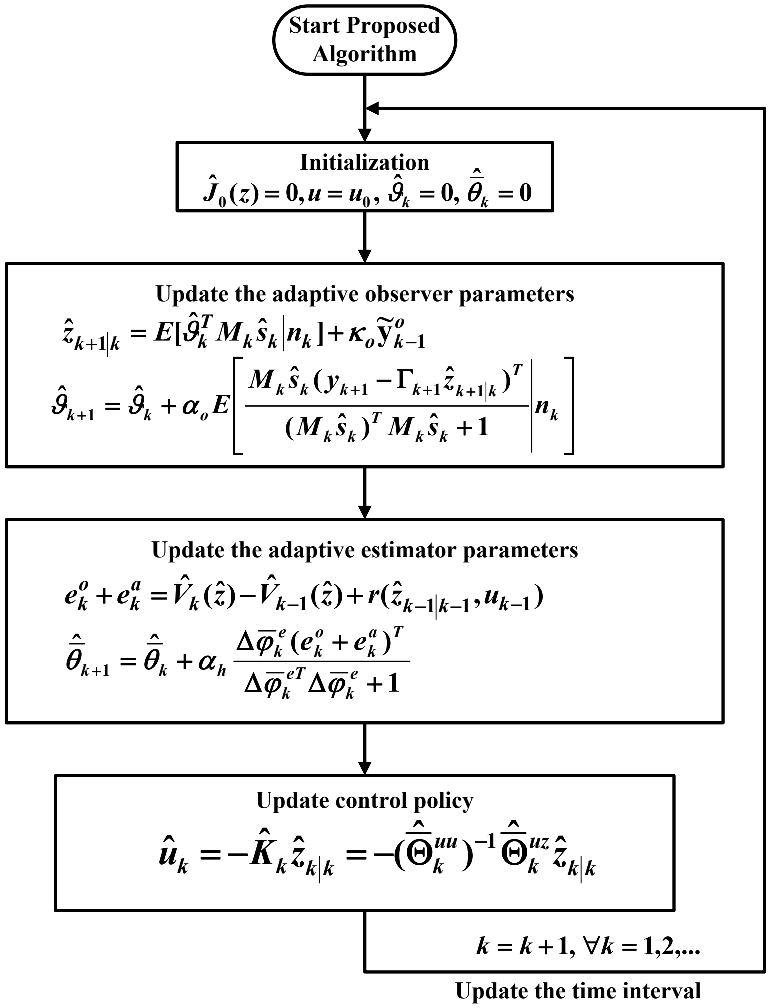


Fig 5. Stochastic optimal regulator for linear NCS under TCP

Here the initial system states are assumed to reside in the set $\mathbf{\Omega}$ stabilized by using the initial admissible control input u_0. Then, sufficient conditions for the observer and value function estimator tuning gains α_o, α_h are derived to guarantee all the future states will approach to zero asymptotically. Eventually, it can be shown that actual control inputs converge to the optimal control asymptotically.

Before introducing the convergence proof, the following lemma is needed to establish the bounds on the optimal closed-loop dynamics while the optimal control is implemented on the NCS under TCP with network imperfections.

Lemma 2: Consider the NCS under TCP, and then there exists an admissible control policy such that the following inequality is satisfied

$$\left\| A_{zk} z_k + B_{zk} u_k \right\|^2 \leq k_a \left\| z_k \right\|^2 \tag{65}$$

where $0 < k_a < 1/2$ is a constant.

Proof: Proof follows similar to [18] but omitted.

Theorem 3 *(Convergence of Optimal Control Inputs):* Given the initial conditions for the system state z_0, observer parameter estimation vector $\hat{\bar{\vartheta}}_0$, value function parameter vector $\hat{\bar{\theta}}_0$ are bounded in the set $\mathbf{\Omega}$, let u_0 be any initial admissible control policy for NCS under TCP in the presence of network imperfections satisfying the bounds given by (65) for $0 < k_a < 1/2$. Let the observer, value function estimated parameters be tuned and the estimated control policy be provided by (26), (47) and (54) respectively. Then, there exist positive constants α_o, η given by Theorem 1 and 2, and α_h given by Theorem 2 such that the system states z_k, observer parameter estimation error vector $\tilde{\bar{\vartheta}}_k$ and value function parameter estimation error vector $\tilde{\bar{\theta}}_k$ for NCS under TCP are all asymptotically stable in the mean. In other words, as $k \to \infty$, $z_k \to 0$, $\tilde{\bar{\vartheta}}_k \to 0$, $\tilde{z}_{k|k} \to 0$, $\tilde{\bar{\theta}}_k \to 0$, $\hat{V}(\hat{z}_{k|k}) \to V^*(z_k)$ and $u_{1k} \to u_k^*, u_{2k} \to u_k^*$.

Proof: Consider the following positive definite Lyapunov function candidate

$$L = L_D + L_J + L_o \tag{66}$$

where L_D is defined as $L_D = tr\{z_k^T \Omega z_k\}$ with $\Omega = \dfrac{\mathbf{I}}{8B_M^2 K_M^2(\chi_{\min}^2 + 1)}$ is positive definite matrix, L_J (32) with positive matrix Ξ is defined in Theorem 1, L_o (49) with positive matrices Λ, Π are given by Theorem 2 and $\mathbf{I}$ is identity matrix. The first difference of (66) can be expressed as $\Delta L = \Delta L_D + \Delta L_J + \Delta L_o$. Consider the first part $\Delta L_D = tr\{z_{k+1}^T \Omega z_{k+1}\} - tr\{z_k^T \Omega z_k\}$ and by applying NCS under TCP and Cauchy-Schwartz inequality, we have

$$\begin{aligned} \Delta L_D &= tr\{z_{k+1}^T \Omega z_{k+1}\} - tr\{z_k^T \Omega z_k\} \\ &\le \Omega \left\| A_{zk} z_k + B_{zk} u_k + B_{zk} K_k \tilde{z}_{k|k} - B_{zk} \tilde{u}_k \right\|^2 - \Omega \left\| z_k \right\|^2 \end{aligned} \tag{67}$$

$$\le 2\Omega \left\| A_{zk} z_k + B_{zk} u_k \right\|^2 + 4\Omega \left\| B_{zk} K_k \tilde{z}_{k|k} \right\|^2 + 4\Omega \left\| B_{zk} \tilde{u}_k \right\|^2 - \Omega \left\| z_k \right\|^2$$

Using the Lemma 2 and recalling $\hat{u}_{1k} = \hat{u}_{2k}$ from Lemma 1, we know

$$\tilde{u}_k = \hat{u}_{1k} - \hat{u}_{2k} = -(\hat{\Theta}_k^{uu})^{-1} \hat{\Theta}_k^{uz} \hat{z}_{k|k} + \frac{1}{2} R^{-1} E\left(B_{zk}^T \frac{\partial \hat{V}(\hat{z}_{k+1|k+1})}{\partial \hat{z}_{k+1|k+1}} \middle| \psi_k \right) = 0 \tag{68}$$

Therefore, (67) can be expressed as

$$\begin{aligned} \Delta L_D &\le -(1 - 2k_a)\Omega \left\| z_k \right\|^2 + 4\Omega \left\| B_{zk} \tilde{u}_k \right\|^2 + 4\Omega \left\| B_{zk} K_k \tilde{z}_{k|k} \right\|^2 + 9\Omega \left\| \tilde{z}_{k|k} \right\|^2 \\ &\le -(1 - 2k_a)\Omega \left\| z_k \right\|^2 + 4\Omega \left\| B_{zk} \tilde{u}_k \right\|^2 + 4B_M^2 K_M^2 \Omega \left\| \tilde{z}_{k|k} \right\|^2 \\ &\le -(1 - 2k_a)\Omega \left\| z_k \right\|^2 + 4\Omega B_M^2 \left\| -(\hat{\Theta}_k^{uu})^{-1} \hat{\Theta}_k^{uz} \hat{z}_{k|k} + \frac{1}{2} R^{-1} E\left(B_{zk}^T \frac{\partial \hat{V}(\hat{z}_{k+1|k+1})}{\partial \hat{z}_{k+1|k+1}} \middle| \psi_k \right) \right\|^2 \\ &\quad + 4B_M^2 K_M^2 \Omega \left\| \tilde{z}_{k|k} \right\|^2 \\ &\le -(1 - 2k_a)\Omega \left\| z_k \right\|^2 + 4B_M^2 K_M^2 \Omega \left\| \tilde{z}_{k|k} \right\|^2 \end{aligned} \tag{69}$$

Similar to Theorem 1 and 2, we separate the proof into two cases: $\gamma_{k+1}=1$ and $\gamma_{k+1}=0$ as following

Case 1: $\gamma_{k+1}=1$. (No packet losses)

Combining (33), (52) and (69), ΔL can be derived as

$$\begin{aligned}
\Delta L &= \Delta L_D + \Delta L_J + \Delta L_o \\
&\le -(1-2k_a)\Omega\|z_k\|^2 + 4B_M^2K_M^2\Omega\|\tilde{z}_{k|k}\|^2 \\
&- \alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\|\tilde{\tilde{\theta}}_k\|^2 - (16\alpha_o(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\alpha_o)-2)\|\tilde{\vartheta}_k\|^2 \\
&- \frac{(1-\eta^2(2+16\alpha_o+32\alpha_o^2))}{\sigma_0^2+1}\|\tilde{z}_{k|k}\|^2 - 4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\sigma_0^2+1} \\
&- 3\alpha_o^2)-1)\|\tilde{\vartheta}_{k-1}\|^4 - \frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4}(\frac{1}{2}-4\eta^4-276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\|\tilde{z}_{k-1|k-1}\|^4 \\
&\le -(1-2k_a)\Omega\|z_k\|^2 - \frac{(1-2\eta^2(U_M-1)^2(2+16\alpha_o+32\alpha_o^2))}{2(\chi_{\min}^2+1)(U_M-1)^2}\|\tilde{z}_{k|k}\|^2 \\
&- \alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\|\tilde{\tilde{\theta}}_k\|^2 - (16\alpha_o(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\alpha_o)-2)\|\tilde{\vartheta}_k\|^2 \\
&- 4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\chi_{\min}^2+1}-3\alpha_o^2)-1)\|\tilde{\vartheta}_{k-1}\|^4 - \frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4} \\
&\times(\frac{1}{2}-4\eta^4-276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\|\tilde{z}_{k-1|k-1}\|^4
\end{aligned} \tag{70}$$

Case 2: $\gamma_{k+1}=0$. (With packet losses) After applying Assumption 2, (34), (53) and (69), we can express ΔL as

$$\begin{aligned}
\Delta L &= \Delta L_D + \Delta L_J + \Delta L_o \\
&\le -(1-2k_a)\Omega\|z_k\|^2 + 4B_M^2K_M^2\Omega\|\tilde{z}_{k|k}\|^2
\end{aligned}$$

$$-\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\left\|\tilde{\tilde{\theta}}_k\right\|^2-4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\chi_{\min}^2+1}-3\alpha_o^2)-1)\left\|\tilde{\vartheta}_{k-j-1}\right\|^4$$

$$-\frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4}(\frac{1}{2}-4\eta^4-276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i-1|i-1}\right\|^4$$

$$-\frac{(1-\eta^2(U_M-1)^2(2+16\alpha_o+32\alpha_o^2))}{(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2-(16\alpha_o(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\alpha_o)-2)\left\|\tilde{\vartheta}_{k-j-1}\right\|^2$$

$$\le-(1-2k_a)\Omega\left\|z_k\right\|^2-\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\left\|\tilde{\tilde{\theta}}_k\right\|^2 \tag{71}$$

$$-4(70\alpha_o(1-12\alpha_o-\frac{6\alpha_o}{\chi_{\min}^2+1}-3\alpha_o^2)-1)\left\|\tilde{\vartheta}_{k-j-1}\right\|^4-\frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4}$$

$$\times(\frac{1}{2}-4\eta^4-276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2)\eta^4)\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i-1|i-1}\right\|^4$$

$$-\frac{(1-2\eta^2(U_M-1)^2(2+16\alpha_o+32\alpha_o^2))}{2(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^2-(16\alpha_o(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\alpha_o)-2)\left\|\tilde{\vartheta}_{k-j-1}\right\|^2$$

Since $0<k_a<\frac{1}{2}$, $\max\{\frac{1}{2}\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}-\sqrt{\frac{\chi_{\min}^2-3}{2(\chi_{\min}^2+1)}}\right),\frac{\chi_{\min}^2+1}{72}\}<\alpha_o<\min\{\frac{1}{2}\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}+\sqrt{\frac{\chi_{\min}^2-3}{2(\chi_{\min}^2+1)}}\right),\frac{\chi_{\min}^2+1}{36}\}$, $0<\alpha_h<\frac{2\Delta\phi_{\min}^2}{3(\Delta\phi_{\min}^2+1)}$ and $0<\eta<\min\{\sqrt[4]{\frac{1}{2(4+276(U_M-1)^4\alpha_o(1+6\alpha_o+9\alpha_o^2))}},\ \frac{1}{\sqrt{(U_M-1)^2(2+16\alpha_o+32\alpha_o^2)}}\}$, ΔL is negative definite and L is positive definite for both two cases. It is important to note that $\left|\sum_{k=k_0}^{\infty}\Delta L_k\right|=\left|L_\infty-L_0\right|<\infty$ since $\Delta L<0$ as long as (70) and (71) hold. Therefore, system state z_k, observed system state $\hat{z}_{k|k}$, observer errors $\tilde{z}_{k|k}$, observer parameter estimation error vector $\tilde{\vartheta}_k$ and value function parameter estimation error $\tilde{\tilde{\theta}}_k$ for NCS under TCP are all asymptotically stable in the mean. In other words, for the two cases, as $k \to \infty$, $z_k \to 0$, $\hat{z}_{k|k} \to z_k$, $\tilde{z}_{k|k} \to 0$, $\tilde{\vartheta}_k \to 0$ and $\tilde{\tilde{\theta}}_k \to 0$, then $\hat{V}(\hat{z}_{k|k}) \to V^*(z_k)$ and $\hat{u}_{1k} \to u_k^*, u_{2k} \to u_k^*$.

IV. OPTIMAL CONTROL DESIGN FOR NCS UNDER UDP

In this section, stochastic optimal control of NCS under UDP with uncertain system dynamics due to unknown network imperfections is derived by using a novel observer and ADP [17] technique. Unlike TCP protocol, there is no-feedback acknowledgement of received packets generated in UDP and therefore the development of the controller design is different. First, a novel observer is proposed to estimate augment system state at the controller by using UDP information set $\zeta_k = \{\mathbf{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\tau}_{k-1}\}$. Second, similar to the case of TCP, we propose a model-free online tuning of the parameters-based value function estimation algorithm under UDP with augmented observed states. Finally, the convergence proof is given.

A. Novel Observer Design for NCS under UDP

Compared with TCP protocol, UDP does not support the acknowledgements. Therefore, packet transmission information between the controller and actuator υ_k is not known at the controller. However, according to Assumption 1, certain statistical information on packet loss indicator between the controller and actuator υ_k is known i.e. mean value $\bar{\upsilon}$ and variance σ_υ^2 is known. This information of υ_k (i.e. $\bar{\upsilon}$) is used to design the observer for NCS under UDP . Next, the details of observer design are given.

a) Innovation step at time kT_s:

In this step, future system states $\hat{z}_{k+1|k}$ can be predicted as

$$\hat{z}_{k+1|k} = E[z_{k+1}|\zeta_k] = E[\hat{\vartheta}_k^T \overline{M}\,\hat{s}_k|n_k] + \kappa_o \tilde{\mathbf{y}}_{k-1}^o \tag{72}$$

where $\overline{M} = \begin{bmatrix} I^n & 0 & \cdots & 0 \\ 0 & \overline{\upsilon} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{\upsilon} \end{bmatrix}$ and I^n is $n \times n$ identity matrix. The prediction error $\tilde{z}_{k+1|k}$ can be expressed as

$$\tilde{z}_{k+1|k} = z_{k+1} - \hat{z}_{k+1|k} \tag{73}$$

$$= [U_z - \kappa_o (D_y^o)^+]\tilde{z}_{k|k} + [1 - \kappa_o (D_y^o)^+] E[\vartheta^T (M_k - \overline{M})\hat{s}_k|n_k] + [1 - \kappa_o (D_y^o)^+] E[\tilde{\vartheta}_k^T \overline{M}\hat{s}_k|n_k]$$

b) Correction step at time $(k+1)T_s$:

The update law for the parameter vector of the observer $\hat{\vartheta}_k$ for NCS under UDP is defined as

$$\begin{aligned} \hat{\vartheta}_{k+1} &= \hat{\vartheta}_k + \alpha_o E\left[\frac{\overline{M}\,\hat{s}_k (y_{k+1} - \Gamma_{k+1}\hat{z}_{k+1|k})^T}{(\overline{M}\,\hat{s}_k \Gamma_{k+1})^T \overline{M}\,\hat{s}_k \Gamma_{k+1} + 1} \middle| n_k \right] \\ &= \hat{\vartheta}_k + \alpha_o \gamma_{k+1} E\left[\frac{\overline{M}\,\hat{s}_k \Gamma_{k+1} \tilde{z}_{k+1|k}^T}{(\overline{M}\,\hat{s}_k \Gamma_{k+1})^T \overline{M}\,\hat{s}_k \Gamma_{k+1} + 1} \middle| n_k \right] \end{aligned} \tag{74}$$

where α_o is the tuning parameter for the observer satisfying $0 < \alpha_o < 1$. Meantime, the parameter estimation error dynamics $\tilde{\vartheta}_k$ can be expressed as

$$\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_k - \alpha_o \gamma_{k+1} E\left[\frac{\overline{M}\,\hat{s}_k \Gamma_{k+1} \tilde{z}_{k+1|k}^T}{(\overline{M}\,\hat{s}_k \Gamma_{k+1})^T \overline{M}\,\hat{s}_k \Gamma_{k+1} + 1} \middle| n_k \right] \tag{75}$$

Eventually, the observed NCS state $\hat{z}_{k+1|k+1}$ and estimation error dynamics $\tilde{z}_{k+1|k+1}$ in this step can be derived as

$$\hat{z}_{k+1|k+1} = E[z_{k+1}|\psi_{k+1}] = E[\hat{\vartheta}_{k+1}^T \overline{M}\,\hat{s}_k|n_k] + \kappa_o \tilde{\mathbf{y}}_{k-1}^o \tag{76}$$

$$\tilde{z}_{k+1|k+1} = z_{k+1} - \hat{z}_{k+1|k+1}$$
$$= [U_z - \kappa_o (D_y^o)^+] e_{k|k} + [1 - \kappa_o (D_y^o)^+] E[\vartheta^T (M_k - \overline{M}) \hat{s}_k | n_k] + [1 - \kappa_o (D_y^o)^+] E[\tilde{\vartheta}_k^T \overline{M} \hat{s}_k | n_k]]$$

Next, the stability of the observer parameter estimation error $\tilde{\vartheta}_k$ and state estimation error $\tilde{z}_{k|k}$ dynamics are analyzed.

Theorem 4: *(Convergence of observer parameter estimation and error dynamics for NCS):* Given the observer (72), estimation error dynamics (76) and parameter update law (74), according to Assumption 2, there exists positive constant α_o and η satisfying $0 < \eta < \sqrt{\dfrac{1}{3((1 - \frac{\chi_{\min}^2}{\chi_{\min}^2 + 1}\alpha_o)^2 - \alpha_o(1 + \alpha_o))}}$ and $\dfrac{\chi_{\min}^2 + 1}{2(\chi_{\min}^2 - 1)} < \alpha_o < 1$ computable positive constants B_ϑ, B_{eo} such that estimation errors $e_{k|k}$ and parameter estimation errors $\tilde{\vartheta}_k$ (66) for NCS under UDP uniformly ultimately bounded (*UUB*) in the mean with ultimate bounds given by $\|\tilde{z}_{k|k}\| \le B_{eo}$ and $\|\tilde{\vartheta}_k\| \le B_\vartheta$, where $\chi_{\min}^2 < \|\Gamma_{k+1} \overline{M} \hat{s}_k\|$.

Proof: Select the Lyapunov candidate function as $L_o = L_{\tilde{z}} + L_\vartheta$ where $L_{\tilde{z}} = tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i|i}^T \Xi \tilde{z}_{i|i}\}$ and $L_\vartheta = 3tr\{\sum_{i=k-N_0}^{k} \tilde{\vartheta}_i^T \tilde{\vartheta}_i\}$ with $\Xi = \dfrac{1}{(\Psi_M^2 + 1)(U_M - 1)^2}\mathbf{I}$ is a positive definite matrix. Then the first difference of this Lyapunov candidate function can be represented as $\Delta L_o = \Delta L_{\tilde{z}} + \Delta L_\vartheta$.

First, we consider first difference of $L_{\tilde{z}}$ (i.e. $\Delta L_{\tilde{z}} = tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i+1|i+1}^T \Xi \tilde{z}_{i+1|i+1}\} - tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i|i}^T \Xi \tilde{z}_{i|i}\}$). Applying Cauchy-Schwartz inequality and equation (75) and (76), $\Delta L_{\tilde{z}}$ can be derived as

$$\Delta L_{\tilde{z}} = tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i+1|i+1}^T \Xi \tilde{z}_{i+1|i+1}\} - tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i|i}^T \Xi \tilde{z}_{i|i}\}$$

$$= tr\{\sum_{i=k-N_0}^{k} \{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i]$$
$$+[1-\kappa_o (D_y^o)^+] E[(\tilde{\vartheta}_{i+1}^T \overline{M}\hat{s}_i)|n_i]\}^T \Xi\{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o (D_y^o)^+]$$
$$\times E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i] + [1-\kappa_o (D_y^o)^+] E[(\tilde{\vartheta}_{i+1}^T M_i \hat{s}_i)|n_i]\}\} - tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i|i}^T \Xi \tilde{z}_{i|i}\}$$

$$= tr\{\sum_{i=k-N_0}^{k} \{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i] + [1-\kappa_o (D_y^o)^+]$$
$$\times E[((\tilde{\vartheta}_i - \alpha_o \gamma_{i+1} \frac{\overline{M}\hat{s}_i \tilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i)^T \overline{M}\hat{s}_i + 1})^T \overline{M}\hat{s}_i)|n_i]\}^T \Xi\{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} \tag{77}$$
$$+[1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i]$$

$$+[1-\kappa_o (D_y^o)^+] E[((\tilde{\vartheta}_i - \alpha_o \gamma_{i+1} \frac{\overline{M}\hat{s}_i \Gamma_{i+1} \tilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i \Gamma_{i+1})^T \overline{M}\hat{s}_i \Gamma_{i+1} + 1})^T M_i \hat{s}_i)|n_i]\}\} - tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i|i}^T \Xi \tilde{z}_{i|i}\}$$

$$\le tr\{\sum_{i=k-N_0}^{k} \{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i]$$
$$+[1-\kappa_o (D_y^o)^+] E[\tilde{\vartheta}^T \overline{M}\hat{s}_i|n_i] - \frac{\chi_{\min}^2}{\chi_{\min}^2 + 1} \alpha_o \gamma_{i+1} \tilde{z}_{i+1|i}]^T \Xi\{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o (D_y^o)^+]$$
$$* E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i] + E[\tilde{\vartheta}^T \overline{M}\hat{s}_i|n_i] - \frac{\chi_{\min}^2}{\chi_{\min}^2 + 1} \alpha_o \gamma_{i+1} \tilde{z}_{i+1|i}]\} - tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i|i}^T \Xi \tilde{z}_{i|i}\}$$

$$\le tr\{\sum_{i=k-N_0}^{k} \{[U_z - \kappa_o (D_y^o)^+] e_{i|i} + [1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i]$$

$$+[1-\kappa_o (D_y^o)^+] E[\tilde{\vartheta}^T \overline{M}\hat{s}_i|n_i] - \frac{\chi_{\min}^2}{\chi_{\min}^2 + 1} \gamma_{i+1} \alpha_o \{[U_z - \kappa_o (D_y^o)^+] e_{i|i} + [1-\kappa_o (D_y^o)^+]$$

$$\times E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i] + [1-\kappa_o (D_y^o)^+] E[\tilde{\vartheta}^T \overline{M}\hat{s}_i|n_i]\}\}^T \Xi\{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i}$$
$$+[1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i] + [1-\kappa_o (D_y^o)^+] E[\tilde{\vartheta}^T \overline{M}\hat{s}_i|n_i]$$

$$-\frac{\chi_{\min}^2 \alpha_o \gamma_{i+1}}{\chi_{\min}^2 + 1}\{[U_z - \kappa_o (D_y^o)^+]\tilde{z}_{i|i} + [1-\kappa_o (D_y^o)^+] E[(\vartheta^T (M_i - \overline{M})\hat{s}_i)|n_i]$$

$$+[1-\kappa_o(D_y^o)^+]E[\widetilde{\vartheta}^T\overline{M}\hat{s}_i|n_i]\}\}\}-tr\{\sum_{i=k-L}^{k}\widetilde{z}_{i|i}^T\Xi\widetilde{z}_{i|i}\}$$

$$\le tr\{\sum_{i=k-N_0}^{k}\{(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})[U_z-\kappa_o(D_y^o)^+]\widetilde{z}_{i|i}+(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})[1-\kappa_o(D_y^o)^+]$$
$$*E[(\vartheta^T(M_i-\overline{M})\hat{s}_i)|n_i]+(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})[1-\kappa_o(D_y^o)^+]E[\widetilde{\vartheta}^T\overline{M}\hat{s}_i|n_i]\}^T$$
$$\times\Xi\{(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})[U_z-\kappa_o(D_y^o)^+]\widetilde{z}_{i|i}+(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})[1-\kappa_o(D_y^o)^+]$$

$$*E[(\vartheta^T(M_i-\overline{M})\hat{s}_i)|n_i]+(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})[1-\kappa_o(D_y^o)^+]E[\widetilde{\vartheta}^T\overline{M}\hat{s}_i|n_i]\}\}$$

$$-tr\{\sum_{i=k-D_0}^{k}e_{i|i}^T\Xi e_{i|i}\}$$

$$\le-\Xi\sum_{i=k-N_0}^{k}(1-3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\eta^2)\left\|U_z-\kappa_o(D_y^o)^+\right\|^2\left\|\widetilde{z}_{i|i}\right\|^2$$
$$+\sum_{i=k-N_0}^{k}(3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\left\|M_i\hat{s}_i\right\|^2\left\|1-\kappa_o(D_y^o)^+\right\|^2\Xi tr\{\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\})$$
$$+tr\{\sum_{i=k-N_0}^{k}3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2E[(\frac{\overline{M}}{M_i}-1)\vartheta^T\vartheta(\frac{\overline{M}}{M_i}-1)^T|n_i]\}$$

Secondly, using observer parameter estimation error dynamics (75), the term ΔL_ϑ can be expressed as

$$\Delta L_\vartheta=3tr\{\sum_{i=k-N_0}^{k}\widetilde{\vartheta}_{i+1}^T\widetilde{\vartheta}_{i+1}\}-3tr\{\sum_{i=k-D_0}^{k}\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\}$$
$$=3tr\{\sum_{i=k-N_0}^{k}(\widetilde{\vartheta}_i-\alpha_o\gamma_{i+1}\frac{\overline{M}\hat{s}_i\Gamma_{i+1}\widetilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1})^T(\widetilde{\vartheta}_i-\alpha_o\gamma_{i+1}\times\frac{\overline{M}\hat{s}_i\Gamma_{i+1}\widetilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1})\}\quad(78)$$
$$-3tr\{\sum_{i=k-N_0}^{k}\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\}$$

$$=3tr\{\sum_{i=k-N_0}^{k}\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\}-6tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}\widetilde{\vartheta}_i^T\frac{\overline{M}\hat{s}_i\Gamma_{i+1}\widetilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}\}$$

$$+3tr\{\sum_{i=k-D_0}^{k}\frac{\alpha_o^2\gamma_{i+1}^2\tilde{z}_{i+1|i}(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}\tilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}\}-3tr\{\sum_{i=k-D_0}^{k}\tilde{\vartheta}_i^T\tilde{\vartheta}_i\}$$

$$=-6tr\{\sum_{i=k-D_0}^{k}\alpha_o\gamma_{i+1}\tilde{\vartheta}_i^T\frac{\overline{M}\hat{s}_i\Gamma_{i+1}\tilde{z}_{i+1|i}^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}\}+3tr\{\sum_{i=k-D_0}^{k}\frac{\alpha_o^2\gamma_{i+1}^2\tilde{z}_{i+1|i}(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}\tilde{z}_{i+1|i}^T}{((\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1)^2}\}$$

$$-6tr\{\sum_{i=k-N_0}^{k}\frac{\alpha_o\gamma_{i+1}}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}[\tilde{\vartheta}_i^T\overline{M}\hat{s}_i\Gamma_{i+1}([U_z-\kappa_o(D_y^o)^+]\tilde{z}_{i|i}$$

$$+[1-\kappa_o(D_y^o)^+]E[(\vartheta^T(M_i-\overline{M})\hat{s}_i)|n_i]+[1-\kappa_o(D_y^o)^+]E[\tilde{\vartheta}^T\overline{M}\hat{s}_i|n_i])^T)]\}$$

$$+9tr\{\sum_{i=k-N_0}^{k}\frac{\alpha_o^2\gamma_{i+1}^2}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}+[1-\kappa_o(D_y^o)^+]E[(\vartheta^T(M_i-\overline{M})\hat{s}_i)|n_i]$$

$$+[1-\kappa_o(D_y^o)^+]E[\tilde{\vartheta}^T\overline{M}\hat{s}_i|n_i])^T)]\}+9tr\{\sum_{i=k-N_0}^{k}\frac{\alpha_o^2\gamma_{i+1}^2\Gamma_{i+1}}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}$$

$$*([U_z-\kappa_o(D_y^o)^+]e_{i|i}+[1-\kappa_o(D_y^o)^+]E[(\vartheta^T(M_i-\overline{M})\hat{s}_i)|n_i]+[1-\kappa_o(D_y^o)^+]E[\tilde{\vartheta}^T\overline{M}\hat{s}_i|n_i])$$

$$\times([U_z-\kappa_o(D_y^o)^+]\tilde{z}_{i|i}+[1-\kappa_o(D_y^o)^+]E[(\vartheta^T(M_i-\overline{M})\hat{s}_i)|n_i]$$

$$+[1-\kappa_o(D_y^o)^+]E[\tilde{\vartheta}^T\overline{M}\hat{s}_i|n_i])^T\}$$

$$\le-6tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1})E[\frac{\tilde{\vartheta}_i^T\overline{M}\hat{s}_i\Gamma_{i+1}\tilde{z}_{i|i}^T[U_z-\kappa_o(D_y^o)^+]}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}|n_i]\}$$

$$-6tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1})E[\frac{\tilde{\vartheta}_i^T\overline{M}\hat{s}_i\Gamma_{i+1}\Gamma_{i+1}^T\hat{s}_i^TM_i^T\vartheta}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}|n_i]\}-3\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}$$

$$\times(2-\alpha_o\gamma_{i+1})tr\{\tilde{\vartheta}_i^T\tilde{\vartheta}_i\}+6tr\{\sum_{i=k-N_0}^{k}\alpha_o^2\gamma_{i+1}^2E[\frac{[U_z-\kappa_o(D_y^o)^+]^T\Gamma_{i+1}\tilde{z}_{i|i}\hat{s}_i^TM_i\vartheta)^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}|n_i]\}+3\alpha_o^2\gamma_{i+1}^2\sigma_\upsilon^2\|\vartheta\|^2$$

$$+3tr\{\sum_{i=k-N_0}^{k}\alpha_o^2\gamma_{i+1}^2E[\frac{[U_z-\kappa_o(D_y^o)^+]^T\tilde{z}_{i|i}\Gamma_{i+1}\Gamma_{i+1}^T\tilde{z}_{i|i}^T[U_z-\kappa_o(D_y^o)^+])^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}|n_i]\}$$

$$\le-3\alpha_o\gamma_{i+1}(2-\alpha_o\gamma_{i+1})tr\{\sum_{i=k-N_0}^{k}\tilde{\vartheta}_i^T\tilde{\vartheta}_i\}+3tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1})\tilde{\vartheta}_i^T\tilde{\vartheta}_i\}$$

$$+3tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1})E[\frac{[U_z-\kappa_o(D_y^o)^+]^T\tilde{z}_{i|i}\Gamma_{i+1}\Gamma_{i+1}^T\tilde{z}_{i|i}^T[U_z-\kappa_o(D_y^o)^+]^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}|n_i]\}$$

$$+3tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1})\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\}+3\sum_{i=k-N_0}^{k}\frac{\alpha_o\gamma_{i+1}(1-\alpha_o\gamma_{i+1})}{\overline{\upsilon}^2}tr\{\vartheta^T(M_i-\overline{M})(M_i-\overline{M})^T\vartheta\}$$
$$+6tr\{\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}\frac{\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T(\overline{M}\hat{s}_i\Gamma_{i+1})+1}\}+6tr\{\sum_{i=k-N_0}^{k}\alpha_o^2\gamma_{i+1}^2$$
$$\times E[\frac{[U_z-\kappa_o(D_y^o)^+]^T\widetilde{z}_{i|i}\Gamma_{i+1}\Gamma_{i+1}^T\widetilde{z}_{i|i}^T[U_z-\kappa_o(D_y^o)^+]^T}{(\overline{M}\hat{s}_i\Gamma_{i+1})^T\overline{M}\hat{s}_i\Gamma_{i+1}+1}|n_i]\}$$

$$\leq -3\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(\alpha_o\gamma_{i+1}-\frac{2}{\chi_{\min}^2+1})\left\|\widetilde{\vartheta}_i\right\|^2+3\sum_{i=k-N_0}^{k}\frac{\alpha_o\gamma_{i+1}(1+\alpha_o\gamma_{i+1})}{(\chi_{\min}^2+1)(U_M-1)^2}\eta^2\left\|\widetilde{z}_{i|i}\right\|^2+3\alpha_o\gamma_{i+1}\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\left\|\vartheta\right\|^2$$

Eventually, combing (77) and (78), the overall first difference ΔL_o can be expressed as

$$\Delta L_o=\Delta L_{\widetilde{z}}+\Delta L_{\vartheta}$$
$$=tr\{\sum_{i=k-N_0}^{k}\widetilde{z}_{i+1|i+1}^T\Xi\widetilde{z}_{i+1|i+1}\}-tr\{\sum_{i=k-N_0}^{k}\widetilde{z}_{i|i}^T\Xi\widetilde{z}_{i|i}\}+tr\{\sum_{i=k-N_0}^{k}\widetilde{\vartheta}_{i+1}^T\widetilde{\vartheta}_{i+1}\}-tr\{\sum_{i=k-N_0}^{k}\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\} \tag{79}$$

$$\leq -\Xi\sum_{i=k-N_0}^{k}(1-3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\eta^2)\left\|\widetilde{z}_{i|i}\right\|^2+\sum_{i=k-N_0}^{k}(3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\left\|M_i\hat{s}_i\right\|^2\Xi tr\{\widetilde{\vartheta}_{i+1}^T\widetilde{\vartheta}_{i+1}\})$$

$$+tr\{\sum_{i=k-N_0}^{k}3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\Xi E[\hat{s}_i(\overline{M}-M_i)\vartheta^T\vartheta(\overline{M}-M_i)^Ts_i^T|n_i]\}-3\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(\alpha_o\gamma_{i+1}-\frac{2}{\chi_{\min}^2+1})\left\|\widetilde{\vartheta}_i\right\|^2$$
$$+3\sum_{i=k-N_0}^{k}\frac{\alpha_o\gamma_{i+1}(1+\alpha_o\gamma_{i+1})}{\chi_{\min}^2+1}\times\eta^2\left\|\widetilde{z}_{i|i}\right\|^2+3\alpha_o\gamma_{i+1}\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\left\|\vartheta\right\|^2$$

Similar to NCS under TCP, packet loss indicator between the sensor and controller γ_{k+1} can be equal to 0 or 1. Therefore, ΔL_o needs to be separated into two different scenarios as shown below.

Case 1: $\gamma_{k+1}=1$ (No packet losses)

Substituting γ_{i+1} value into (70), ΔL_o can be represented as

$$\Delta L_o\leq -\Xi\sum_{i=k-N_0}^{k}(1-3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\eta^2)\left\|\widetilde{z}_{i|i}\right\|^2+\sum_{i=k-N_0}^{k}(3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\left\|M_i\hat{s}_i\right\|^2$$

$$*\Xi tr\{\widetilde{\vartheta}_i^T\widetilde{\vartheta}_i\})+tr\{\sum_{i=k-N_0}^{k}3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\Xi E[\hat{s}_i(\overline{M}-M_i)\vartheta^T\vartheta(\overline{M}-M_i)^Ts_i^T|n_i]\}$$

$$-3\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(\alpha_o\gamma_{i+1}-\frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_i\right\|^2+\sum_{i=k-N_0}^{k}\frac{3\alpha_o\gamma_{i+1}(1+\alpha_o\gamma_{i+1})}{\chi_{\min}^2+1}\eta^2\left\|\tilde{z}_{i|i}\right\|^2+3\alpha_o\gamma_{i+1}\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\left\|\vartheta\right\|^2$$

$$\le-\frac{(1-3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2\eta^2(U_M-1)^2))}{(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2+3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2 \tag{80}$$

$$\times\frac{\left\|\tilde{\vartheta}_k\right\|^2}{(U_M-1)^2}+3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2(U_M-1)^2}\left\|\vartheta\right\|^2-3\alpha_o(\alpha_o-$$

$$\frac{2}{\chi_{\min}^2+1})\left\|\tilde{\vartheta}_k\right\|^2+3\frac{\alpha_o(1+\alpha_o)}{\chi_{\min}^2+1}\eta^2\left\|\tilde{z}_{k|k}\right\|^2+3\alpha_o\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\left\|\vartheta\right\|^2$$

$$\le-\frac{(1-3\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-\alpha_o(1+\alpha_o)))}{(\sigma_0^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2-6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2})\left\|\tilde{\vartheta}_k\right\|^2$$

$$+3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\left\|\vartheta\right\|^2$$

$$\le-\frac{(1-3\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-\alpha_o(1+\alpha_o)))}{(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2-6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2})\left\|\tilde{\vartheta}_k\right\|^2+\varepsilon_M^o$$

where $\varepsilon_M^o=3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\left\|\vartheta\right\|^2$.

Case 2: $\gamma_{k+1}=0$. (With packet lost)

According to Assumption 2, observer stability region needs to be satisfied by packets lost probability between the sensor and controller. Hence, if $\gamma_{k+1}=0$, then there exists $j\in[k-N_0,k]$ such that $\gamma_{k+1}=\gamma_k=\cdots==\gamma_{k-j+1}=0$ and $\gamma_{k-j}=1$. In this case, first difference ΔL_o can be derived as

$$\Delta L_o\le-\Xi\sum_{i=k-N_0}^{k}(1-3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\eta^2)\left\|\tilde{z}_{i|i}\right\|^2+\sum_{i=k-N_0}^{k}(3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\left\|M_i\hat{s}_i\right\|^2$$

$$*\Xi tr\{\tilde{\vartheta}_i^T\tilde{\vartheta}_i\}) + tr\{\sum_{i=k-N_0}^{k} 3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\gamma_{i+1})^2\,\Xi E[\hat{s}_i(\overline{M}-M_i)\vartheta^T\vartheta(\overline{M}-M_i)^T s_i^T|n_i]\}$$

$$-3\sum_{i=k-N_0}^{k}\alpha_o\gamma_{i+1}(\alpha_o\gamma_{i+1}-\frac{2}{\chi_{\min}^2+1})\|\tilde{\vartheta}_i\|^2 + \sum_{i=k-N_0}^{k}\frac{3\alpha_o\gamma_{i+1}(1+\alpha_o\gamma_{i+1})}{\chi_{\min}^2+1}\eta^2\|\tilde{z}_{i|i}\|^2 + 3\alpha_o\gamma_{i+1}\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\|\vartheta\|^2 \quad (81)$$

$$\le -\frac{(1-3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2\eta^2(U_M-1)^2)\sum_{i=k-D_0}^{k}\|\tilde{z}_{i|i}\|^2}{(\chi_{\min}^2+1)(U_M-1)^2} + 3\alpha_o\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\|\vartheta\|^2$$

$$+3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2\frac{\|\tilde{\vartheta}_{k-j-1}\|^2}{(U_M-1)^2} + 3(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\|\vartheta\|^2 - 3\alpha_o(\alpha_o-\frac{2}{\chi_{\min}^2+1})\|\tilde{\vartheta}_{k-j-1}\|^2$$

$$+3\frac{\alpha_o(1+\alpha_o)}{\chi_{\min}^2+1}\eta^2\sum_{i=k-D_0}^{k}\|\tilde{z}_{i|i}\|^2$$

$$\le -\frac{(1-3\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-\alpha_o(1+\alpha_o)))}{(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\|\tilde{z}_{i|i}\|^2 - 6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2})\|\tilde{\vartheta}_{k-j-1}\|^2$$

$$+3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\|\vartheta\|^2$$

$$\le -\frac{(1-3\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-\alpha_o(1+\alpha_o)))}{(\chi_{\min}^2+1)(U_M-1)^2}\|\tilde{z}_{k|k}\|^2$$

$$-6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2})\|\tilde{\vartheta}_{k-j-1}\|^2 + \varepsilon_M^o$$

where $\varepsilon_M^o = 3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\|\vartheta\|^2$.

Since α_o and η are positive constants which satisfy $\frac{\chi_{\min}^2+1}{2(\chi_{\min}^2-1)} < \alpha_o < 1$ and $0<\eta<\sqrt{\frac{1}{3(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-\alpha_o(1+\alpha_o))}}$, and $\varepsilon_M^o = 3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\overline{\upsilon}^2}\|\vartheta\|^2$ is a bounded constant, then based on (80) and (81), the first difference of ΔL_o is less than zero provided the following inequalities hold

$$\left\| e_{k|k} \right\| > \sqrt{\frac{(\chi_{\min}^2 + 1)\varepsilon_M^o}{1 - 3\eta^2 (U_M - 1)^2 ((1 - \frac{\chi_{\min}^2}{\chi_{\min}^2 + 1}\alpha_o)^2 - \alpha_o (1 + \alpha_o))}} \equiv B_{eo} \text{ or}$$

$$\left\| \tilde{\vartheta}_k \right\| > \sqrt{\frac{\varepsilon_M^o}{(\frac{\chi_{\min}^2 - 1}{\chi_{\min}^2 + 1}\alpha_o - \frac{1}{2})}} \equiv B_{\vartheta} \tag{82}$$

Using the standard Lyapunov extension [10], the observer error and its parameter estimation errors are *UUB* in the mean.

Next, we will propose stochastic optimal control for NCS under UDP with observed system state $\hat{z}_{k|k}$ by using value function estimation.

B. Value Function Definition for NCS under UDP

Based on the optimal control theory [7], NCS description (3) and UDP protocol, the stochastic value function for NCS under UDP can be rewritten as

$$V^*(z_k) = E[z_k^T P_k z_k | \zeta_k] \tag{83}$$

where ζ_k is network information set for NCS under UDP and $P_k > 0$ is the solution of SRE. Meanwhile, optimal action dependent value function can be expressed as

$$V^*(z_k) = E\{[r(z_k, u_k) + V^*(z_{k+1})] | \zeta_k\} = E\{[z_k^T \ u_k^T]\Theta_k [z_k^T \ u_k^T]^T | \zeta_k\} \tag{84}$$

Similar to NCS under TCP, we can substitute value function into Bellman equation to derive $E(\Theta_k | \psi_k)$ as

$$\overline{\Theta}_k = E(\Theta_k | \zeta_k) = \begin{bmatrix} \overline{\Theta}_k^{zz} & \overline{\Theta}_k^{zu} \\ \overline{\Theta}_k^{uz} & \overline{\Theta}_k^{uu} \end{bmatrix}$$

$$= \begin{bmatrix} O + E(A_{zk}^T P_{k+1} A_{zk} | \zeta_k) & E(A_{zk}^T P_{k+1} B_{zk} | \zeta_k) \\ E(B_{zk}^T P_{k+1} A_{zk} | \zeta_k) & R + E(B_{zk}^T P_{k+1} B_{zk} | \zeta_k) \end{bmatrix} \tag{85}$$

Therefore, optimal control gain for NCS under UDP can be expressed in terms of parameter, $\overline{\Theta}_k$,as

$$K_k = [R + E(B_{zk}^T P_{k+1} B_{zk} | \varsigma_k)]^{-1} E(B_{zk}^T P_{k+1} A_{zk} | \zeta_k) = (\overline{\Theta}_k^{uu})^{-1} \overline{\Theta}_k^{uz} \tag{86}$$

Obviously, if slow-varying parameter $\overline{\Theta}_k$ can be estimated online, optimal control gain can be calculated without known system dynamics by using (86).

C. Value Function Estimation of NCS under UDP

Based on value function estimator used in stochastic optimal control of NCS under TCP (Section 2.3), we can define the value function with observed system states for NCS under UDP as

$$V^*(\hat{z}_{k|k}) = \varphi_k^{eT} \overline{\Theta}_k \varphi_k^e = \overline{\theta}_k^T \overline{\varphi}_k^e \tag{87}$$

While the value function estimation can be represented as

$$\hat{V}(\hat{z}_{k|k}) = \varphi_k^{eT} \hat{\overline{\Theta}}_k \varphi_k^e = \hat{\overline{\theta}}_k^T \overline{\varphi}_k^e \tag{88}$$

where the augment state $\varphi_k^e = [\hat{z}_{k|k}^T \; u_k^T]^T \in \mathbb{R}^{n+bm=l}$, $\overline{\varphi}_k^e = ((\overline{\varphi}_{k1}^e)^2, \cdots, \overline{\varphi}_{k1}^e \overline{\varphi}_{kl}^e, (\overline{\varphi}_{k2}^e)^2, \cdots, \overline{\varphi}_{kl-1}^e \overline{\varphi}_{kl}^e, (\overline{\varphi}_{kl}^e)^2)$ is nothing but Kronecker product quadratic polynomial stochastic independent basis vector, and $\hat{\overline{\theta}}_k = vec(\hat{\overline{\Theta}}_k)$.

Using the value function estimation algorithm that is proposed as part of stochastic optimal control for NCS under TCP in Section III.B, the parameter update law of value function estimation $\hat{\overline{\theta}}_k$ for NCS under UDP can be derived as

$$\hat{\bar{\theta}}_{k+1} = \hat{\bar{\theta}}_k + \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^o + e_k^a)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1} \tag{89}$$

where the residual error $e_{k+1}^a + e_{k+1}^o = r(\hat{z}_{k|k}, u_k) + \hat{\bar{\theta}}_{k+1}^T \Delta\bar{\varphi}_k^e$ and α_h is tuning parameter for value function estimation.

Next dynamics of parameter estimation errors of value function estimation can be expressed as

$$\tilde{\bar{\theta}}_{k+1} = \tilde{\bar{\theta}}_k - \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^o + e_k^a)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1} \tag{90}$$

Since the estimated value function with observed system states results in estimation errors for value function $V^*(z_k)$, stability of the estimation error dynamics need to be ensured.

Theorem 5 *(Boundness of the Value Function Estimation Errors).* Given the initial conditions for the parameter vector $\hat{\bar{\theta}}_0$ of value function estimation is bounded in the set $\mathbf{\Omega}$, let u_{0k} be an initial admissible control policy for the linear NCS under UDP. Let observer parameter update law be given by (89). Then there exists positive constants α_h, α_o , η satisfying $0 < \alpha_h < \dfrac{2\Delta\phi_{\min}^2}{3(\Delta\phi_{\min}^2 + 1)}$, $0 < \alpha_o < \dfrac{(\chi_{\min}^2 + 1)(\chi_{\min}^2 - 5)}{21(\chi_{\min}^2 + 1)^2 + 3}$ with $0 < \eta < \sqrt[4]{\dfrac{3\alpha_o^3}{(1 + 6\alpha_o + 9\alpha_o^2 + \alpha_o^3)(U_M - 1)^4}}$, $0 < \Delta\phi_{\min} < \left\|\Delta\bar{\varphi}_k^e\right\|$,where $\Delta\phi_{\min}$ is the lower bound of $\left\|\Delta\bar{\varphi}_k^e\right\|$,and computable positive constant B_θ, such that the value function estimation parameter estimation errors (90) are *UUB* in the mean with ultimate bounds given by $\left\|\tilde{\bar{\theta}}_k\right\| \le B_\theta$, where $\chi_{\min}^2 < \left\|\Gamma_{k+1}\bar{M}\hat{s}_k\right\|^2$.

Proof: Consider the positive definite Lyapunov candidate function

$$L_J = L_\theta + L_{ao} \tag{91}$$

where $L_\theta = tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\}$ and $L_{ao} = tr\{\sum_{i=k-N_0}^{k} \vartheta_{i-1}^T \vartheta_{i-1} \vartheta_{i-1}^T \vartheta_{i-1}\} + tr\{\sum_{i=k-N_0}^{k} \tilde{z}_{i-1|i-1} \tilde{z}_{i-1|i-1}^T \Lambda \tilde{z}_{i-1|i-1} \tilde{z}_{i-1|i-1}^T\}$ with $\Pi = \dfrac{(\Delta\phi_{\min}^2 + 1)\alpha_o^4}{2\alpha_h^2 O^2 (\chi_{\min}^2 + 1)^2}\mathbf{I}$ and $\Lambda = \dfrac{\alpha_o^4}{4(\psi_M^2 + 1)^2 (U_M - 1)^4}\mathbf{I}$ are positive definite matrices. The first difference of (91) can be derived as $\Delta L_J = \Delta L_\theta + \Delta L_{ao}$. Since Lyapunov candidate function includes observer parameter, we have to separate proofs into two cases which is similar to NCS under TCP.

Case 1: $\gamma_{k+1} = 1$. (No packet losses)

According to (75), (76) and (90), ΔL_J can be expressed as

$$\begin{aligned}
\Delta L_J &= \Delta L_\theta + \Delta L_{ao} \\
&= tr\{(\tilde{\tilde{\theta}}_k - \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})^T \Pi (\tilde{\tilde{\theta}}_k - \alpha_h \frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})\} - tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + \Lambda \left\|\tilde{z}_{k|k}\right\|^4 \\
&\quad - \Lambda \left\|\tilde{z}_{k-1|k-1}\right\|^4 + \left\|\tilde{\vartheta}_k\right\|^4 - \left\|\tilde{\vartheta}_{k-1}\right\|^4
\end{aligned} \tag{92}$$

$$\begin{aligned}
&\le tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} - \alpha_h tr\{(\frac{(e_k^a + e_k^o)\Delta\bar{\varphi}_k^{eT}}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1}) \Pi \tilde{\tilde{\theta}}_k\} - \alpha_h tr\{\tilde{\tilde{h}}_k^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})\} \\
&\quad + \alpha_h^2 tr\{(\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})\} - tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + 4\Lambda \left\|U_z - \kappa_o (D_y^o)^+\right\|^4 \left\|e_{k-1|k-1}\right\|^4 \\
&\quad + 4\Lambda \left\|\tilde{\vartheta}_{k-1}\right\|^4 \left\|\overline{M}\hat{s}_{k-1}\right\|^4 + 4\Lambda \frac{\sigma_\upsilon^4}{\upsilon^4} \left\|\vartheta\right\|^4 \left\|\overline{M}\hat{s}_{k-1}\right\|^4 - \Lambda \left\|e_{k-1|k-1}\right\|^4 - \alpha_o (\frac{\chi_{\min}^2 - 5}{\chi_{\min}^2 + 1} - (7 + \\
&\quad \frac{1}{(\chi_{\min}^2 + 1)^2})\alpha_o - (\frac{6}{(\chi_{\min}^2 + 1)^2} + 12)\alpha_o^2 - 4\alpha_o^3) \left\|\tilde{\vartheta}_{k-1}\right\|^4 \\
&\quad + \frac{\alpha_o (1 + 6\alpha_o + 9\alpha_o^2)}{(\chi_{\min}^2 + 1)^2} \left\|U_z - \kappa_o (D_y^o)^+\right\|^4 \left\|e_{k-1|k-1}\right\|^4 + (\frac{3\alpha_o^2 (2 + 3\alpha_o)}{(\chi_{\min}^2 + 1)^2} + 1) \left\|\vartheta\right\|^4 \frac{\sigma_\upsilon^4}{\upsilon^4}
\end{aligned}$$

$$\le tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} - \alpha_h tr\{(\frac{(e_k^a + e_k^o)\Delta\bar{\varphi}_k^{eT}}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1}) \Pi \tilde{\tilde{\theta}}_k\} - \alpha_h tr\{\tilde{\tilde{h}}_k^T \Pi (\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT} \Delta\bar{\varphi}_k^e + 1})\}$$

$$-\alpha_h tr\{\tilde{\tilde{h}}_k^T \Pi(\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\} + \alpha_h^2 tr\{(\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})^T \Pi(\frac{\Delta\bar{\varphi}_k^e (e_k^a + e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1})\}$$

$$-tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} - \frac{(4\alpha_o^4 - \alpha_o(1+6\alpha_o + 9\alpha_o^2 + \alpha_o^3)(U_M - 1)^4\eta^4)}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$-\alpha_o(\frac{\chi_{\min}^2 - 5}{\chi_{\min}^2 + 1} - (7 + \frac{1}{(\chi_{\min}^2 + 1)^2})\alpha_o - (\frac{6}{(\chi_{\min}^2 + 1)^2} + 12)\alpha_o^2 - 5\alpha_o^3)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$
$$+(\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2 + 1)^2} + 1 + \alpha_o^4)\left\|\vartheta\right\|^4 \frac{\sigma_\upsilon^4}{\upsilon^4}$$

$$\le -2\alpha_h tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} - 2\alpha_h tr\{\frac{e_k^o \Pi \Delta\varphi_k^{eT}\tilde{\tilde{\theta}}_k}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\} + 2\alpha_h^2 tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + 2\alpha_h^2 tr\{\frac{(e_k^o)^2}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e + 1}\}$$
$$+2\alpha_h tr\{\frac{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k}{\Delta\phi_{\min}^2 + 1}\} - \frac{(4\alpha_o^4 - \alpha_o(1+6\alpha_o + 9\alpha_o^2 + \alpha_o^3)(U_M - 1)^4\eta^4)}{(\sigma_0^2 + 1)^2 (U_M - 1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$-\alpha_o(\frac{\chi_{\min}^2 - 5}{\chi_{\min}^2 + 1} - (7 + \frac{1}{(\chi_{\min}^2 + 1)^2})\alpha_o - (\frac{6}{(\chi_{\min}^2 + 1)^2} + 12)\alpha_o^2 - 5\alpha_o^3)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$+(\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2 + 1)^2} + 1 + \alpha_o^4)\left\|\vartheta\right\|^4 \frac{\sigma_\upsilon^4}{\upsilon^4}$$
$$\le -\alpha_h(2 - 3\alpha_h - \frac{2}{\Delta\phi_{\min}^2 + 1}) tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} + \frac{2\alpha_h^2 O^2 \Pi}{\Delta\phi_{\min}^2 + 1}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$
$$-\frac{(4\alpha_o^4 - \alpha_o(1+6\alpha_o + 9\alpha_o^2 + \alpha_o^3)(U_M - 1)^4\eta^4)}{(\sigma_0^2 + 1)^2 (U_M - 1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$
$$-\alpha_o(\frac{\chi_{\min}^2 - 5}{\chi_{\min}^2 + 1} - (7 + \frac{1}{(\chi_{\min}^2 + 1)^2})\alpha_o - (\frac{6}{(\chi_{\min}^2 + 1)^2} + 12)\alpha_o^2$$

$$-5\alpha_o^3)\left\|\tilde{\vartheta}_{k-1}\right\|^4 + (\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2 + 1)^2} + 1 + \alpha_o^4)\left\|\vartheta\right\|^4 \frac{\sigma_\upsilon^4}{\upsilon^4}$$

$$\le -\alpha_h(2 - 3\alpha_h - \frac{2}{\Delta\phi_{\min}^2 + 1}) tr\{\tilde{\tilde{\theta}}_k^T \Pi \tilde{\tilde{\theta}}_k\} - \frac{1}{(\chi_{\min}^2 + 1)^2 (U_M - 1)^4}(3\alpha_o^4 - \alpha_o(1 + 6\alpha_o + 9\alpha_o^2 + \alpha_o^3)$$
$$*(U_M - 1)^4\eta^4)\left\|\tilde{z}_{k-1|k-1}\right\|^4 - \alpha_o(\frac{\chi_{\min}^2 - 5}{\chi_{\min}^2 + 1} - (7 + \frac{1}{(\chi_{\min}^2 + 1)^2})\alpha_o$$
$$-(\frac{6}{(\chi_{\min}^2 + 1)^2} + 12)\alpha_o^2 - 5\alpha_o^3)\left\|\tilde{\vartheta}_{k-1}\right\|^4 + (\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2 + 1)^2} + 1 + \alpha_o^4)\left\|\vartheta\right\|^4 \frac{\sigma_\upsilon^4}{\upsilon^4}$$

$$\leq -\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\left\|\tilde{\tilde{\theta}}_k\right\|^2-\frac{1}{(\chi_{\min}^2+1)^2(U_M-1)^4}(3\alpha_o^4-\alpha_o(1+6\alpha_o$$
$$\frac{+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}{(\chi_{\min}^2+1)^2(U_M-1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4-\alpha_o(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1}-(7+\frac{1}{(\chi_{\min}^2+1)^2})\alpha_o$$
$$-(\frac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2-5\alpha_o^3)\left\|\tilde{\vartheta}_{k-1}\right\|^4+\varepsilon_M^{AE}$$

where $\varepsilon_M^{AE}=(\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2}+1+\alpha_o^4)\left\|\vartheta\right\|^4\frac{\sigma_\upsilon^4}{\underline{\upsilon}^4}$.

Case 2: $\gamma_{k+1}=0$. (With packet lost)

Using the Assumption 2 and update law for observer estimated parameter, the first difference of ΔL_J can be expressed as

$$\Delta L_J=\Delta L_\theta+\Delta L_{ao}$$
$$=tr\{(\tilde{\tilde{\theta}}_k-\alpha_h\frac{\Delta\bar{\varphi}_k^e(e_k^a+e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e+1})^T\Pi(\tilde{\tilde{\theta}}_k-\alpha_h\frac{\Delta\bar{\varphi}_k^e(e_k^a+e_k^o)^T}{\Delta\bar{\varphi}_k^{eT}\Delta\bar{\varphi}_k^e+1})\}$$
$$-tr\{\tilde{\tilde{\theta}}_k^T\Pi\tilde{\tilde{\theta}}_k\}+\Lambda\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i|i}\right\|^4-\Lambda\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i-1|i-1}\right\|^4+\sum_{i=k-N_0}^{k}\left\|\tilde{\vartheta}_i\right\|^4-\sum_{i=k-N_0}^{k}\left\|\tilde{\vartheta}_{i-1}\right\|^4 \tag{93}$$
$$\leq -\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})\Pi\left\|\tilde{\tilde{\theta}}_k\right\|^2-\frac{(3\alpha_o^4-\alpha_o(1+6\alpha+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)_o}{(\chi_{\min}^2+1)^2(U_M-1)^4}\sum_{i=k-N_0}^{k}\left\|\tilde{z}_{i-1i-1}\right\|^4$$
$$-\alpha_o(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1}-(7+\frac{1}{(\chi_{\min}^2+1)^2})\alpha_o-(\frac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2-5\alpha_o^3)\left\|\tilde{\vartheta}_{k-j-2}\right\|^4+\varepsilon_M^{AE}$$

Since α_h,α_o and η are positive constants which satisfy $0<\alpha_h<\frac{2\Delta\phi_{\min}^2}{3(\Delta\phi_{\min}^2+1)}$, $0<\alpha_o<\frac{(\chi_{\min}^2+1)(\chi_{\min}^2-5)}{21(\chi_{\min}^2+1)^2+3}$ and $0<\eta<\sqrt[4]{\frac{3\alpha_o^3}{(U_M-1)^4(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)}}$ and ε_M^{AE} is defined as $\varepsilon_M^{AE}=(\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2}+1+\alpha_o^4)\left\|\vartheta\right\|^4\frac{\sigma_\upsilon^4}{\underline{\upsilon}^4}$. Then, according to (92) and (93), the first difference of ΔL_J is less than zero provided the following inequalities hold

$$\left\|\tilde{\tilde{\theta}}_k\right\| > \sqrt{\frac{\varepsilon_M^{AE}}{\alpha_h(2-3\alpha_h-\dfrac{2}{\Delta\phi_{\min}^2+1})\Pi}} \equiv B_\theta \qquad \text{or} \tag{94}$$

$$\left\|\tilde{\tilde{z}}_{k-1|k-1}\right\| > \sqrt[4]{\frac{(\chi_{\min}^2+1)^2\varepsilon_M^{AE}}{(3\alpha_o^4-\alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}} \quad \text{or}$$

$$\left\|\tilde{\tilde{\vartheta}}_{k-1}\right\| > \sqrt[4]{\frac{\varepsilon_M^{AE}}{\alpha_o(\dfrac{\chi_{\min}^2-5}{\chi_{\min}^2+1}-(7+\dfrac{1}{(\chi_{\min}^2+1)^2})\alpha_o-(\dfrac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2-5\alpha_o^3)}}$$

Therefore, based on Lyapunov theory [10], value function estimator parameter estimation errors are *UUB* in the mean.

Next, with the estimated $\hat{\bar{\Theta}}_k$ and equation (86), stochastic optimal control for NCS under UDP can be obtained as

$$\hat{u}_k = -(\hat{\bar{\Theta}}_k^{uu})^{-1}\hat{\bar{\Theta}}_k^{uz} z_k \tag{95}$$

D. Closed-Loop System Stability for NCS under UDP

In this section, it will be shown that observer error dynamics and its parameter estimation errors and value function parameter estimation errors are *UUB* in the mean. Further, the closed-loop NCS under UDP will be proven to be *UUB* in the mean.

Theorem 6 *(Convergence of the Optimal Control Signal).* Given the initial conditions for the system states z_0, observer parameter vectors $\hat{\bar{\vartheta}}_0$, value function and its parameter vectors $\hat{\bar{\theta}}_0$ be bounded in the set $\mathbf{\Omega}$, let u_0 be the initial admissible control policy for NCS under UDP with the bounds given by $\|A_{zk}z_k + B_{zk}u_k\|^2 \le k_a\|z_k\|^2$ and $0 < k_a < 1/2$. Let the observer, value function estimator parameters be tuned and estimated control policy be provided by (74), (89) and (95) respectively. Then, there exists positive constants α_o, η given by Theorem 4 and 5, α_h given by Theorem 5 and positive constants b_z, b_e, b_ϑ and b_θ such that the system states z_k, observation error $\tilde{z}_{k|k}$, observer parameter estimation error $\tilde{\tilde{\vartheta}}_k$ and stochastic value function parameter estimation errors $\tilde{\tilde{\bar{\theta}}}_k$, respectively, are all UUB in the mean for all $k \geq k_0 + T$ with ultimate bounds given by $\|z_k\| \leq b_z$, $\|e_{k|k}\| \leq b_z$, $\|\tilde{\tilde{\vartheta}}_k\| \leq b_\vartheta$ and $\|\tilde{\tilde{\bar{\theta}}}_k\| \leq b_\theta$.

Proof: Consider the positive definite Lyapunov candidate function as

$$L = L_D + L_J + L_o \tag{96}$$

where L_D is defined as $L_D = tr\{z_k^T \Omega z_k\}$ with positive definite matrix $\Omega = \dfrac{1}{8B_M^2 K_M^2 (\chi_{\min}^2 + 1)} \mathbf{I}$, L_J (77) with positive matrix Ξ is defined in Theorem 1, L_o (91) with positive matrices Λ, Π are given by Theorem 5 and $\mathbf{I}$ is identity matrix. The first difference of (96) can be represented as $\Delta L = \Delta L_D + \Delta L_J + \Delta L_o$, then considering the first part ΔL (i.e. ΔL_D) by applying NCS under UDP and Cauchy-Schwartz inequality, we have

$$\begin{aligned}
\Delta L_D &= tr\{z_{k+1}^T \Omega z_{k+1}\} - tr\{z_k^T \Omega z_k\} \\
&\leq \Omega \left\| A_{zk} z_k + B_{zk} u_k + B_{zk} K_k \tilde{z}_{k|k} - B_{zk} \tilde{u}_k \right\|^2 - \Omega \|z_k\|^2 \\
&\leq 2\Omega \|A_{zk} z_k + B_{zk} u_k\|^2 + 4\Omega \left\| B_{zk} K_k \tilde{z}_{k|k} \right\|^2 + 4\Omega \|B_{zk} \tilde{u}_k\|^2 - \Omega \|z_k\|^2 .
\end{aligned} \tag{97}$$

Similar to NCS under TCP, there are two methods to obtain stochastic optimal control, and both of them are same. Namely,

$$\tilde{u}_k = \hat{u}_{1k} - \hat{u}_{2k} = -(\hat{\bar{\Theta}}_k^{uu})^{-1} \hat{\bar{\Theta}}_k^{uz} \hat{z}_{k|k} + \frac{1}{2} R^{-1} E\left(B_{zk}^T \frac{\partial \hat{V}(\hat{z}_{k+1|k+1})}{\partial \hat{z}_{k+1|k+1}} \middle| \xi_k \right) = 0 \tag{98}$$

Meanwhile, (98) can be written as

$$\Delta L_D \le -(1-2k_a)\Omega\|z_k\|^2 + 4\Omega\|B_{zk}\tilde{u}_k\|^2 + 4\Omega\left\|B_{zk}K_k\tilde{z}_{k|k}\right\|^2 + 9\Omega\left\|\tilde{z}_{k|k}\right\|^2$$
$$\le -(1-2k_a)\Omega\|z_k\|^2 + 4\Omega\|B_{zk}\tilde{u}_k\|^2 + 4B_M^2K_M^2\Omega\left\|\tilde{z}_{k|k}\right\|^2$$

$$\le -(1-2k_a)\Omega\|z_k\|^2 + 4\Omega B_M^2\left\|-(\hat{\overline{\Theta}}_k^{uu})^{-1}\hat{\overline{\Theta}}_k^{uz}\hat{z}_{k|k} + \frac{1}{2}R^{-1}E\left(B_{zk}^T\frac{\partial\hat{V}(\hat{z}_{k+1|k+1})}{\partial\hat{z}_{k+1|k+1}}\middle|\psi_k\right)\right\|^2 \tag{99}$$

$$+4B_M^2K_M^2\Omega\left\|\tilde{z}_{k|k}\right\|^2$$
$$\le -(1-2k_a)\Omega\|z_k\|^2 + 4B_M^2K_M^2\Omega\left\|\tilde{z}_{k|k}\right\|^2$$

Similar to NCS under TCP, we separate the proof into two cases $\gamma_{k+1}=1$ and $\gamma_{k+1}=0$ as following.

Case 1: $\gamma_{k+1}=1$. (No packet losses)

Combing (77), (93), and (99), then ΔL can be expressed as

$$\Delta L = \Delta L_D + \Delta L_J + \Delta L_o$$
$$\le -(1-2k_a)\Omega\|z_k\|^2 + 4B_M^2K_M^2\Omega\left\|\tilde{z}_{k|k}\right\|^2 - \alpha_h(2-3\alpha_h - \frac{2}{\Delta\phi_{\min}^2+1})tr\{\tilde{\overline{\theta}}_k^T\Pi\tilde{\overline{\theta}}_k\}$$
$$-\frac{(3\alpha_o^4-\alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}{(\chi_{\min}^2+1)^2(U_M-1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4 - \alpha_o(\frac{\sigma_0^2-5}{\sigma_0^2+1} - (7+\frac{1}{(\sigma_0^2+1)^2})\alpha_o$$
$$-(\frac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2 - 5\alpha_o^3)\left\|\tilde{\vartheta}_{k-1}\right\|^4 + (\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2}+1+\alpha_o^4)\|\vartheta\|^4\frac{\sigma_\upsilon^4}{\upsilon^4}$$
$$-\frac{(1-3\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-\alpha_o(1+\alpha_o)))}{(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2$$

$$-6\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2}\right)\left\|\tilde{\vartheta}_k\right\|^2+3\left(\left(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\right)^2+\alpha_o\right)\frac{\sigma_\upsilon^2}{\upsilon^2}\|\vartheta\|^2$$

$$\le -(1-2k_a)\Omega\|z_k\|^2-\alpha_h\left(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1}\right)\left\|\tilde{\tilde{\theta}}_k\right\|^2$$

$$-\frac{(3\alpha_o^4-\alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}{(\chi_{\min}^2+1)^2(U_M-1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$-\alpha_o\left(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1}-\left(7+\frac{1}{(\chi_{\min}^2+1)^2}\right)\alpha_o-\left(\frac{6}{(\chi_{\min}^2+1)^2}+12\right)\alpha_o^2-5\alpha_o^3\right)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$+\left(\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2}+1+\alpha_o^4\right)\|\vartheta\|^4\frac{\sigma_\upsilon^4}{\upsilon^4}$$

$$-\frac{(1-6\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-2\alpha_o(1+\alpha_o)))}{2(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2 \quad (100)$$

$$-6\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2}\right)\left\|\tilde{\vartheta}_k\right\|^2+3\left(\left(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\right)^2+\alpha_o\right)\frac{\sigma_\upsilon^2}{\upsilon^2}\|\vartheta\|^2$$

$$\le -(1-2k_a)\Omega\|z_k\|^2-\alpha_h\left(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1}\right)\left\|\tilde{\tilde{\theta}}_k\right\|^2$$

$$-\frac{(3\alpha_o^4-\alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}{(\chi_{\min}^2+1)^2(U_M-1)^4}\left\|\tilde{z}_{k-1|k-1}\right\|^4$$

$$-\alpha_o\left(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1}-\left(7+\frac{1}{(\chi_{\min}^2+1)^2}\right)\alpha_o-\left(\frac{6}{(\chi_{\min}^2+1)^2}+12\right)\alpha_o^2-5\alpha_o^3\right)\left\|\tilde{\vartheta}_{k-1}\right\|^4$$

$$-6\left(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2}\right)\left\|\tilde{\vartheta}_k\right\|^2-\frac{(1-6\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-2\alpha_o(1+\alpha_o)))}{2(\chi_{\min}^2+1)(U_M-1)^2}\left\|\tilde{z}_{k|k}\right\|^2+\varepsilon_{TM}$$

where $\varepsilon_{TM}=3\left(\left(1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o\right)^2+\alpha_o\right)\frac{\sigma_\upsilon^2}{\upsilon^2}\|\vartheta\|^2+\left(\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2}+1+\alpha_o^4\right)\|\vartheta\|^4\frac{\sigma_\upsilon^4}{\upsilon^4}$.

Case 2: $\gamma_{k+1}=0$. (With packet lost)

After applying assumption 2, (78), (93), and (99), we can derive the ΔL as

$$\Delta L = \Delta L_D + \Delta L_J + \Delta L_o$$

$$\begin{aligned}
&\le -(1-2k_a)\Omega\|z_k\|^2 - \alpha_h(2-3\alpha_h - \frac{2}{\Delta\phi_{\min}^2+1})\|\tilde{\tilde{\theta}}_k\|^2 \\
&- \frac{(3\alpha_o^4 - \alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}{(\chi_{\min}^2+1)^2(U_M-1)^4}\sum_{i=k-N_0}^{k}\|\tilde{z}_{i-1|i-1}\|^4 \\
&- \alpha_o(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1} - (7+\frac{1}{(\chi_{\min}^2+1)^2})\alpha_o - (\frac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2 \\
&- 5\alpha_o^3)\|\tilde{\vartheta}_{k-j-1}\|^4 + (\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2} + 1 + \alpha_o^4)\|\vartheta\|^4\frac{\sigma_\upsilon^4}{\upsilon^4} \qquad (101) \\
&- \frac{(1-6\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2 - 2\alpha_o(1+\alpha_o)))}{2(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\|\tilde{z}_{i|i}\|^2 \\
&- 6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o - \frac{1}{2})\|\tilde{\vartheta}_{k-j-1}\|^2 + 3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\upsilon^2}\|\vartheta\|^2 \\
&\le -(1-2k_a)\Omega\|z_k\|^2 - \alpha_h(2-3\alpha_h - \frac{2}{\Delta\phi_{\min}^2+1})\|\tilde{\tilde{\theta}}_k\|^2 \\
&- \frac{(3\alpha_o^4 - \alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}{(\chi_{\min}^2+1)^2(U_M-1)^4}\sum_{i=k-N_0}^{k}\|\tilde{z}_{i-1|i-1}\|^4 \\
&- \alpha_o(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1} - (7+\frac{1}{(\chi_{\min}^2+1)^2})\alpha_o - (\frac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2 \\
&- 5\alpha_o^3)\|\tilde{\vartheta}_{k-j-2}\|^4 - 6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o - \frac{1}{2})\|\tilde{\vartheta}_{k-j-1}\|^2 \\
&- \frac{(1-6\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2 - 2\alpha_o(1+\alpha_o)))}{2(\chi_{\min}^2+1)(U_M-1)^2}\sum_{i=k-N_0}^{k}\|\tilde{z}_{i|i}\|^2 + \varepsilon_{TM}
\end{aligned}$$

where $\varepsilon_{TM} = 3((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2+\alpha_o)\frac{\sigma_\upsilon^2}{\upsilon^2}\|\vartheta\|^2 + (\frac{3\alpha_o^2(2+3\alpha_o)}{(\chi_{\min}^2+1)^2} + 1 + \alpha_o^4)\|\vartheta\|^4\frac{\sigma_\upsilon^4}{\upsilon^4}$ is a positive constant.

Therefore, ΔL is less than zero when the following inequalities hold

$$\|z_k\| > \sqrt{\frac{\varepsilon_{TM}}{(1-2k^*)}} \equiv b_z \qquad \text{or}$$

$$\|\tilde{z}_{k|k}\| > \sqrt{\frac{2(\chi_{\min}^2+1)\varepsilon_{TM}}{(1-6\eta^2(U_M-1)^2((1-\frac{\chi_{\min}^2}{\chi_{\min}^2+1}\alpha_o)^2-2\alpha_o(1+\alpha_o)))}} \equiv b_e \text{ or}$$

$$\|\tilde{z}_{k-1|k-1}\| > \sqrt[4]{\frac{(\chi_{\min}^2+1)^2\varepsilon_{TM}}{(3\alpha_o^4-\alpha_o(1+6\alpha_o+9\alpha_o^2+\alpha_o^3)(U_M-1)^4\eta^4)}} \qquad \text{or}$$

$$\|\tilde{\vartheta}_k\| > \sqrt{\frac{\varepsilon_{TM}}{6(\frac{\chi_{\min}^2-1}{\chi_{\min}^2+1}\alpha_o-\frac{1}{2})}} \equiv b_\vartheta \qquad \text{or} \qquad (102)$$

$$\|\tilde{\vartheta}_{k-1}\| > \sqrt[4]{\frac{\varepsilon_{TM}}{\alpha_o(\frac{\chi_{\min}^2-5}{\chi_{\min}^2+1}-(7+\frac{1}{(\chi_{\min}^2+1)^2})\alpha_o-(\frac{6}{(\chi_{\min}^2+1)^2}+12)\alpha_o^2-5\alpha_o^3)}} \text{ or}$$

$$\|\tilde{\theta}_k\| > \sqrt{\frac{\varepsilon_{TM}}{\alpha_h(2-3\alpha_h-\frac{2}{\Delta\phi_{\min}^2+1})}} \equiv b_\theta$$

Using the Lyapunov theory [10], the system states, observed state and its parameter estimation errors and value function estimator parameter estimation errors are *UUB* in the mean.

V. SIMULATION RESULTS

In this section, the performances of proposed stochastic optimal control of NCS under both TCP and UDP are evaluated with a single protocol at a time. Meanwhile, the standard optimal controls of NCS under TCP or UDP with known system dynamics and network imperfections are also simulated for comparison.

Example: The continuous-time version of a batch reactor system dynamics are given as [20]:

$$\dot{x} = \begin{bmatrix} 1.38 & -0.2077 & 6.715 & -5.676 \\ -0.5814 & -4.29 & 0 & 0.675 \\ 1.067 & 4.273 & -6.654 & 5.893 \\ 0.048 & 4.273 & 1.343 & -2.104 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 5.679 & 0 \\ 1.136 & -3.146 \\ 1.136 & 0 \end{bmatrix} u \tag{103}$$

$$y = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix} x$$

with $x \in \mathbb{R}^{4\times 1}$ and $u \in \mathbb{R}^{2\times 1}$. Note that this batch reactor example has been developed over years as a benchmark example for NCS, see e.g. [20][21][22].

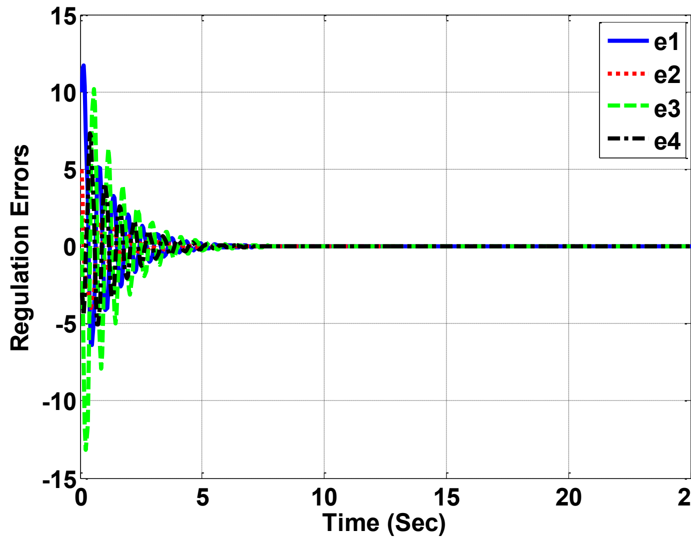


Fig 6. Performance of standard control without network imperfection.

The NCS parameters under TCP are selected as [18]:

1) The sampling time: $T_s = 50ms$;

2) The delay bound is selected as two, i.e. $b = 2$;

3) The random delays: $E(\tau_{sc}) = 35ms$, $E(\tau) = 75ms$;

4) Packet losses follow Bernoulli distribution with $\bar{\gamma} = 0.3$ and $\bar{\upsilon} = 0.2$.

First, we consider the effective of network imperfections (i.e. random delays and packet losses in this paper) for NCS under TCP or UDP. In Figure 5, the standard control inputs $u_k = -\begin{bmatrix} 3.78 & 1.82 & 0.50 & 4.27 \\ -0.28 & 0.98 & -0.91 & 6.48 \end{bmatrix} x_k$ designed by pole placement method can maintain stability of the batch reactor system without any network imperfection as shown in Figure 6. However, this standard control cannot maintain system stable in the mean in presence of network imperfection caused by TCP or UDP protocol as shown in Figures 7 and 8.

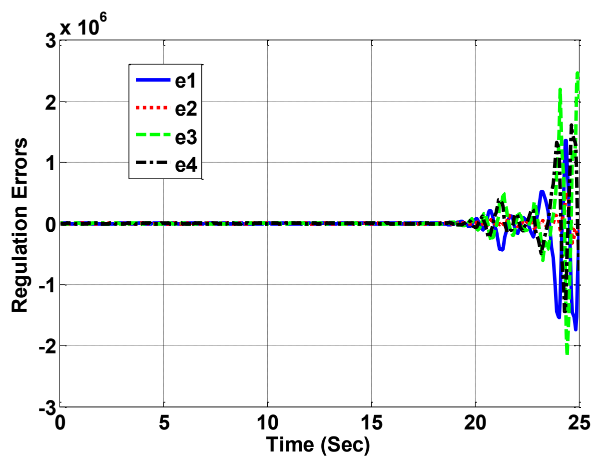


Fig 7. State regulation errors of standard control when network imperfections are present for NCS under TCP

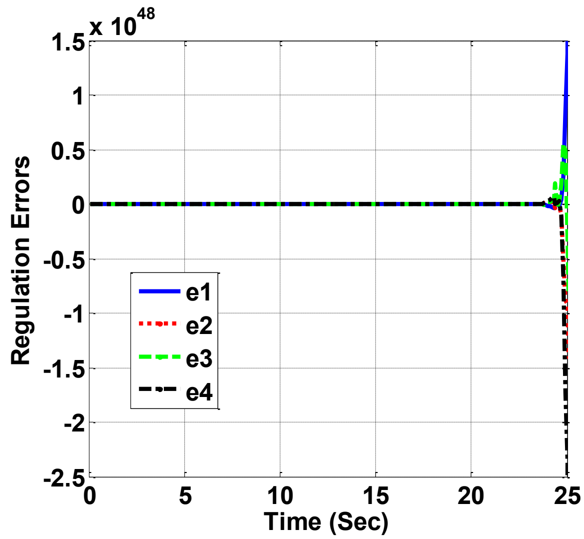


Fig 8. State regulation errors of standard control when network imperfections are present for NCS under UDP

Next, proposed stochastic optimal controller and novel observer designs are applied on NCS under TCP and UDP with unknown system dynamics in presence of network imperfections respectively. The augment state z_k is derived as $z_k = [x_k \ u_{k-1} \ u_{k-2}]^T \in \mathbb{R}^{8\times 1}$ and $\varphi^e = [\hat{z} \ u]^T \in \mathbb{R}^{10\times 1}$. The initial admissible policy for proposed algorithm is selected as $u_0 = -\begin{bmatrix} 0.87 & 0.85 & -0.1 & 1.24 & 0.03 & 0 & 0.13 & 0.01 \\ -1.51 & 0.09 & -2.55 & 2.47 & 0 & 0.08 & -0.05 & 0.52 \end{bmatrix} \hat{z}_{k|k}$ while regression functions for value function estimation is generated as $\{\varphi_1^{e2}, \varphi_1^e \varphi_2^e, \varphi_1^e \varphi_3^e, \ldots, \varphi_2^{e2}, \ldots, \varphi_{10}^{e2}\}$ as [18]. The designed tuning rate for value function estimator is selected as $\alpha_h = 10^{-4}$ for NCS under TCP, and $\alpha_h = 10^{-3}$ for NCS under UDP while initial parameters are set to zeros at the beginning of simulation.

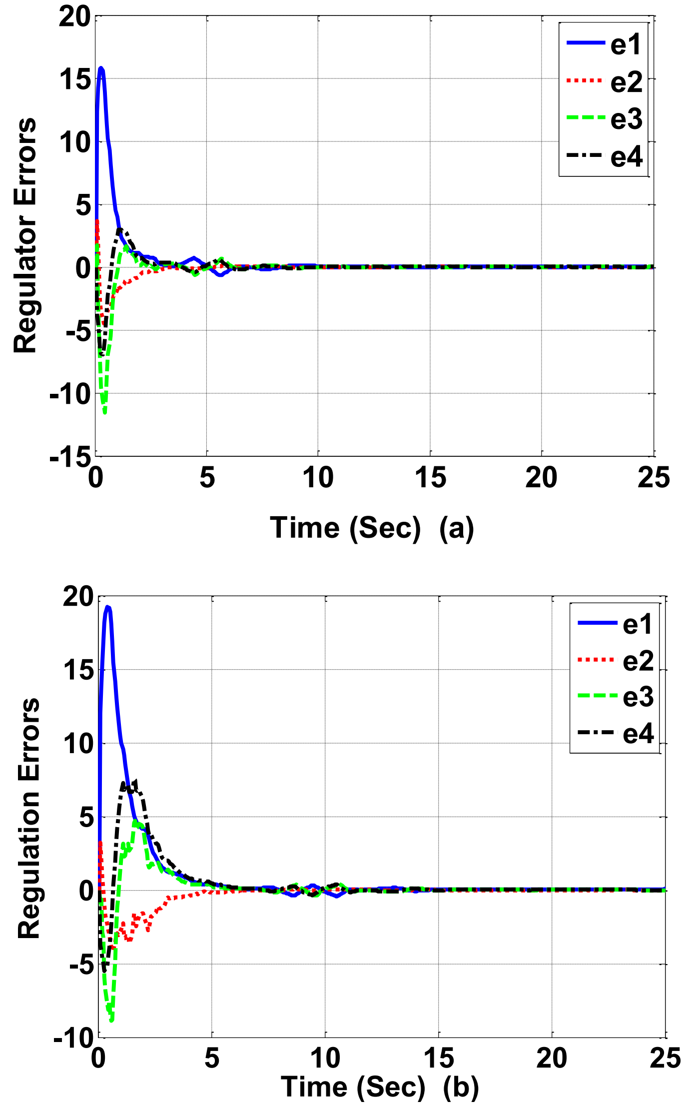


Fig 9. Performance of the proposed stochastic optimal control: state regulation errors of NCS under a) TCP, b) UDP.

The initial parameters of control estimator are chosen to reflect the initial admissible control. On the other hand, regression function for observer is defined as equation (6) and (72), and designed learning rate is defined as $\alpha_o = 10^{-3}$ for NCS under TCP and $\alpha_o = 10^{-2}$ for NCS under UDP when initial parameters of observer are set to zeros. The simulation was run for 500 time steps, and first 100 time steps, exploration noise was added to maintain the persistency of excitation (PE) condition holds (Remark 3).

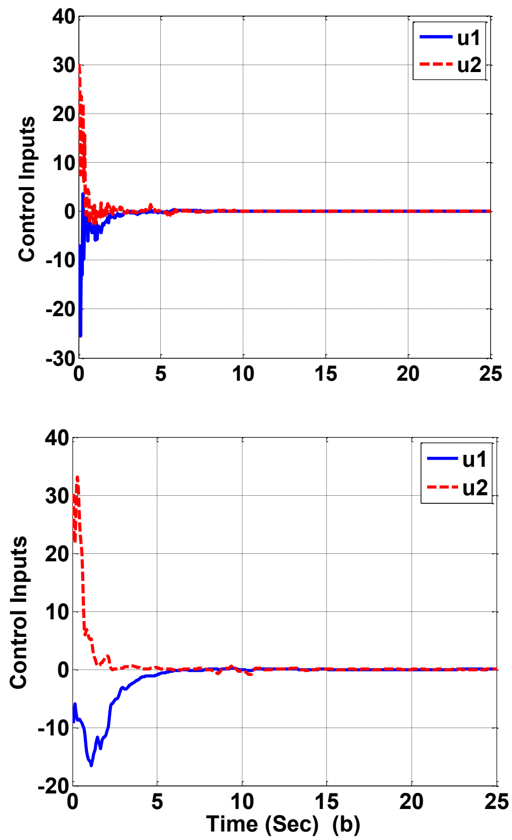


Fig 10. Comparison of control inputs $u = (u_1\ u_2)^T \in \mathbb{R}^{2\times 1}$ for the proposed controller of NCS under: a) TCP, b) UDP.

In Figures 9, 10 and 11, we evaluate the performance of proposed value function estimator and observer based optimal control for NCS under TCP and UDP respectively. Even the dynamics of NCS under TCP and UDP are unknown, the proposed value function estimator and observer based optimal control can still force NCS under TCP and UDP states regulation errors convergence to zero as shown in Figure 9 (a) and (b). However, since acknowledge scheme is used in TCP protocol, compared with UDP protocol more network information (i.e. packet losses indicator between controller and actuator $\boldsymbol{\upsilon}_{k-1}$) can be used in controller design. Meanwhile, only statistical information of packet losses between controller and actuator (i.e. mean value of $\boldsymbol{\upsilon}_k$: $\bar{\boldsymbol{\upsilon}}$ and variance of $\boldsymbol{\upsilon}_k$: σ_{υ}^2.) can be used in stochastic optimal control design of NCS under UDP. Since this statistical information for the UDP is not completely accurate due to lack of feedback, system states of NCS under TCP converge quicker than NCS under UDP.

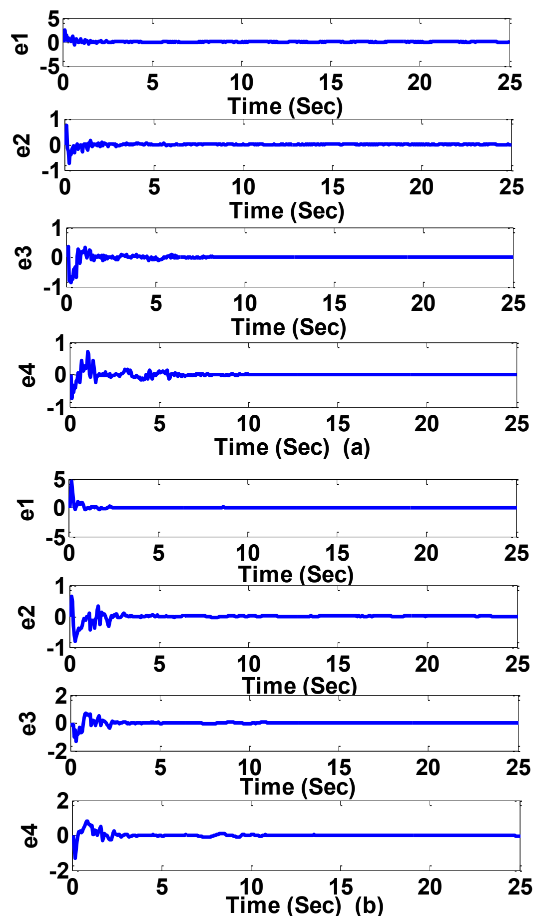


Fig 11. Performance of proposed observer for NCS under: a) TCP, b) UDP.

In Figure 10 (a) and (b), control inputs of the proposed value function estimator and observer based optimal control for NCS under TCP and UDP are shown and compared. Because TCP protocol provides more network information for control inputs design when compared with UDP, NCS under TCP can use smaller control inputs to maintain the system stable in the mean. Note that there is a slightly overshoot at the beginning of Figures 9 (a)(b) and 10 (a)(b) for proposed algorithm. It is because the initial online learning phase needed time to tune the observer and value function estimator to obtain the optimal control performance for NCS under TCP and UDP.

After a short time, proposed scheme will have similar performance even when NCS system dynamics are unknown no matter which ever TCP or UDP protocols are utilized. On the other hand, performance of proposed observer for NCS under TCP and NCS under UDP are evaluated in Figure 11 (a)(b). The proposed observer can force observed system state to converge to actual system state quickly for both NCS under TCP and NCS under UDP. However, performance of proposed observer for NCS under UDP is not on par with the case with TCP which alluded earlier due to inaccurate feedback information with UDP protocol.

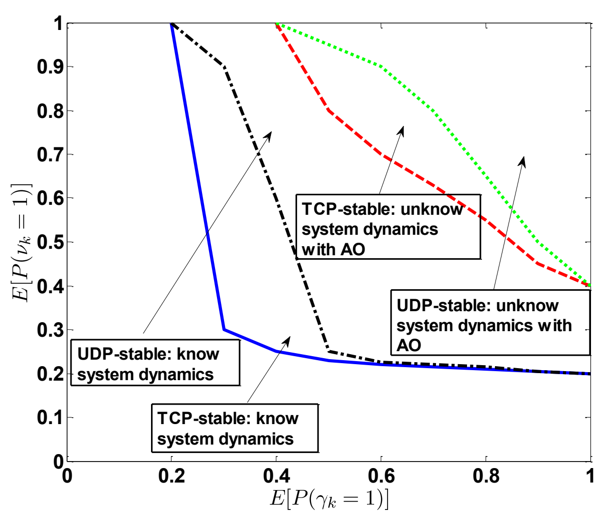


Fig 12. Stability regions of observer for NCS under TCP and NCS under UDP

It is important to know that there is a critical domain of values for the parameters of the Bernoulli arrival processes (i.e. $\bar{\gamma}$ and $\bar{\upsilon}$) which cannot ensure the stability of observer holds. In Figure 12, this stability region for the proposed observer is shown and compared with results in [12]. Since proposed observer does not require knowledge of system dynamics which is needed in [12], the region of proposed method is much tighter than [12]. Also since TCP protocol can provide more network information for the observer design, the stability region of UDP is smaller than TCP.

Based on the results presented in Figures 5 through 12, after a short initial tuning time, proposed value function estimator and observer based stochastic optimal control for NCS under TCP and UDP with uncertain dynamics and network imperfection will have nearly the same performance as that of the conventional optimal control for NCS under TCP and UDP with known system dynamics and network imperfection.

VI. CONCLUSION

In this paper, a novel adaptive dynamics programming scheme consisting of a novel observer, value function estimator is utilized to solve the Bellman equation in real time for obtaining optimal control of NCS under TCP and UDP. By using past input and estimated states, the system dynamics requirement was relaxed while using the estimated states and value function estimation, stochastic optimal control inputs were derived.

An initial admissible control ensured that the system is kept stable in the mean while the observer and value function estimator is tuned. Initial overshoots are observed due to the online tuning phase of observer and value function estimator while they disappear with time quickly. All observer and value function estimator parameters $\hat{\vartheta}_k, \hat{\overline{\Theta}}_k$ were tuned online using proposed update law and Lyapunov theory demonstrated the asymptotically stability in the mean of the closed-loop system for NCS under TCP and the uniformly ultimately boundedness in the mean of the closed-loop system for NCS under UDP respectively.

REFERENCES

[1] Y. Halevi and A. Ray. “Integrated communication and control systems: Part I---Analysis,” J. Dynamic Syst., Measure. Contr., vol. 110, pp. 367-373, 1988.

[2] W. Zhang, M. S. Branicky, and S. Phillips, “Stability of networked control systems,” IEEE Contr. Syst. Mag., vol. 21, pp. 84–99, 2001.

[3] F. Lian, J. Moyne, and D. Tilbury. “Optimal controller design and evaluation for a class of networked control systems with distributed constant delays,” in Proc. Amer. Contr. Conf., pp.3009-3014, 2002.

[4] J. Nilsson, B. Bernhardsson, and B. Wittenmark. “Stochastic analysis and control of real-time systems with random time delays,” Automatica, vol.1, pp. 57-64, 1998.

[5] K. J. Åstrom, Introduction to stochastic control theory, Academic Press, New York, 1970.

[6] P.J. Werbos, “A menu of designs for reinforcement learning over time,” J. Neural Networks Contr., vol. 3, pp. 835-846, 1983.

[7] F.L. Lewis and V.L. Syrmos. Optimal Control. 2nd ed., Wiley, New York, 1995.

[8] J. Si, A.G. Barto, W.B. Powell, and D. Wunsch, Handbook of learning and approximate dynamic programming. Wiley, 2004.

[9] T. Dierks, and S. Jagannathan, “Optimal control of affine nonlinear discrete-time systems with unknown internal dynamics,” in Proc. IEEE Conf. on Decis. and Contr., pp. 6750-6755, 2009.

[10] S. Jagannathan, Neural network control of nonlinear discrete-time systems, CRC Press, 2006.

[11] W.M. Wonham, “On a matrix Riccati equation of stochastic control,” SIAM J. Contr., vol. 6, pp. 681-697, 1968.

[12] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, “Foundations of control and estimation over lossy networks,” Proc. IEEE, vol. 95, no.1, pp. 163-187, 2007.

[13] P. Antsaklis, and J. Baillieul, “Special issue on networked control systems,” IEEE Trans. Automat. Contr., vol. 49, pp. 1421-1423, 2004.

[14] S. S. Hu, and Q. X. Zhu, “Stochastic optimal control and analysis of stability of networked control systems with long delay,” Automatica, vol. 39, pp. 1877-1884, 2003.

[15] T. Dierks, and S. Jagannathan, “Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update,” IEEE Trans. Neur. Net., conditionally accepted, 2011.

[16] A. I. Petros, and S. Jing, Robust adaptive control, PTR Prentice-Hall Press, 1996.

[17] P. J. Werbos, “Approximate dynamic programming for real-time and neural modeling,” Handbook of Intelligent Control, New York, Van Nostrand Reinhold, 1992.

[18] H. Xu, and S. Jagannathan, “Stochastic optimal control of unknown linear networked control system using Q-learning methodology,” in Proc. Amer. Contr. Conf., pp.2819-2824, 2011.

[19] M. Green, and J. B. Moore, “Persistency of excitation in linear systems,” Syst. and Contr. Lett., vol. 7, pp. 351-360, 1986.

[20] D. Carnevale, A. R. Teel, and D. Nesic, "A Lyapunov proof of improved maximium allowable transfer interval for networked control systems," IEEE. Trans. Automa. Contr., vol. 52, pp. 892-897, 2007.
[21] D. B. Dacic, and D. Nesic, "Quadratic stabilization of linear networked control system via simultaneous protocol and controller design," Automatica, vol. 43, pp. 1145-1155, 2007.
[22] G. C. Walsh, H. Ye, and L. Bushnell, "Stability analysis of networked control systems," in Proc. Amer. Contr. Conf., pp. 2876-2880, 1999.
[23] A. Goldsmith, Wireless Communication. Cambridge University Press, Cambridge, U.K. 2003.
[24] N. Boizot, and E. Busvelle, "Adaptive-gain observers and application," Nonlinear Observers and Application, vol. 363, pp. 71-114, 2007.
[25] D. Tse, and P. Viswanath, Fundementals of wireless communication, Cambridge university press, 2005.
[26] 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, and Cybern. B., vol. 41, pp. 14-25, 2011.
[27] D. P. Bertsekas, Dynamic programming and optimal control/approximate dynamic programming, Athena Scientific, 2007

PAPER V

A CROSS LAYER APPROACH TO THE NOVEL DISTRIBUTED SCHEDULING PROTOCOL AND EVENT-TRIGGERED CONTROLLER DESIGN FOR CYBER PHYSICAL SYSTEMS

H. Xu and S. Jagannathan

Abstract - In the next generation Cyber Physical Systems (CPS), multiple real-time dynamic systems are connected through a shared communication network. For such CPS, the existing network protocols (e.g. Centralized/Distributed Scheduling) cannot be implemented since the behavior of real-time dynamic systems is ignored though it needs to be considered during the protocol design. Therefore, in this paper, a novel distributed scheduling protocol design via cross-layer approach is proposed to optimize the performance of CPS by maximizing the utility function which is generated by using the information from both application layer (i.e. event-triggered controllers for each real-time system) and network layer. Subsequently, a novel adaptive model based optimal event-triggered control scheme is derived for each real-time dynamic system with unknown system dynamics in the application layer. Compared with traditional scheduling algorithms, the proposed distributed scheduling scheme via cross-layer approach can not only allocates the network resources efficiently but also improves the performance of the overall real-time dynamic system. Simulation results are included to illustrate the proposed cross-layer co-design.

Keywords- Cyber Physical Systems (CPS); Distributed Scheduling; Event-triggered Control; Cross Layer

I. INTRODUCTION

In past few years, intelligent control [1] and communication network [2] have been two of the fast-growing research areas. Most recently, a number of researchers [3-4] realize that combining these two areas can bring more significant advantages for both modern control and communication network such as saving installation costs, increasing adaptability, reliability and usability. For distinguishing from traditional control or communication network, this novel class of system has been referred as Cyber Physical System (CPS) [5-7]. In CPS, since the control and communication subsystems are tied together closely, novel CPS-based control and communication schemes have to be designed by considering the linkage between control and communication subsystems. Therefore, incorporating the effects from fixed communication network, authors in [6] proposed a cyber physical control scheme to maintain the stability of control system part of the CPS. In [7], from communication network protocol side, authors evaluated the performance of widely used protocol (i.e. IEEE 802.15.4) for CPS.

However, most of these works [5-7] have not considered real-time interactions between control and communication subsystem. A revolutionary algorithm for CPS should utilize real-time interaction to optimize the performance of both control and communication subsystems. For well known open systems interconnection (OSI) representation [8], control subsystem belongs to application layer while the communication network protocol is included in the network and data link layers. Therefore, to consider the interaction among different layers properly, cross-layer design [9-12] is necessary. In [9-10], authors have shown that cross-layer design can attain performance gains by exploiting the dependence between protocol layers compared with traditional individual layered protocol design. However, most cross-layer designs are implemented for data link and physical layers [11-12] where application layer is neglected. For CPS, application and data link layers should be designed jointly via cross-layer approach. In other words, the control design at the application layer and the protocol design for communication at the data link layer have to be considered jointly.

On the other hand, the distributed scheduling is critical for the communication protocol design [13]. Compared with traditional centralized scheduling [14], the main advantage of distributed scheduling is that it does not require a central processor to deliver the schedules after collecting information from all the communication links in the network. According to IEEE 802.11 standard [13], carrier sense multiple access (CSMA) protocol is introduced to schedule communication links in a distributed manner where a communication link wishing to transmit does so only if it does not hear an on-going transmission from the network. Further, authors [15] derived a throughput-optimal distributed scheduling algorithm and proven that even distributed scheduling scheme can still achieve the throughput maximization. However, since random access scheme is widely used in most CSMA-based distributed scheduling [15-18] and these schemes focus on improving data link layer performance alone which in turn affect the performance of application layer (i.e. control system), these protocols are neither optimal nor suitable for CPS since they can degrade the performance of CPS overall.

Meanwhile, for the application layer, the optimal control design is the most challenging issue. It is important to note there are two main drawbacks for traditional optimal control schemes [19]. First drawback is that full knowledge of system dynamics are needed and optimal control is solved backward-in-time which is not suitable for hardware implementation. Second, these traditional optimal control schemes are sampled periodically which require significant network resources (e.g. bandwidth) for communication to transmit information between a sensor and a controller potentially causing congestion. The performance of control system in the application layer degrades significantly due to congestion.

Therefore, this paper proposed a novel cross layer scheme for CPS which includes an event triggered controller design for control system design in the application layer and a distributed scheduling algorithm for communication network in the data link layer. The main contributions of this paper include: 1) the distributed scheduling via cross layer approach which improves the performance of CPS by minimizing the cost function from both data link and application layers; 2) novel adaptive model-based optimal event-triggered control scheme which is designed in a forward-in-time manner and without the knowledge of system dynamics. Compared with time-based periodic sampling, the proposed event-triggered scheme is based on events which are initiated not only by the control system but also the shared communication network performance.

This paper is organized as follows. First, Section II presents the background for multiple pairs CPS and event-triggered control schemes. A novel cross-layer design which includes distributed scheduling and adaptive model-based optimal event-triggered control is proposed for multiple CPS pairs in Section III. Section IV illustrates the effectiveness of proposed schemes via numerical simulation and Section V provides concluding remarks.

II. BACKGROUND

A. Multiple Cyber Physical System Pairs

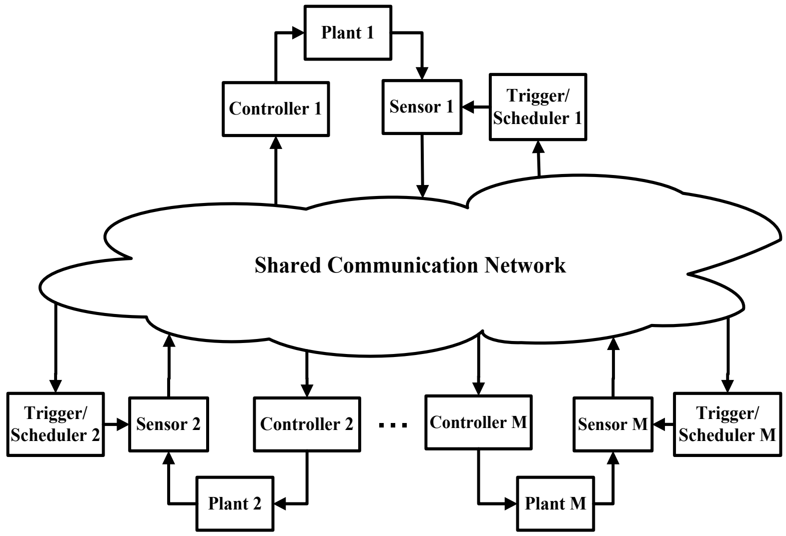


Figure 1. Multiple pairs CPS

The basic structure of multiple CPS pairs is shown in Figure 1, where multiple pairs of real-time subsystems communicate to their respective controllers through a shared communication network (e.g. IEEE 802.11). Obviously, the shared communication network can affect performance of control systems. For instance, when shared communication network is congested due to improper scheduling, system cannot even maintain stability since the information from the subsystem cannot be transmitted to the controller successfully and may experience undue delays. Therefore, a novel cross-layer design is needed for CPS. Without loss of generality, multiple pairs CPS are assumed to be homogeneous in this paper. On the other hand, to save the network resources, event-triggered system is used instead of a traditional time-driven sampling. Next, the background of event-triggered system is given.

B. Event-Triggered System

Recently, event-triggered control system has been a topic of significant interest for CPS due to its network benefits [20-22]. In Figure 2, the basic structure of Zero-Order-Hold (ZOH) event-triggered system is shown first. Compared with traditional system, a trigger is included in the sensor to decide when to sense and transmit the system information (e.g. system state x_k). For sake of simplicity, the CPS subsystem is considered to be linear discrete-time invariant. Since multiple CPS pairs are homogeneous, *lth* CPS pair can be represented as

$$x_{l,k+1} = A_l x_k + B_l u_{l,k} \tag{1}$$

where $x_{l,k} \in \Re^n, u_{l,k} \in \Re^m$ are *lth* CPS system states and control inputs respectively, and $A_l \in \Re^{n \times n}, B_l \in \Re^{n \times m}$ denote system matrices for *lth* CPS pair.

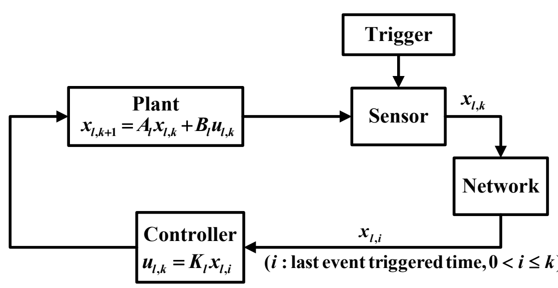


Figure 2. ZOH Event-triggered System

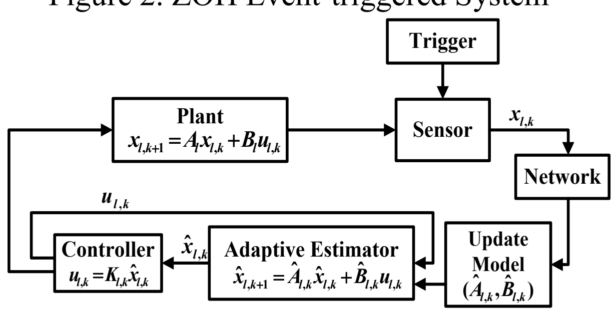


Figure 3. Adaptive model-based event-triggered system

It is important to note that controller will hold the last received system state vector until a new system state vector is received. Obviously, the error between the system states used in the controller and actual values might increase quickly in ZOH event-triggered control scheme. For overcoming this drawback, a novel adaptive model-based event-triggered system is proposed and the basic structure is shown in Figure 3. Besides event trigger, an adaptive model $(\hat{A}_{l,k}, \hat{B}_{l,k})$ is used for the controller to estimate the system state vector when controller has not received any information from the sensor. The estimated system state vector can be represented as

$$\hat{x}_{l,k+1} = \hat{A}_{l,k}\hat{x}_{l,k} + \hat{B}_{l,k}u_{l,k} \tag{2}$$

with $\hat{A}_{l,k} \in \Re^{n\times n}, \hat{B}_{l,k} \in \Re^{n\times m}$ are the adaptive model of lth CPS pair at time kT_s . It is important to note that the adaptive model will be updated once when the most recent system state vector is received at the controller. Eventually, the adaptive model and estimated system states will converge close to the actual system and actual system states respectively, which in turn improves the performance of event-triggered control system.

III. NOVEL CROSS-LAYER DESIGN FOR CPS

In this section, novel cross-layer design is proposed for multiple pairs of CPS. First the main idea of cross-layer design and classical ZOH event-triggered control scheme are introduced. Subsequently, a novel adaptive model-based optimal event-triggered control is derived. Eventually, distributed scheduling is proposed to optimize multiple CPS pairs by minimizing a cost function which includes the information from both application and data link layers.

A. Cross-Layer Design

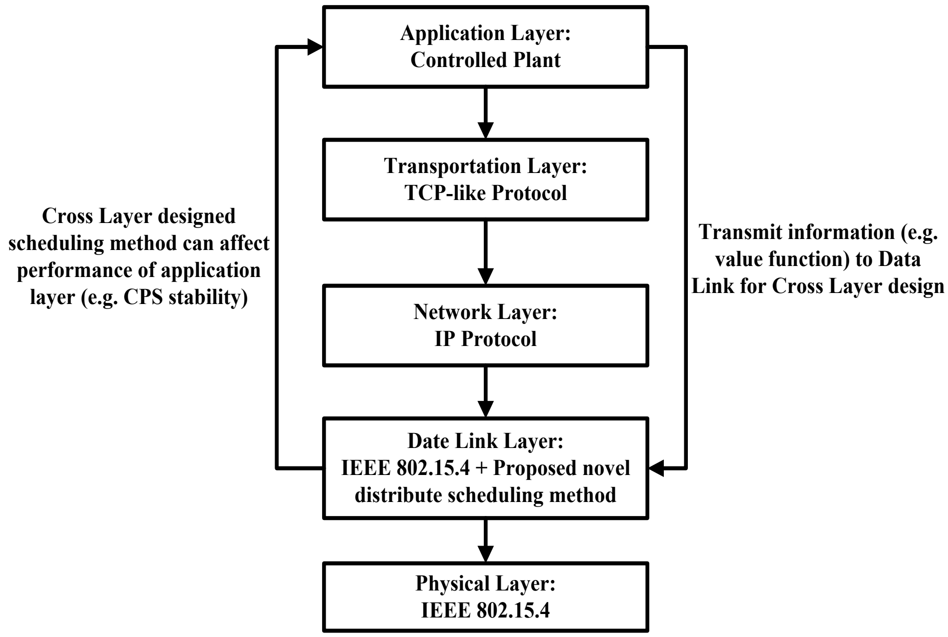


Figure 4. The framework of multiple pairs CPS cross-layer design

In the proposed scheme, the novel event-triggered control design and distributed scheduling protocol are implemented into all CPS pairs which are sharing the communication network. Each CPS pair tunes its adaptive model-based optimal event-triggered controller design by using the proposed distributed scheduling algorithm, computes its value function based on tuned control design, and transmits the information to the data link layer. Then, data link layer can update the scheduling of the CPS pair based on network traffic payload from data link layer and the value function information received from the application layer. The cross-layer design framework of multiple CPS pair is shown in Figure 4.

B. ZOH Event-Triggered Control Design

Different from periodic sampling scheme, the ZOH event-triggered controller might not receive the system state at every sampling time instant. Hence the controller will hold the latest received system state vector for control input design until a new state vector is received due to an event caused by condition when state measurement error e_k^{ZOH} exceeds the threshold. Now, without loss of generality, consider the linear discrete-time (1) of lth CPS pair, and the ZOH event-triggered control input given by

$$u_{l,k} = \begin{cases} K_l x_{l,k} & \text{event is initiated} \\ K_l x_{l,i} & \text{event is not initiated} \end{cases}, \quad 0 < i \le k \tag{3}$$

where $x_{l,i}$ is the latest lth CPS state measurement at the time iT_s due to an event and K_l is a stabilizing control gain matrix for lth CPS system given by (1) with known system matrices (A_l, B_l). After substituting control input $u_{l,k}$ from (3), the closed-loop system dynamics of lth CPS pair due to the ZOH event-triggered control input can be expressed as

$$x_{l,k+1} = (A_l + B_l K_l)x_{l,k} - B_l K_l e_{l,k}^{ZOH} \tag{4}$$

with $e_{l,k}^{ZOH} = \begin{cases} 0 & \text{event is initiated} \\ x_{l,k} - x_{l,i} & \text{event is not initiated} \end{cases}$.

Obviously, if holding time of ZOH event-triggered control system is longer, state measurement error $e_{l,k}^{ZOH}$ in (4) might be larger which can affect the stability of the system. Therefore, a threshold [28] is derived for $e_{l,k}^{ZOH}$ to ensure the stability.

Theorem 1 *(ZOH Event Triggering Condition)*: In linear discrete-time event-triggered CPS system (1), the event should be triggered and controller should be updated when the following is not satisfied

$$\left\| e_{l,k}^{ZOH} \right\|^2 \le \sigma_l \frac{\left\| Q_l \right\|}{\left\| B_l K_l \right\|^2 \left\| P_l \right\|} \left\| x_{l,k} \right\|^2, \quad \text{for} \quad i = 1,2,...,M \tag{5}$$

with $0 < \sigma_l < 1$ and Q_l, P_l, K_l are positive definite matrices and designed control gain for lth CPS pair which are also the solution of following equation

$$2(A_l + B_l K_l)^T P_l (A_l + B_l K_l) - P_l = -Q_l \tag{6}$$

Proof: Consider Lyapunov function candidate as $V_{ZOH} = x_{l,k}^T P_l x_{l,k}$. Using lth CPS system representation (4) and Cauchy-Schwartz inequality, the first difference of Lyapunov function candidate can be derived as

$$\begin{aligned} \Delta V_{ZOH} &= x_{l,k+1}^T P_l x_{l,k+1} - x_{l,k}^T P_l x_{l,k} \\ &= [(A_l + B_l K_l) x_{l,k} - B_l K_l e_{l,k}^{ZOH}]^T P_l [(A_l + B_l K_l) x_{l,k} \\ &\quad - B_l K_l e_{lk}^{ZOH}] - x_{l,k}^T P_l x_{l,k} \end{aligned} \tag{7}$$

$$\begin{aligned} &\le x_{l,k}^T [2(A_l + B_l K_l)^T P_l (A_l + B_l K_l) - P_l] x_{l,k} \\ &\quad + 2(e_{l,k}^{ZOH})^T K_l^T B_l^T P_l B_l K_l e_{lk}^{ZOH} \\ &\le -\left\| Q_l \right\| \left\| x_{l,k} \right\|^2 + 2\left\| B_l K_l \right\|^2 \left\| P_l \right\| \left\| e_{l,k}^{ZOH} \right\|^2 \end{aligned}$$

Applying the introduced event triggering condition (5), the first difference of the Lyapunov function candidate is given by

$$\begin{aligned} \Delta V_{ZOH} &\le -\left\| Q_l \right\| \left\| x_{l,k} \right\|^2 + 2\left\| B_l K_l \right\|^2 \left\| P_l \right\| \left\| e_{l,k}^{ZOH} \right\|^2 \\ &\le -(1 - \sigma_l) \left\| Q_l \right\| \left\| x_{l,k} \right\|^2 \end{aligned} \tag{8}$$

after applying the event trigger condition (5) where $0 < \sigma_l < 1$ and Q_l is obtained from (6). Then ΔV_{ZOH} is negative definite while V_{ZOH} is positive definite for all CPS pairs (i.e. $l = 1,2,...,M$). Therefore, the ZOH event-triggered closed-loop system is globally asymptotically stable all CPS pairs. In other words, as $k \to \infty$, $x_{l,k} \to 0$ for $l = 1,2,...,M$.

Remark 1: In ZOH, the system is running open-loop based on control input derived by the system state vector which is received previously.

It is important to note that although ZOH event-triggered control can save the communication network resource compared with time-driven control, its efficiency is low since measured state and control input are held as constants during any two events. On the other hand, the optimality of event-triggered control design should be considered carefully which is neglected in this section. Therefore, a novel adaptive model-based optimal event-triggered control scheme is proposed in next section. It is important to note that proposed scheme cannot only estimate system state vector and update control inputs during any two events by using adaptive model, but also consider the optimality of event-triggered system.

C. Adaptive Model-Based Optimal Event-Triggered Control Design

In this subsection, a novel adaptive model-based optimal event-triggered control is derived. First, using estimation [23] and adaptation [24] techniques, the dynamic system state vector is estimated by using an adaptive model which is updated when an event is initiated. Subsequently, adaptive model-based optimal control is proposed for CPS based on the state vector which can maintain the stability even with unknown system dynamics. Eventually, the convergence proof is given for proposed algorithm.

Without loss of generality, lth CPS pair is selected to explain the adaptive model-based optimal event-triggered control as follows.

1) Adaptive State Estimator Design

According to event-triggered control schemes [20-22], the system dynamics of the adaptive state estimator will be updated only when the event is initiated and sensed system states are received at the controller. Recalling (1) and (2), for lth CPS pair the event-triggered control system and adaptive state estimator with received information can be represented as

$$x_{l,k+1} = A_l x_{l,k} + B_l u_{l,k} = \theta_l^T z_{l,k} \tag{9}$$

$$\hat{x}_{l,k+1} = \hat{A}_{l,k} x_{l,k} + \hat{B}_{l,k} u_{l,k} = \hat{\theta}_{l,k}^T z_{l,k} \tag{10}$$

with $\theta_l = [A_l \;\; B_l]^T$ and $\hat{\theta}_l = [\hat{A}_{l,k} \;\; \hat{B}_{l,k}]^T$ denote the target and estimated system dynamics of the lth CPS pair respectively, and $z_{l,k} = [x_{l,k}^T \;\; u_{l,k}^T]^T$ represents the augmented state vector.

Then, the state estimator error dynamics $e_{l,k+1}$ can be derived as

$$e_{l,k+1} = x_{l,k+1} - \hat{x}_{l,k+1} = \theta_k^T z_{l,k} - \hat{\theta}_{l,k}^T z_{l,k} = \widetilde{\theta}_{l,k}^T z_{l,k} \tag{11}$$

where $\widetilde{\theta}_{l,k} = \theta_l - \hat{\theta}_{l,k} = [\widetilde{A}_{l,k} \;\; \widetilde{B}_{l,k}]^T$ is the parameter estimator error. Next, define the update law for lth CPS pair estimated unknown parameter vector $\hat{\theta}_{l,k}$ as

$$\hat{\theta}_{l,k+1} = \hat{\theta}_{l,k} + \alpha_{l,e} \gamma_{l,k+1} z_{l,k} e_{l,k+1}^T \tag{12}$$

with $\alpha_{l,e}$ is the tuning parameter satisfying $0 < \alpha_{l,e} < 1$ and $\gamma_{l,k+1}$ is an indicator to present event trigger condition, i.e.

$$\gamma_{l,k+1} = \begin{cases} 1 & \text{event is initiated} \\ 0 & \text{event is not initiated} \end{cases} \tag{13}$$

Meanwhile, lth pair CPS adaptive parameter estimation error dynamics $\widetilde{\theta}_{l,k}$ can be expressed as

$$\widetilde{\theta}_{l,k+1} = \widetilde{\theta}_{l,k} - \alpha_{l,e} \gamma_{l,k+1} z_{l,k} e_{l,k+1}^T \tag{14}$$

Compared with traditional adaptive estimator schemes [23] where the updates are taken periodically, event-based non-periodic tuning law is used here which tunes the state estimator in a non periodic manner. Next, the convergence of lth CPS adaptive parameter estimation error $\widetilde{\theta}_{l,k}$ is demonstrated.

Theorem 2 *(Convergence of lth CPS adaptive parameter estimation errors):* Let the proposed adaptive estimator for the lth CPS pair be defined as (10), and its parameter update law be given by (12). Then, there exist positive constant tuning parameter $\alpha_{l,e}$ for lth CPS pair such that the parameter estimation errors $\tilde{\theta}_{l,k}$ (14) converge to zero asymptotically as time $k \to \infty$.

Proof: Consider the positive definite Lyapunov candidate function as

$$V_{l,\theta} = tr\{\sum_{i=k-\tau}^{k} \tilde{\theta}_{l,i}^T \tilde{\theta}_{l,i}\} \tag{15}$$

with $tr\{.\}$ is the matrix trace operator, $t = (k+1)T_s$ is the latest update time of the adaptive estimator for the lth CPS pair during the interval $[(k-\tau)T_s, (k+1)T_s]$ (i.e. the latest event-triggered time, and $\gamma_{l,k-\tau} = \gamma_{l,k-\tau+1} = \cdots = \gamma_{l,k} = 0, \gamma_{l,k+1} = 1$). Then the first difference can be expressed as $\Delta V_{l,\theta} = tr\{\sum_{i=k+1-\tau}^{k+1} \tilde{\theta}_{l,i}^T \tilde{\theta}_{l,i}\} - tr\{\sum_{i=k-\tau}^{k} \tilde{\theta}_{l,i}^T \tilde{\theta}_{l,i}\}$.

Using (15), and applying the Cauchy-Schwartz inequality reveals that

$$\begin{aligned}
\Delta V_{l,\theta} &= tr\{\sum_{i=k+1-\tau}^{k+1} \tilde{\theta}_{l,i}^T \tilde{\theta}_{l,i}\} - tr\{\sum_{i=k-\tau}^{k} \tilde{\theta}_{l,i}^T \tilde{\theta}_{l,i}\} \\
&= tr\{\sum_{i=k-\tau}^{k} (\tilde{\theta}_{l,i} - \alpha_{l,e}\gamma_{l,i+1} z_{l,i} e_{l,i+1}^T)^T (\tilde{\theta}_{l,i} - \alpha_{l,e}\gamma_{l,i+1} z_{l,i} e_{l,i+1}^T)\} \\
&\quad - tr\{\sum_{i=k-\tau}^{k} \tilde{\theta}_{l,i}^T \tilde{\theta}_{l,i}\} \\
&= tr\{(\tilde{\theta}_{l,k} - \alpha_{l,e} z_{l,k} e_{l,k+1}^T)^T (\tilde{\theta}_{l,k} - \alpha_{l,e} z_{l,k} e_{l,k+1}^T)\} - tr\{\tilde{\theta}_{l,k-\tau}^T \tilde{\theta}_{l,k-\tau}\}
\end{aligned} \tag{16}$$

Since $\gamma_{l,k-\tau} = \gamma_{l,k-\tau+1} = \cdots = \gamma_{l,k} = 0, \gamma_{l,k+1} = 1$ and equation (14), the parameter vector of the adaptive estimator has not been updated during $[(k-\tau)T_s, kT_s]$, i.e.

$$\tilde{\theta}_{l,k-\tau} = \tilde{\theta}_{l,k-\tau+1} = \cdots = \tilde{\theta}_{l,k} \tag{17}$$

Next, according to parameter estimation error dynamics (11) for the lth CPS pair and equation (17), the first difference of Lyapunov function candidate $\Delta V_{l,\theta}$ can be derived as

$$\Delta V_{l,\theta} = tr\{(\tilde{\theta}_{l,k} - \alpha_{l,e} z_{l,k} e_{l,k+1}^T)^T (\tilde{\theta}_{l,k} - \alpha_{l,e} z_{l,k} e_{l,k+1}^T)\} - tr\{\tilde{\theta}_{l,k-\tau}^T \tilde{\theta}_{l,k-\tau}\}$$
$$= tr\{\tilde{\theta}_{l,k}^T \tilde{\theta}_{l,k}\} - 2\alpha_{l,e} tr\{\tilde{\theta}_{l,k}^T z_{l,k} e_{l,k+1}^T\} + \alpha_{l,e}^2 tr\{(z_{l,k} e_{l,k+1}^T)^T z_{l,k} e_{l,k+1}^T\}$$

$$= -2\alpha_{l,e} tr\{\tilde{\theta}_{l,k}^T z_{l,k} e_{l,k+1}^T\} + \alpha_{l,e}^2 tr\{(z_{l,k} e_{l,k+1}^T)^T z_{l,k} e_{l,k+1}^T\}$$
$$= -2\alpha_{l,e} tr\{\tilde{\theta}_{l,k}^T z_{l,k} z_{l,k}^T \tilde{\theta}_{l,k}\} + \alpha_e^2 tr\{(z_{l,k} z_{l,k}^T \tilde{\theta}_{l,k})^T z_{l,k} z_{l,k}^T \tilde{\theta}_{l,k}\}$$

$$- tr\{\tilde{\theta}_{l,k}^T \tilde{\theta}_{l,k}\}$$
$$= -2\alpha_{l,e} tr\{\tilde{\theta}_{l,k}^T z_{l,k} z_{l,k}^T \tilde{\theta}_{l,k}\} + \alpha_{l,e}^2 tr\{\tilde{\theta}_{l,k}^T z_{l,k}^T z_{l,k} z_{l,k} z_{l,k}^T \tilde{\theta}_{l,k}\}$$

$$\begin{aligned} &\le -2\alpha_{l,e} \|z_{l,k}\|^2 tr\{\tilde{\theta}_{l,k}^T \tilde{\theta}_{l,k}\} + \alpha_{l,e}^2 \|z_{l,k}\|^4 tr\{\tilde{\theta}_{l,k}^T \tilde{\theta}_{l,k}\} \\ &\le -\alpha_{l,e} \|z_{l,k}\|^2 (2 - \alpha_{l,e} \|z_{l,k}\|^2) \|\tilde{\theta}_{l,k}\|^2 \end{aligned} \tag{18}$$

Since the tuning parameter $\alpha_{l,e}$ is positive value which satisfies $0 < \alpha_{l,e} < \dfrac{1}{\|z_{l,k}\|^2 + 1}$, then according to (18), $\Delta V_{l,\theta}$ is negative definite and $V_{l,\theta}$ is positive definite. Therefore, the lth CPS pair adaptive parameter estimation errors $\tilde{\theta}_{l,k}$ are asymptotically stable. In other words, as $k \to \infty$, $\tilde{\theta}_{l,k} \to 0$.

2) Optimal Event-triggered Control System Design

In this part, optimal design for adaptive model-based event-triggered control is given in a detailed manner. First, a value function is defined and approximated adaptively. Subsequently, the optimal event-triggered control scheme is derived by using the information from the estimated value function and adaptive model. Eventually, the convergence proof of proposed scheme is derived.

a) Value Function Setup

According to the optimal control theory [19], the infinite-horizon value function of the adaptive model-based event-triggered control system for lth CPS pair can be defined as

$$V(x_{l,k}) = \sum_{i=k}^{\infty}(x_{l,i}^T Q_l x_{l,i} + u_{l,i}^T S_l u_{l,i}) = x_{l,k}^T P_l x_{l,k} \tag{19}$$

where P_l being the solution to the Algebraic Riccati Equation (ARE) [19] of lth CPS pair adaptive model-based event-triggered control system. Then, the Hamiltonian for the system is represented as

$$H(x_{l,k}, u_{l,k}) = r(x_{l,k}, u_{l,k}) + V(x_{l,k+1}, u_{l,k+1}) - V(x_{l,k}, u_{l,k}) \tag{20}$$

with $r(x_{l,k}, u_{l,k}) = x_{l,k}^T Q_l x_{l,k} + u_{l,k}^T S_l u_{l,k}$ is a one-step cost-to-go function. Based on the standard optimal control theory [19], the optimal control input is given by the stationary condition, $\partial H(x_{l,k}, u_{l,k}) / \partial u_{l,k} = 0$ which yields

$$u_{l,k}^* = K_l^* x_k = -(R_l + B_{l,k}^T P_l B_{l,k})^{-1} B_{l,k}^T P_l A_{l,k} x_{l,k} \tag{21}$$

where K_l^* is the optimal control gain.

Remark 2: It is important to note that solving optimal control in (20) requires the system dynamics which is not typically known accurately in practical systems.

On the other hand, we can define the optimal action dependent value function as

$$V(x_{l,k}) = [x_{l,k}^T \;\; u_{l,k}^T] \Theta_l [x_{l,k}^T \;\; u_{l,k}^T]^T \tag{22}$$

Then, similar to [26], using Bellman equation and value function definition (13), substituting value function into the Bellman equation results in

$$\begin{bmatrix} x_{l,k} \\ u_{l,k} \end{bmatrix}^T \Theta_l \begin{bmatrix} x_{l,k} \\ u_{l,k} \end{bmatrix} = r(x_{l,k}, u_{l,k}) + V(x_{l,k+1})$$
$$= \begin{bmatrix} x_{l,k} \\ u_{l,k} \end{bmatrix}^T \begin{bmatrix} A_{l,k}^T P_l A_{l,k} + Q_l & A_{l,k}^T P_l B_{l,k} \\ B_{l,k}^T P_l A_{l,k} & B_{l,k}^T P_l B_{l,k} + S_l \end{bmatrix} \begin{bmatrix} x_{l,k} \\ u_{l,k} \end{bmatrix} \tag{23}$$

Therefore, matrix Θ_l can be expressed as

$$\Theta_l = \begin{bmatrix} \Theta_{l,k}^{xx} & \Theta_{l,k}^{xu} \\ \Theta_{l,k}^{ux} & \Theta_{l,k}^{uu} \end{bmatrix} = \begin{bmatrix} A_{l,k}^T P_l A_{l,k} + Q_l & A_{l,k}^T P_l B_{l,k} \\ B_{l,k}^T P_l A_{l,k} & B_{l,k}^T P_l B_{l,k} + S_l \end{bmatrix} \tag{24}$$

Next, according to [19] and (24), the optimal control gain for lth adaptive model-based event-triggered control system pair can be represented in terms of value function parameters, Θ_l , as

$$K_l^* = -(S_l + B_{l,k}^T P_l B_{l,k})^{-1} B_{l,k}^T P_l A_{l,k} = -(\Theta_{l,k}^{uu})^{-1} \Theta_{l,k}^{ux} \tag{25}$$

It is important to note that if the parameter vector Θ_l can be estimated online, then system dynamics are not needed to calculate the optimal control gain for lth adaptive model-based event-triggered control system pair.

b) Model-free online tuning of the value function

In this subsection, the adaptive model-based optimal control is developed by using estimated states from adaptive model without using value and policy iterations.

After incorporating the estimated system states from adaptive model, the value function can be represented in vector form as

$$V(\hat{x}_{l,k}, u_{l,k}) = \hat{z}_{l,k}^T \Theta_l \hat{z}_{l,k} = \rho_l^T w_{l,k} \tag{26}$$

where the regression functions $\hat{z}_{l,k} = [\hat{x}_{l,k}^T \;\; u_{l,k}^T]^T \in \Re^{n+m=j}$, $w_{l,k} = (\hat{z}_{l,k1}^2, \ldots, \hat{z}_{l,k1}\hat{z}_{l,kj}, \hat{z}_{l,k2}^2, \ldots, \hat{z}_{l,kj-1}\hat{z}_{l,kj}, \hat{z}_{l,kj}^2)$ is the Kronecker product quadratic polynomial independent basis vector, and $\rho_l = vec(\Theta_l)$

with $vec(\bullet)$ is a vectorized function constructed by stacking the columns of the square matrix into a one column vector with off-diagonal elements summed as $\Theta_{nm}+\Theta_{mn}$.

According to standard optimal control theory [19], the Bellman equation is given in terms of the value function by $V(x_{l,k+1},u_{l,k+1})-V(x_{l,k},u_{l,k})+r(x_{l,k},u_{l,k})=0$. However, this equality cannot be satisfied while the estimated system stats $\hat{x}_{l,k}$ and the estimated term $\hat{\rho}_{l,k}$ are used.

After incorporating the estimated system states, we have the estimated Bellman equation as

$$\hat{V}(\hat{x}_{l,k+1},u_{l,k+1})-\hat{V}(\hat{x}_{l,k},u_{l,k})+r(\hat{x}_{l,k},u_{l,k})=e_{l,k+1} \tag{27}$$

with $e_{l,k+1}$ is the temporary difference (TD) error [13] in the Bellman equation. The estimated value function for lth CPS adaptive model-based event-triggered control system pair can be expressed similar to (26) as

$$\hat{V}(\hat{x}_{l,k},u_{l,k})=\hat{z}_{l,k}^T\hat{\Theta}_{l,k}\hat{z}_{l,k}=\hat{\rho}_{l,k}^T w_{l,k}^e \tag{28}$$

Using the delayed values for convenience, the TD error in (27) can be represented as

$$\begin{aligned} e_{l,k} &= r(\hat{x}_{l,k-1},u_{l,k-1})+\hat{\rho}_{l,k}^T w_{l,k}^e-\hat{\rho}_{l,k}^T w_{l,k-1}^e \\ &= r(\hat{x}_{l,k-1},u_{l,k-1})+\hat{\rho}_{l,k}^T (w_{l,k}^e-w_{l,k-1}^e) \\ &= r(\hat{x}_{l,k-1},u_{l,k-1})+\hat{\rho}_{l,k}^T \Delta W_{l,k-1}^e \end{aligned} \tag{29}$$

with $\Delta W_{l,k-1}^e = w_{l,k}^e - w_{l,k-1}^e$.

Then, the dynamics of the TD error can be rewritten as

$$e_{l,k+1}=r(\hat{x}_{l,k},u_{l,k})+\hat{\rho}_{l,k+1}^T \Delta W_{l,k}^e \tag{30}$$

Next, an auxiliary error vector incorporating the history of pervious cost-to-go is defined as

$$\mathbf{\Pi}_{l,k} = \mathbf{\Gamma}_{l,k-1} + \hat{\rho}_{l,k}^{T} \mathbf{\Delta W}_{l,k-1}^{e} \tag{31}$$

where $\Gamma_{l,k-1} = [r(\hat{x}_{l,k-1}, u_{l,k-1}), r(\hat{x}_{l,k-2}, u_{l,k-2}), \cdots, r(\hat{x}_{l,k-1-j}, u_{l,k-1-j})]$ and $\mathbf{\Delta W}_{l,k-1}^{e} = [\Delta W_{l,k-1}^{e} \ \Delta W_{l,k-2}^{e} \ \cdots \ \Delta W_{l,k-1-j}^{e}]$. Note that (31) indicates a time history of pervious $j+1$ TD error (29) recalculated by using the most recent $\hat{\rho}_{l,k}$. As a consequence, the value and policy iterations are not needed.

The dynamics of the auxiliary vector are generated similar to (30) as

$$\mathbf{\Pi}_{l,k+1} = \mathbf{\Gamma}_{l,k} + \hat{\rho}_{l,k+1}^{T} \mathbf{\Delta W}_{l,k}^{e} \tag{32}$$

Next, the update law of matrix $\hat{\Theta}_l$ for lth pair CPS can be defined as

$$\hat{\rho}_{l,k+1} = \Delta \mathbf{W}_{l,k}^{e} (\Delta \mathbf{W}_{l,k}^{eT} \Delta \mathbf{W}_{l,k}^{e})^{-1} (\alpha_{l,o} \mathbf{\Pi}_{l,k}^{T} - \mathbf{\Gamma}_{l,k}^{T}) \tag{33}$$

with $0 < \alpha_{l,o} < 1$ is a tuning parameter for lth pair CPS. Substituting (33) into (32) yields

$$\mathbf{\Pi}_{l,k+1} = \alpha_{l,o} \mathbf{\Pi}_{l,k} \tag{34}$$

or

$$e_{l,k+1} = \alpha_{l,o} e_{l,k} \tag{35}$$

Remark 3: It is observed while the system states have converged to zeros, the estimated value function is no longer updated. It can be seen as a persistency of excitation (PE) requirement for the inputs to the value function estimator wherein the adaptive model-based event-triggered control system states must be persistently existing long enough for the estimator to learn the value function. The PE condition is well known in adaptive control theory and can be satisfied by adding exploration noise [25].

c) *Estimation of the Optimal Event-triggered Control*

According to [19], the optimal control can be obtained by minimizing the value function. Recalling (25), the optimal control gain can be developed by using the estimated states as

$$\hat{u}_{l,k} = \hat{K}_{l,k}\hat{x}_{l,k} = -(\hat{\Theta}^{uu}_{l,k})^{-1}\hat{\Theta}^{ux}_{l,k}\hat{x}_{l,k} \tag{36}$$

Based on (36), optimal control gain can be obtained in terms of $\hat{\Theta}_{l,k}$ matrix, which is solved by approximating value function with adaptive model-based event-triggered control system states. It relaxes the requirement of the system dynamics when (33) eliminates the value and policy iterations.

Next, the convergence proof of the optimal control inputs will be given in the following theorem.

Theorem 3 *(Convergence of the Optimal Control Inputs)*: Let $u_{l,0}(k)$ be an initial admissible control policy for lth pair CPS adaptive model-based event-triggered control system. Let the adaptive update law be given by (33). Then there exists a positive constant $0 < \alpha_{l,o} < 1$ such that the estimated parameter and optimal signal converges to the actual parameter and optimal signal respectively, i.e. $\hat{\rho}_{l,k} \to \rho_l$, $\hat{u}_{l,k} \to u^*_{l,k}$.

Proof: according to the Bellman equation, we have

$$\hat{\rho}^T_{l,k}\Delta W^e_{l,k-1} + r(\hat{x}_{l,k-1}, u_{l,k-1}) - \rho^T_l \Delta W_{l,k-1} - r(x_{l,k-1}, u_{l,k-1}) = e_{l,k} \tag{37}$$

In the other words,

$$\tilde{\rho}_{l,k}^{T}\Delta W_{l,k-1} + \rho_l^T \Delta\tilde{W}_{l,k-1} + \tilde{r}_{l,k-1} = e_{l,k} \tag{38}$$

with $\tilde{\rho}_{l,k} = \rho_l - \hat{\rho}_{l,k}$, $\tilde{r}_{l,k-1} = r(x_{l,k-1}, u_{l,k-1}) - r(\hat{x}_{l,k-1}, u_{l,k-1})$ and $\Delta\tilde{W}_{l,k} = \Delta W_{l,k}^{e} - \Delta W_{l,k}$. Then, selecting the Lyapunov function candidate as

$$V_{ae}(\tilde{\rho}_{l,k}, \tilde{r}_{l,k}, \Delta\tilde{W}_{l,k}) = (\tilde{\rho}_{l,k}^{T}\Delta W_{k-1}^{e} + \rho_l^T \Delta\tilde{W}_{k-1} + \tilde{r}_{l,k-1})^2 \tag{39}$$

Next, the first difference of Lyapunov function candidate as

$$\begin{aligned}\Delta V_{ae} &= (\tilde{\rho}_{l,k+1}^{T}\Delta W_{l,k}^{e} + \rho_l^T \Delta\tilde{W}_{l,k} + \tilde{r}_{l,k})^2 \\ &\quad - (\tilde{\rho}_{l,k}^{T}\Delta W_{l,k-1}^{e} + \rho_l^T \Delta\tilde{W}_{l,k-1} + \tilde{r}_{l,k-1})^2 \\ &= -(1-\alpha_{l,o}^2)(\tilde{\rho}_{l,k}^{T}\Delta W_{l,k-1}^{e} + \rho_l^T \Delta\tilde{W}_{l,k-1} + \tilde{r}_{l,k-1})^2\end{aligned} \tag{40}$$

When $0 < \alpha_{l,o} < 1$, the first difference ΔV_{ae} is negative definite. Meanwhile, the term $(\tilde{\rho}_{l,k}^{T}\Delta W_{k-1}^{e} + \rho_l^T \Delta\tilde{W}_{k-1} + \tilde{r}_{l,k-1})$ can be represented as

$$\tilde{\rho}_{l,k}^{T}\Delta\psi_{l,k-1}^{e} + \rho_l^T \Delta\tilde{\psi}_{l,k-1} + \tilde{r}_{l,k-1} = [\Delta W_{l,k-1}^{e} \quad \rho_l^T \quad 1]\begin{bmatrix}\tilde{\rho}_{l,k}^{T} \\ \Delta\tilde{W}_{l,k-1} \\ \tilde{r}_{l,k-1}\end{bmatrix} \tag{41}$$

Define $\psi_{l,k} = [\Delta W_{l,k-1}^{e} \quad \rho_l^T \quad 1]$ and $Z_{l,k} = [\tilde{\rho}_{l,k} \quad \Delta\tilde{W}_{l,k-1}^{T} \quad \tilde{r}_{l,k-1}^{T}]^T$. Using the adaptive control theory, we know $\psi_{l,k}\psi_{l,k}^{T} \neq 0$ provided $\psi_{l,k}$ satisfies the PE condition. Therefore, we have

$$Z_{l,k} = \begin{bmatrix}\rho_{l,k}^{T} \\ \Delta\tilde{W}_{l,k-1} \\ \tilde{r}_{l,k-1}\end{bmatrix} \to \mathbf{0} \tag{42}$$

Therefore, all the signal errors will converge to zero asymptotically. Namely, for lth pair CPS adaptive model-based event-triggered control, $\hat{\rho}_{l,k} \to \rho_l$, $\hat{u}_{l,k} \to u_{l,k}^{*}$ when $k \to \infty$

.

Until now, the adaptive model-based optimal event-triggered control is derived for multiple pairs CPS. Next, the novel distributed scheduling algorithm will be proposed.

D. Novel Distributed Scheduling Algorithm for Multiple Pair CPS

In this section, optimal distributed scheduling design is derived at data link layer mainly. Without loss of generality, traditional wireless ad-hoc network protocol [27] is implemented into the other layers. For optimizing the performance of multiple CPS pairs which included performance from both application layer and data link layer, a novel optimal cross-layer distributed scheduling algorithm is proposed by incorporating control system information from application layer. Similar to above sections, without loss of generality, lth CPS pair is considered here.

Firstly, the cost function for lth CPS pair is represented as

$$J_{l,k} = x_{l,k}^T Q_l x_{l,k} + u_{l,k}^T S_l u_{l,k} + \beta_l R_{l,k} \tag{43}$$

where $R_{l,k}$ is lth CPS pair average traffic payload during $[0, kT_s]$ and β_l is the weight of average traffic payload for lth adaptive model-based event-triggered control system pair. While β_l is large, it indicates the average traffic payload will affect the total cost more.

Subsequently, the entire cost function for multiple CPS adaptive model-based event-triggered control system pairs can be represented as

$$J_k = \sum_{l=1}^{M} J_{l,k} = \sum_{l=1}^{M} (x_{l,k}^T Q_l x_{l,k} + u_{l,k}^T S_l u_{l,k} + \beta_l R_{l,k}) \tag{44}$$

with M denotes the number of CPS pairs.

Next, the optimal design of multiple CPS pairs should minimize the cost function (44), i.e.

$$J_k^* = \min_{u,\pi} \sum_{l=1}^{M} (x_{l,k}^T Q_l x_{l,k} + u_{l,k}^T S_l u_{l,k} + \beta_l R_{l,k}) \tag{45}$$

where u is the control design and π is the scheduling policy.

According to (45), minimizing the above cost function (45) requires two parts: 1) the optimal control input design and 2) novel distributed scheduling design. For the control input designing part l, the above adaptive model-based optimal event-triggered control scheme have already shown to provide a best performance. For the distributed scheduling part, novel scheme will be derived in this section.

Obviously, each pair CPS has two options for scheduling: 1) CPS pair is scheduled; and 2) CPS pair is not scheduled. It is important to note that whether or not each CPS pair is scheduled depends upon which option can bring more benefits (i.e. large cost value). For instance,

Case 1: lth CPS pair has been scheduled

$$\begin{aligned} J_{l,k}^{S,1} &= x_{l,k}^T Q_l x_{l,k} + x_{l,k}^T \hat{K}_{l,k}^T S_l \hat{K}_{l,k} x_{l,k} + \beta_l R_{l,k}^Y \\ &= x_{l,k}^T Q_l x_{l,k} + x_{l,k}^T \Lambda_l x_{l,k} + \beta_l R_{l,k}^Y \end{aligned} \tag{46}$$

where $\Lambda_l = \hat{K}_{l,k}^T S_l \hat{K}_{l,k}$ and $R_{l,k}^Y$ is the average traffic payload when lth CPS pair has been scheduled.

Case 2: lth CPS pair has not been scheduled

$$\begin{aligned} J_{l,k}^{S,2} &= x_{l,k}^T Q_l x_{l,k} + \hat{x}_{l,k}^T \hat{K}_{l,k}^T S_l \hat{K}_{l,k} \hat{x}_{l,k} + \beta_l R_{l,k}^N \\ &= x_{l,k}^T Q_l x_{l,k} + \hat{x}_{l,k}^T \Lambda_l \hat{x}_{l,k} + \beta_l R_{l,k}^N \end{aligned} \tag{47}$$

with $R_{l,k}^{N}$ is the average traffic payload when lth CPS pair has not been scheduled.

Then, the difference between these two cases can be considered as utility function and expressed as

$$\begin{aligned}\Delta J_{l,k}^{S} &= J_{l,k}^{S,1} - J_{l,k}^{S,2} \\ &= (x_{l,k}^{T} Q_l x_{l,k} + x_k^{T} \Lambda_l x_{l,k} + \beta_l R_{l,k}^{Y}) \\ &\quad - (x_{l,k}^{T} Q_l x_{l,k} + \hat{x}_{l,k}^{T} \Lambda_l \hat{x}_{l,k} + \beta_l R_{l,k}^{N}) \\ &= \varphi(e_{l,k}) + \beta_l D_{l,k}\end{aligned} \tag{48}$$

where $D_{l,k} = R_{l,k}^{Y} - R_{l,k}^{N}$ is the difference of average traffic payload for lth CPS pair between two cases which can be represented as

$$D_{l,k} = R_{l,k}^{Y} - R_{l,k}^{N} = (\frac{N_k + 1}{kT_s} - \frac{N_k}{kT_s}) N_{l,bit} = \frac{1}{kT_s} N_{l,bit} \tag{49}$$

with $N_{l,bit}$ is the number of bits for packetizing the sensed event of lth CPS pair, $e_{l,k} = x_{l,k} - \hat{x}_{l,k}$ and $\varphi(e_{l,k}) = (x_{l,k}^{T} \Lambda_l x_{l,k} - \hat{x}_{l,k}^{T} \Lambda_l \hat{x}_{l,k})$. Obviously, when $\Delta J_{l,k}^{S} > 0$, it indicates that scheduling lth CPS pair can obtain more benefits. Otherwise, scheduling lth CPS pair will degrade the performance. Therefore, when $\Delta J_{l,k}^{S} > 0$, this CPS pair can be considered as scheduled. It is important to note that there are multiple CPS pairs (i.e. M CPS pairs), and probably several CPS pairs' utility function are higher than zero which indicates that all of these CPS pairs have to be scheduled. However, according to communication network literature [13], only one CPS pair can access the communication network. For optimizing the performance of network, the optimal scheduling policy should maximize the total utility function, that is

$$(\Delta J_k^{S})^{*} = \max_{\pi} \sum_{l \in \mathbf{G}_k} \Delta J_{l,k}^{S} \tag{50}$$

where $\mathbf{G}_k$ is the CPS pair set with positive value of utility function at time kT_s (i.e. $\Delta J_{l,k}^S > 0$ for $l \in \mathbf{G}_k$). Obviously, for centralized scheduling design, the optimal scheduler has to select the CPS pair that has the maximum value of $\Delta J_{l,k}^S$. However, in the centralized scheduling scheme, finding the maximum value $\Delta J_{l,k}^S$ requires significant information from every CPS pair which might be too complex to be implemented into the practical system. Therefore, novel distributed scheduling scheme is needed to solve this drawback.

In this paper, the main idea of proposed novel distributed scheduling algorithm is to separate the transmission time of different CPS pairs by using backoff interval (BI) [27] based on related utility function in a distributed manner. In Figure 5, the framework is shown.

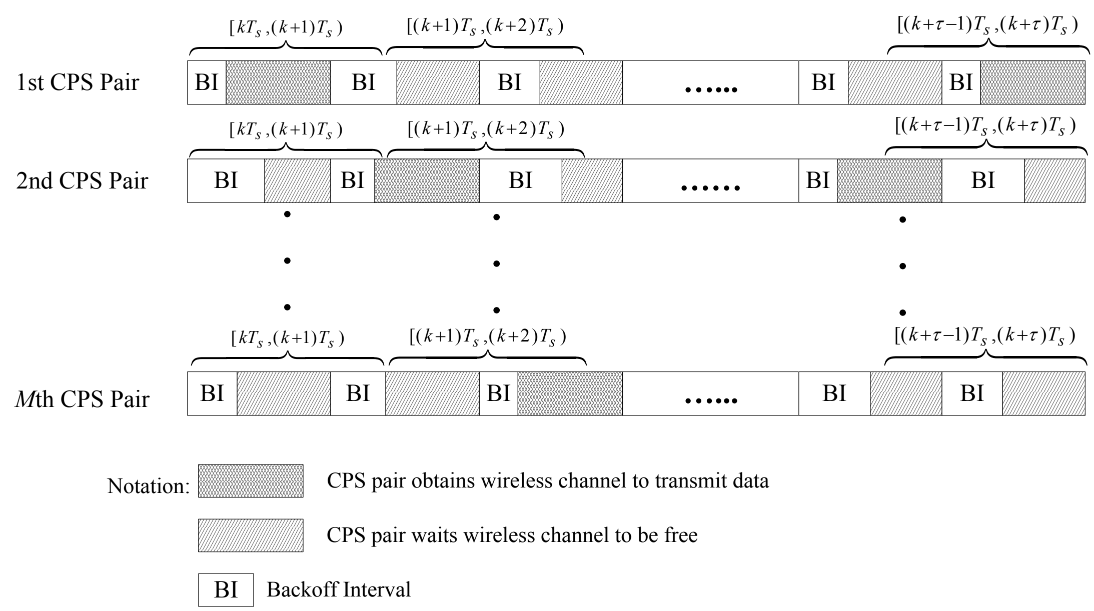


Figure 5. Framework of proposed cross-layer distributed scheduling scheme

To solving optimal scheduling problem (50) for multiple pairs of CPS, the BI can be designed as

$$BI_{l,k} = \varsigma \times (e^{-\Delta J_{l,k}} + n_{l,k}) \quad \text{for } l \in \mathbf{G}_k \tag{51}$$

with ς is the scaling factor and $n_{l,k}$ is a random variable which satisfies Gaussian distribution (i.e. $n_{l,k} \sim L * N(0,\sigma^2)$), and $L = \min\limits_{l,j \in G_k}(e^{-\Delta J_{l,k}} - e^{-\Delta J_{j,k}})$ is the range of the random value $n_{l,k}$. Next, the proposed novel distributed scheduling algorithm steps are shown as following:

Algorithm 1 Novel optimal distributed scheduling scheme

1: **Initialize:** The utility function are initialized as $\Delta J_{l,0} = 0, \forall l = 1,2,\ldots,M$

2: **While** $\{ kT_s \le t < (k+1)T_s \}$ **do**

3: **Calculate** backoff interval (BI) by different pair of CPS adaptive model based event triggered control system

$$BI_{l,k} = \varsigma \times (e^{-\Delta J_{l,k}} + n_{l,k}) \text{ for } l \in \mathbf{G}_k .$$

4: **Contend** shared communication network resource.

5: **If** lth pair of CPS has the smallest BI **then**

6: **Schedule** lth event triggered pair and transmit lth CPS pair's data through shared communication network.

7: **If** transmission is over, **then**

8: **Update** the scheduled CPS pair's utility function $\Delta J_{l,k}$.

9: **end if**

10: **else**

11: **Wait** for shared communication network channel to be free.

12: **end if**

13: **Update** time stamp: $t = t + BI_{l,k} + T_{l,k}$ ($BI_{l,k}$ is the backoff interval of scheduled CPS pair. $T_{l,k}$ is the transmission time of scheduled CPS pair.)

14: **end while**

15: **Update and broadcast** utility function $\Delta J_{l,k}$ from each pair of CPS.

16: **Go to** next time period $[(k+1)T_s,(k+2)T_s)$ (i.e. $k=k+1$), and go back to line 2.

Remark 4: Since each CPS pair decides its scheduling by only using local information from application and data link layers, proposed novel cross-layer scheduling scheme is distributed.

Remark 5: Compared with other distributed scheduling schemes [14-18], the proposed algorithm generates the backoff interval intelligently by optimizing utility function instead of selecting it randomly as in [14-18], which can be considered as main contribution of developed novel distributed scheduling algorithm in this paper.

Next, the optimality of proposed novel distributed scheduling is shown in Theorem 4.

Theorem 4: *(Optimal distributed scheduler performance)* Given the multiple CPS pairs and event triggered control scheme, the proposed distributed scheduling scheme selects the adaptive model based event-triggered CPS pair with highest utility function value since it has the shortest backoff interval (i.e. BI) and highest priority to access the shared communication network. In addition, the proposed algorithm can render best performance schedules for every CPS pair.

Proof: Assume lth CPS pair has the highest utility function value (i.e. $\Delta J_{l,k}=\max_{i\in\mathbf{G}_k}\Delta J_{i,k}$), then we have $\Delta J_{l,k}>\Delta J_{i,k}$ for any $i=\mathbf{G}_k, i\neq l$. Therefore,

$$e^{-\Delta J_{l,k}} < e^{-\Delta J_{i,k}} \qquad i = \mathbf{G}_k, i \neq l \tag{52}$$

Next, for any $i = \mathbf{G}_k, i \neq l$, the backoff interval (BI) can be expressed as

$$\begin{aligned} BI_{l,k} &= \varsigma \times (e^{-\Delta J_{l,k}} + n_{l,k}) \\ &< \varsigma \times [e^{-\Delta J_{l,k}} + \min_{l,j \in [1,M]} (e^{-\Delta J_{l,k}} - e^{-\Delta J_{j,k}})] \\ &< \varsigma \times (e^{-\Delta J_{l,k}} + e^{-\Delta J_{i,k}} - e^{-\Delta J_{l,k}}) \\ &< \varsigma \times e^{-\Delta J_{i,k}} < \varsigma \times (e^{-\Delta J_{i,k}} + n_{i,k}) \\ &< BI_{i,k} \end{aligned} \tag{53}$$

Hence, $BI_{l,k} < BI_{i,k}$ for any $i = \mathbf{G}_k, i \neq l$. Based on proposed distributed scheduling algorithm, lth CPS pair can be scheduled to use shared communication network due to its shortest backoff interval (BI).

Next, the cost-optimality of proposed scheme will be proven by using contradiction method.

Assume that there exists another jth CPS pair, and scheduling jth CPS pair can render better performance than scheduling lth CPS pair even lth CPS pair has shortest BI (i.e. cost function value $J_k^{sj} < J_k^{sl}$, but $BI_{l,k} < BI_{j,k}$). According to the definition of cost function (45), J_k^{sj} can be defined as

$$J_k^{sj} = (\sum_{i=1}^{M} J_{i,k}) - \Delta J_{j,k} \text{ for } j = 1,2,\ldots,M \tag{54}$$

Since $BI_{l,k} < BI_{j,k}$ is given in assumption above, we have $\Delta J_{j,k} < \Delta J_{l,k}$ by using (51) and (53). Meanwhile, the cost function of jth CPS pair can be derived as

$$J_k^{sj} = (\sum_{i=1}^{M} J_{i,k}) - \Delta J_{j,k} > (\sum_{i=1}^{M} J_{i,k}) - \Delta J_{l,k} = J_k^{sl} \tag{55}$$

It is important to note that $J_k^{sj} > J_k^{sl}$ in (55) is contradicted with assumption that $J_k^{sj} < J_k^{sl}$. According to contradiction method, there does not exist any other scheduled CPS pair can obtain the better performance than scheduling lth CPS pair which has the shortest backoff interval. In the other words, proposed distributed scheduling algorithm can render the best performance by scheduling the CPS pair with shortest backoff interval.

On the other hand, any CPS pair with negative utility function value (i.e. $\Delta J_{i,k} < 0$) should not contend the shared communication network resource since it will degrade the performance. Next, the proof about this is given in details.

Assume pth CPS pair with a negative utility function (i.e. $\Delta J_{p,k} < 0$) is scheduled, then the cost function with this scheduling decision can be expressed as

$$J_k^{sp} = \sum_{i=1}^{M} J_{i,k} - \Delta J_{p,k} > \sum_{i=1}^{M} J_{i,k} = J_k (\text{no pair is scheduled}) \tag{56}$$

Therefore, scheduling a CPS pair with negative utility function will degrade the performance.

Remark 5: Fairness is an important factor to evaluate the performance of scheduling schemes. For proposed distributed scheduling algorithm, a fairness index is defined as $FI = \frac{\left(\sum_{i=1}^{M} \frac{R_i}{\sum_{j=0}^{\infty}(x_{i,j}^T Q_i x_{i,j} + u_{i,j}^T R_i u_{i,j}) + \beta_i R_i}\right)^2}{M * \sum_{i=1}^{M}\left(\frac{R_i}{\sum_{j=0}^{\infty}(x_{i,j}^T Q_i x_{i,j} + u_{i,j}^T R_i u_{i,j}) + \beta_i R_i}\right)^2}$ to measure the fairness among different CPS pairs.

IV. SIMULATION RESULTS

In this section, the proposed cross-layer CPS co-design included novel optimal adaptive model-based event-triggered control and cross-layer distributed scheduling algorithm is evaluated in the following example. The CPS includes six pairs which are located within 300m*300m square area randomly. For maintain the homogeneous property, all six pairs are using the similar control system as [22]. The discrete-time model is given as

$$x_{k+1} = \begin{bmatrix} 1.1138 & -0.0790 \\ 0.0592 & 0.8671 \end{bmatrix} x(t) + \begin{bmatrix} 0.2033 \\ 0.1924 \end{bmatrix} u(t) \tag{57}$$

with sampling interval $T_s = 0.15$ seconds, the number of bits for the six quantized sensed data for the CPS pairs are defined as $N_{bit} = [10\ 8\ 6\ 7\ 8\ 4]$. The initial system states are given by $x_{1,0} = [20\ 7]^T$, $x_{2,0} = [12\ 5]^T$, $x_{3,0} = [10\ \ 3]^T$, $x_{4,0} = [8\ 4]^T$, $x_{5,0} = [10\ 5]^T$ and $x_{6,0} = [0.1\ 6]^T$.

First, the performance of proposed optimal adaptive model-based event-triggered control is shown. Due to the page limitation and without loss of generality, an average value of state regulation errors for the six CPS pairs is shown in the Figure 6. The results indicates that proposed optimal adaptive model-based event-triggered control design can not only force the regulation errors converge to zero asymptotically, but also make the state regulation errors converge to zero quickly while ensuring all CPS pairs are stable. It is important to note that overshoots observed at the beginning because the optimal control and adaptive model tuning needs a short time.

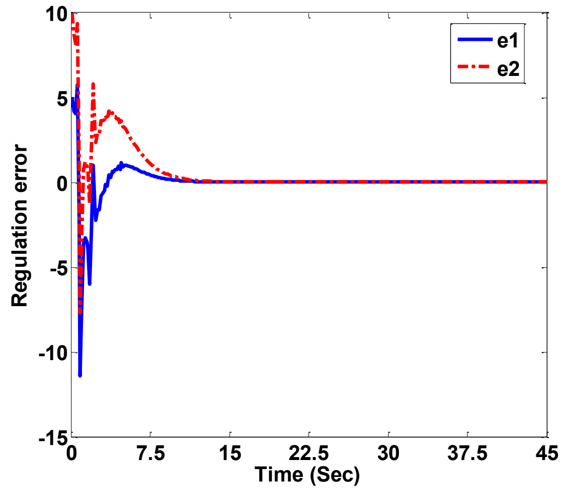


Figure 6. State regulation errors with optimal adaptive model-based event-triggered control system.

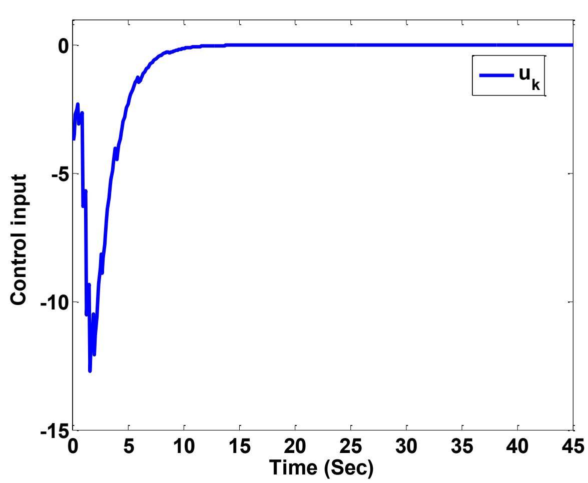


Figure 7. Performance of optimal control for CPS with unknown system dynamics.

Next, the performance of proposed CPS optimal adaptive model-based event-triggered control is evaluated. Without loss of generality, the design of 5^{th} CPS pair's adaptive model-based event-triggered control input is shown in Figure 7. It is important to note that the proposed optimal control inputs can make the CPS state regulation error converge to zero when the CPS dynamics are unknown which implies that the proposed controller can make the CPS closed-loop system stable. Meanwhile, proposed optimal control design has a small overshoot initially since the optimal controller needs a short tuning phase.

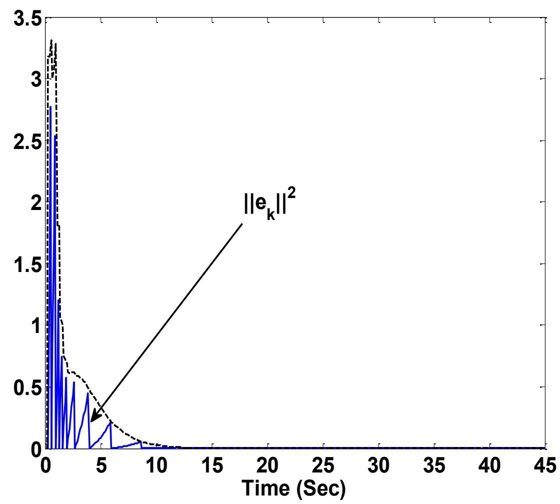


Figure 8. Performance of the adaptive model-based event-triggered CPS estimation error norm $\|e_k\|^2$

Then, the performance of adaptive model-based event-triggered CPS's error event is shown. Without loss of generality, we pick 5^{th} adaptive model-based event-triggered CPS to evaluate the performance. As shown in Figure 8, when error event is triggered and scheduled, the estimation error will be reset to zero since actual CPS state will be received at the controller. On the other hand, if the error event is not triggered and scheduled, the estimation error increases due to inaccurate adaptive model. Once the adaptive parameters are estimated accurately, the estimation error converges to zero. It is important to note that the state estimation error is large and error event are triggered and scheduled more frequently at the beginning since adaptive model needs to be tuned. After a short period error events are obviously triggered and scheduled much less frequently which would reduce communication network traffic.

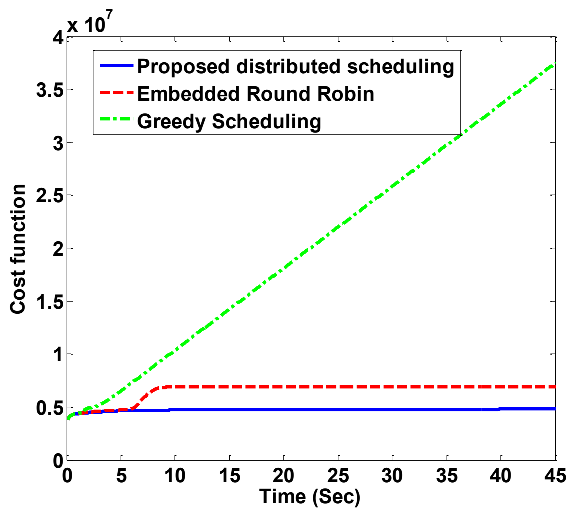


Figure 9. The cost function comparison for different scheduling schemes

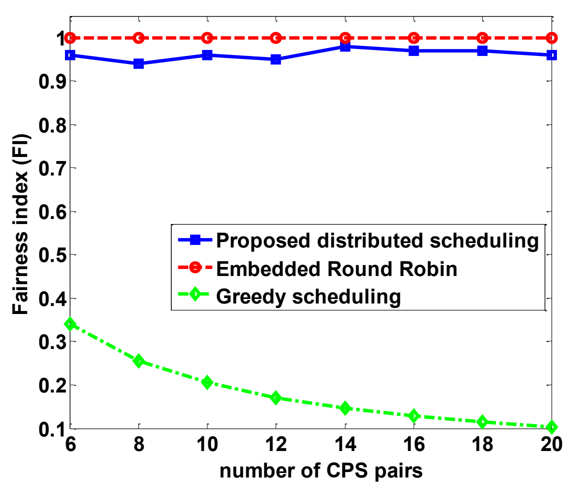


Figure 10. The fairness comparison for different scheduling schemes

Next, the performance of proposed cross-layer distributed scheduling has been evaluated. For comparison, classical widely used embedded round robin (ERR) [27] and Greedy scheduling [27] are added. In Figure 9, the cost function of multiple pairs of CPS with three different scheduling schemes is compared. Proposed novel cross-layer distributed scheduling maintain a lowest value while costs of multiple pairs of CPS with ERR and Greedy scheduling are much more than proposed scheduling scheme. It indicates that the proposed distributed scheduling scheme can improve the performance of multiple pairs CPS much better than widely used ERR and Greedy scheduling. It is important to note that: 1) Since ERR only guarantees that each CPS pair can have the same probability to access the shared communication network resource and does not consider to efficient the usage of network resource for the multiple pairs CPS, it cannot optimized the multiple pairs CPS performance, and 2) Since Greedy scheduling only focuses on data link layer performance optimization and cost function of multiple pairs CPS is defined from both data link layer and application layer, it also cannot optimize the performance of the multiple CPS pairs.

Eventually, the fairness of different scheduling schemes has been evaluated. As shown in Figure 10, fairness indices of proposed cross-layer distributed scheduling and widely used ERR schemes are close and equal to one, whereas that of Greedy scheduling is much less than one thus indicating fair allocation of shared communication network resource for the proposed one while meeting the overall performance. The ERR method though is fair has higher cost than the proposed one.

According to above results (Figure 6 through 10), the proposed cross-layer co-design included optimal adaptive model-based event-triggered control scheme and a cross-layer distributed scheduling scheme to optimize the performance of both communication network and the CPS subsystems.

V. CONCLUSION

In this work through a novel cross-layer co-design for multiple pairs Cyber-Physical System, it is demonstrated that the proposed scheme can optimize not only the performance of control system, but also the shared communication network. The novel optimal adaptive model-based event-triggered control does not require system dynamics and use event-triggered instead of an inefficient time-driven sampling which is quite useful for hardware implementation. The novel scheduling algorithm is distributed, simple and requires less computation than centralized scheduling algorithms.

REFERENCES

[1] A.M. Meystel and J.S. Albus, *Intelligent systems: architecture, design, and control*, Wiley, John & Sons, New York, 2002.

[2] I.F. Akyildiz, W.Y Lee, M.C. Vuran and S. Mohanty, “NeXt generation/ dynamic spectrum access/ cognitive radio wireless networks: A survey,” *Computer Networks*, vol. 50, pp. 2127-2159, 2006.

[3] J. Baillieul, P.J. Ansaklis, “Control and communication challenges in networked real-time systems,” in *Proceeding of IEEE*, vol. 95, pp. 9-28, 2007.

[4] G. Tian, Y.C. Tian and C. Fidge, “Performance analysis of IEEE 802.11 DCF based WNCS networks,” in *Procceding of Local Computer Networks*, pp. 496-503, 2010.

[5] R. Poovendran, K. Sampigethaya, K.S. Gupta, I. Lee, K.V. Prasad, D. Corman and J.L. Paunicka, “Special issue on Cyber-Physical Systems,” in *Proceeding of IEEE*, in press, 2012.

[6] F. Xia, X, Kong and Z. Xu, “Cyber-Physical control over wireless sensor and actuator networks with packet loss,” in Book: *Wireless networking based control*, Springer, 2012.

[7] F. Xia, A. Vinel, R. Gao, L. Wang and T. Qiu, “Evaluating IEEE 802.15.4 for Cyber-Physical Systems,” *EURASIP Journal on wireless communication and networking*, vol. 2011, pp. 1-14, 2011.

[8] J.D. Day and H. Zimmermann, “The OSI reference model,” in *Proceeding of IEEE*, vol. 71, pp. 1334-1340, 1983.

[9] V. Srivastava and M. Motani, “Cross-layer design: a survey and the road ahead,” *IEEE Communication Magazine*, vol. 43, pp. 112-119, 2005.

[10] S. Shakkottai, T.S. Rappaport and P.C. Karlsson, “Cross-layer design for wireless networks,” *IEEE Communication Magazine*, vol. 41, pp. 74-80, 2003.

[11] Q. Liu, S. Zhou and G.B. Giannakis, “Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design,” *IEEE Transactions on wireless communication,* vol. 4, pp. 1142-1153, 2005.

[12] P. Reena and L. Jacob, “Joint congestion and power control in UWB based wireless sensor network,” in *Proceeding of 32nd IEEE Conference on Local Computer Network (LCN)*, pp. 911-918, 2007 .

[13] IEEE Standard 802.11a 1999 (R2003), “Wireless LAN medium access control (MAC) and Physical layer (PHY) specification, high-speed physical layer in 5GHz band,” 2003.

[14] J.G. Dai and B. Prabhakar, “The throughput of data switchs with and without speedup,” in *Proceeding of 31st IEEE International Conference on Computer Communications*, pp. 556-564, 2000.

[15] L. Jiang and J. Walrand, “A distributed CSMA algorithm for throughput and utility maximization in wireless networks,” *IEEE/ACM Transactions on Networking*, vol. 18, pp. 960-972, 2010.

[16] N. Vaidy, A. Dugar, S. Gupta and P. Bahl, “Distributed fair scheduling in wireless LAN,” *IEEE Transactions on Mobile Computing*, vol.4, pp. 616-629, 2005.

[17] D. Zheng, W. Ge and J. Zhang, “Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach,” *IEEE Transactions on Information Theory*, vol. 55, pp. 205-222, 2009.

[18] L. Qian and R. Negi, “Distributed throughput-optimal scheduling in ad hoc wireless networks,” in *Proceeding of 2011 IEEE International Conference on Communicaitons (ICC)*, pp. 1-5, 2011.

[19] F.L. Lewis and V.L. Syrmos, Optimal control, 2nd edition, Wiley, New York, 1995.

[20] M. Mazo Jr. and P. Tabuada, “Decentralized event-triggered control over wireless sensor/actuator netowrks,” *IEEE Transactions on Automatic Control*, vol. 56, pp. 2456-2461, 2011.

[21] D.V. Dimarogonas and K.H. Johanson, “Event-triggered control for multi-agent systems,” in *Proceeding of 48th IEEE Conference on Decision and Control*, pp. 713-7136, 2009.

[22] E. Garcia and P.J. Antsaklis, “Model-based event-triggered control with time-varying network delays,” in Proceeding of 50th IEEE Conference on Decision and Control, pp. 1650-1655, 2011.

[23] G. Luders, and K. Narendra, “An adaptive observer and identifier for a linear system,” *IEEE Transaction. on Automatic Control*, vol. 18, pp. 496-499, 1973.

[24] P. J. Werbos, “A menu of designs for reinforcement learning over time,” *Journal of Neural Networks Control*, vol. 3, pp. 835-846, 1983.

[25] Green, M., and Moore, J.B. “Persistency of excitation in linear systems,” *System and Control Letter*, vol. 7, pp. 351–360, 1989.

[26] H. Xu, S. Jagannathan and F.L. Lewis, “Stochastic optimal control of unknown networked control systems in the presence of random delays and packet losses,” *Automatica,* in Press, 2012.

[27] S. Jagannathan, *Wireless ad-hoc and sensor network: protocols, performance, and control*, CRC Press, Florida, 2007.

[28] A. Eqtami, D. V. Dimarogonas, and K. J. Kyriakopoulos, “Event-triggered control for discrete-time systems,” in *Proc. Amer. Contr. Conf.*, pp. 4719-4724, 2010.

2. CONCLUSIONS AND FUTURE WORK

In this dissertation, adaptive dynamic programming (ADP) techniques are utilized to develop model-free stochastic optimal and suboptimal adaptive design for linear/nonlinear networked control system (NCS) in the presence of uncertain system dynamics and unknown network imperfections. First, a novel system representation has been derived for linear and nonlinear NCS respectively. Subsequently, using ADP techniques, a novel value function estimator design is included and a suite of stochastic optimal adaptive control schemes is derived based on estimated value function. Besides state-feedback, output-feedback based stochastic optimal adaptive control has been derived for both linear and nonlinear NCS. Value and policy iterations are not utilized; instead the value function and control policies are updated once a sampling interval thus making the proposed scheme suitable for hardware implementation.

The proposed optimal adaptive designs are not only able to maintain the stability of the NCS but also force the system performance to attain optimality even when system dynamics and network imperfections are unknown. Finally, the behavior of network protocols is investigated and a cross-layer framework to design both the controller and the protocol is introduced for next generation Cyber Physical Systems (CPS).

2.1. CONCLUSIONS

In the first paper, an online ADP technique based on adaptive value function estimator is introduced to solve the stochastic optimal adaptive and suboptimal regulation of linear NCS with uncertain dynamics in presence of unknown network-induced delays and packet losses. Considering effects of network imperfections, the linear NCS is represented in terms of augmented state vector consisting of past values of state and inputs. Due to unknown network imperfections, the linear NCS dynamics with augmented state vector becomes unknown and time-varying, and all traditional control schemes which neglect the network imperfections cannot maintain linear NCS stable in the mean. Linear NCS dynamics and network imperfections are not needed while estimating the value function and stochastic optimal adaptive control law is derived based on estimated value function by using past values of cost to go errors thus relaxing the value and or policy iterations. In addition, by using the proposed parameter update law, all value function estimator parameters are tuned forward-in-time, online and asymptotic stability of overall closed-loop system is demonstrated by using Lyapunov stability analysis. Exploration noise is shown to provide the needed persistence of excitation condition which is required for parameter convergence.

In the second paper, the optimal adaptive design was extended to solve linear zero-sum games in the presence of network imperfections. First, a novel linear NCS zero-sum games representation with augment states was derived. Subsequently, the optimal adaptive approach that combines value function estimator and ADP is utilized to solve Bellman equation of linear NCS zero-sum games in real-time without the knowledge of system dynamics and network imperfections. Since optimal adaptive scheme balances control and disturbance effects for linear NCS, Pareto optimality has been achieved. Finally, stability proofs guarantee the asymptotic convergence of value function estimator parameters in the mean and closed-loop system while simulation results reaffirm the satisfactory performance of the proposed stochastic optimal adaptive design.

On the other hand, in the third paper, a novel time-based stochastic optimal adaptive control scheme for nonlinear NCS referred to as neuro dynamic programming for NCS has been introduced in presence of unknown system dynamics and network imperfections. First this paper has developed the nonlinear NCS representation with augment states in the input-output form. A neural network (NN)-based identifier relaxes the requirement of input gain matrix for the nonlinear NCS while the action NN does not need information on network imperfections. Since proposed scheme is time-based and forward-in-time, value or policy iterations were not required because a history of cost-to-go errors was utilized in the value function estimator. Using Lyapunov theory, all NN weights and closed-loop signals had been proven uniformly ultimately bounded in the mean, and simulation results indicate the satisfied performance of proposed scheme even when the system is represented in the input-output form.

In the past three chapters, though the effect of delays and packet losses are included in the design, a specific network protocol behavior is not considered. In addition, the above stochastic adaptive optimal control designs for linear NCS are all based on state-feedback design while a network protocol behavior requires output feedback. Therefore, the behaviors of TCP and UDP by using output feedback design are considered in the fourth paper. First, a linear NCS representation under TCP and UDP was developed which is different from the past three chapters. Subsequently, an ADP-based scheme consisting of a novel observer, value function estimator is utilized to solve the Bellman equation in real-time for obtaining optimal control of linear NCS under TCP and UDP. Since past control inputs and estimated states have been used, the requirement of system dynamics and network imperfections was relaxed to obtain the stochastic optimal adaptive design.

In addition, stability region of designed observer is derived and compared with traditional observer. Although stability region of the proposed adaptive observer is smaller, it relaxed the requirements for unknown system dynamics. Eventually, Lyapunov theory is used to prove the observer and its parameters, adaptive value function estimator parameters and closed-loop system are asymptotically stable in the mean for NCS under TCP and uniformly ultimately bounded in the mean for NCS with UDP.

Furthermore, it is important to note that novel event-triggered control technique has attracted significant interest than traditional time-driven control due to its network benefits. The joint network protocol and controller design is necessary for CPS. Therefore, a novel cross-layer co-design is developed in the fifth paper. In this paper, first, an event-triggered control scheme is introduced instead of a traditional time-driven scheme. Subsequently, a novel optimal adaptive model-based event-triggered control scheme is derived. Compared with other event-triggered control schemes (e.g. ZOH, Fixed-model), proposed algorithm can not only save more network resources but also improves the performance of control system while relaxing the requirement of system dynamics. Then, a novel distributed scheduling is derived by maximizing the utility function which is based on information from both the application and network layers.

2.2. FUTURE WORK

As part of the future work, more network imperfections (e.g. quantization errors etc.) could be considered for NCS design. This would complicate the design and brings more challenges for proving stability. However, the major benefit is to make our design more suitable for real-time implementation.

On the other hand, although event-triggered control schemes are derived in the fifth paper, they only focus on linear discrete-time system. Since nonlinear system is more generic than a linear system, proposed event-triggered control schemes should be extended to the case of nonlinear systems. Therefore, the development of an adaptive model-based event-triggered control for nonlinear system can be considered as one part of future work. After that, optimal event-triggered controller scheme development can also be introduced as part of future work.

Another possibility would be efficient CPS co-design. Until now, most control and networking researchers separated the control and network protocol designs for the sake of convenience. However, such designs will degrade the performance of CPS since controller and network protocol designs can influence each other. Therefore, it is necessary to derive a novel co-design framework to overcome this drawback. Besides the fifth paper, more CPS co-designs such as joint intelligent control with novel routing, adaptive modulation/demodulation and encoding/decoding can be considered. Although developing an efficient co-design framework and proving its superiority mathematically are difficult, it is an interesting and promising area to explore.

REFERENCES

[1] P.R. Kumar, "Fundamental Issues in Networked Control System", in *Proceeding of 3rd WIDE Ph.D. School on Networked Control System*, Siena, Italy, 2009.

[2] Y. Halevi, and A. Ray, "Intergrated Communication and Control Systems: Part I---analysis", *Journal of Dynamic Systems and Measure Control,* vol. 110, pp. 367-373, 1988.

[3] P. Antsaklis, and J. Baillieul, "Special Issue on Networked Control Systms", *IEEE Transactions on Automatic Control*, vol. 49, pp. 1421-1423, 2004.

[4] M.S. Branicky, S.M. Phillips, and W. Zhang, "Stability of Networked Control Systems: explicit analysis of delay", in *Proceeding of American Control Conference*, pp. 2352-2357, 2000.

[5] J.P. Hesphanha, P. Naghshtabri, and Y. Xu, "A Survey of Recent Results in Networked Control System", in *Proceeding of IEEE*, vol. 95, pp. 138-162, 2007.

[6] J. Nilsson, B. Bernhardsson, and B. Wittenmark, "Stochastic Analysis and Control of Real-time Systems with Random Time Delays", *Automatica*, vol. 34, pp. 57-64, 1998.

[7] W.P.M.H. Heemels, A.R. van de Wouw, and D. Nesic, "Networked Control Systems with Communication Constraints: Tradeoff between transmission intervals, delays and Performance", *IEEE Transaction on Automatic Control*, vol. 55, pp, 1781-1796, 2010.

[8] W. Zhang, M.S. Branicky, and S. Phillips, "Stability of Networked Control Systems", *IEEE Control System Magazine*, vol. 21, pp. 84-99, 2001.

[9] G.C. Walsh, H. Ye, and L. Bushnell, "Stability Analysis of Networked Control Systems", in *Proceeding of American Control Conference*, pp. 2876-2880, 1999.

[10] F. Lian, J. Moyne, and D. Tilbury, "Modeling and Optimal Controller Design of Networked Control Systems with Multiple Delays", *International Journal of Control*, vol. 76, pp. 591-606, 2003.

[11] L.W. Liou, and A. Ray, "A Stochastic Regulator for Integrated Communication and Control Systems: Part I --- Formulation of Control Law", *ASME Journal of Dynamic System Measurement and Control*, vol. 4, pp. 688-694, 1991.

[12] X.H. Liu, and A. Goldsmith, "Wireless Medium Access Control in Networked Control System", in *Proceeding of American Control Conference*, pp. 688-694, 2004.

[13] S.S. Hu, and Q.X. Zhu, "Stochastic Optimal Control and Analysis of Stability of Networked Control Systems with Long delay", *Automatica*, vol. 39, pp. 1877-1884, 2003.

[14] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of Control and Estimation over Lossy Networks", in *Proceeding of IEEE*, vol. 95, pp. 163-187, 2007.

[15] J. Wu, and T. Chen, "Design of Networked Control Systems with Packet dropouts", *IEEE Transaction on Automatic Control*, vol. 52, pp. 1314-1319, 2007.

[16] M.B.G. Cloosterman, N. van de Wouw, W.P.M.H. Heemels, and H. Nijmeijer, "Stability of Networked Control Systems with Uncertain time-varying delays", *IEEE Transaction on Automatic Control*, vol. 54, pp. 1575-1580, 2009.

[17] F.L. Lewis, and V.L. Syrmos, *Optimal Control*, 2nd Edition, Wiley, New York.

[18] H. Zhang, Y. Luo, and D. Liu, "Neural Network Based Near Optimal Control for a Class of Discrete-time Affine Nonlinear System with Control Constraints", IEEE Transaction on Neural Network, vol. 20, pp. 1490-1530.

[19] J. Song, A.K. Mok, D. Chen, M. Nixon, "Challenges of Wireless Control in Process Control", *Real-time and embedded system workshop*, 2006.

[20] R. Poovendran, K. Sampigethaya, K.S. Gupta, I. Lee, K.V. Prasad, D. Corman and J.L. Paunicka, "Special issue on Cyber-Physical Systems," in *Proceeding of IEEE*, in press, 2012.

[21] F. Xia, X, Kong and Z. Xu, "Cyber-Physical control over wireless sensor and actuator networks with packet loss," in Book: *Wireless networking based control*, Springer, 2012.

[22] F. Xia, A. Vinel, R. Gao, L. Wang and T. Qiu, "Evaluating IEEE 802.15.4 for Cyber-Physical Systems," *EURASIP Journal on wireless communication and networking*, vol. 2011, pp. 1-14, 2011.

VITA

Hao Xu was born June 4, 1984 in Nanjing, China. He earned the Bachelor of Engineering degree in Electrical and Computer Engineering from Nanjing Forestry University and Master of Science degree in Electrical Engineering from Southeast University, China, in 2006 and 2009, respectively. He is a member of IEEE --- Institute of Electrical and Electronics Engineering. Hao Xu received the degree of Doctor of Philosophy at Missouri University of Science and Technology (formerly the University of Missouri-Rolla) in August 2012.