Abstract—Neuro dynamic programming (NDP) techniques for optimal control of nonlinear network control system (NNCS) are not addressed in the literature. Therefore, in this paper, a novel NNCS representation incorporating the unknown system uncertainties and network imperfections is introduced first by using input and output measurements. Then, an online neural network (NN) identifier is introduced to estimate the control coefficient matrix. Subsequently, the critic NN and action NN 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. Instead, value function and control inputs are updated at every sampling instant. Lyapunov theory is used to show that all the closed-loop signals and NN weights are uniformly ultimately bounded (UUB) while the approximated control input converges close to its target value with time.

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. The primary advantages of NCS are reducing system wiring, ease of system diagnosis and maintenance, and increasing system agility. Adding communication network in the feedback loop brings challenging issues such as random delays and packet losses which can cause the system unstable.


The work in [3] extended the controller design to a nonlinear NCS when the dynamics are considered known. However, optimal controller design is generally preferable for NCS and especially for NNCS, which is very difficult to attain. The uncertain dynamics and unknown network imperfections in the case of NNCS further complicates the optimal controller design. By using the stochastic optimal control theory, Nilsson [5] proposed the optimal and suboptimal controller design for linear NCS with random delays. Although these optimal [5] and suboptimal controller designs [5] have resulted in satisfactory performance, they are all based on known linear NCS dynamics and require information on delays which are not known beforehand.

Neuro dynamics programming (NDP) technique, on the other hand, proposed by Werbos [6] intends to solve optimal control problem forward-in-time. In NDP, one combines adaptive critics, a reinforcement learning technique, with dynamic programming. Although NDP is an effective technique to solve the optimal control of NNCS, traditional NDP techniques [6] requires partial knowledge of system dynamics which becomes a problem for NNCS due to network imperfections that are not known. In addition, NDP techniques based on value and/or policy iterations [6-9] are not useful for real-time control since a significant number of iterations may be needed within any sampling interval. Also, in some cases [8-9], a model may be needed to iterate the value and policies. Finally, existing state-feedback based NDP schemes [6-9] are not applicable for NNCS since the unknown network imperfections such as delays and packet losses can cause instability [2] if they are not considered carefully in the controller design.

Thus, in this paper, a novel time-based NDP algorithm is derived for NNCS with uncertain dynamics and in the presence of unknown network imperfections such as random delays and packet losses by using input-output measurements. To learn the partial knowledge of NNCS, an online 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. 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 [8-9] are not used and the value function and control input are updated at each sampling instant making the proposed NDP scheme a time-based model-free optimal controller for uncertain NNCS with unknown network imperfections.

II. NONLINEAR NETWORKED CONTROL SYSTEM (NCS) BACKGROUND

The feedback control loop in the NNCS considered in this paper is closed over a wireless network. Due to unreliability of wireless network, two types of networked-induced delay and one type of packet losses are included in NCS: (1) \( \tau_{c}(t) \): sensor-to-controller delay, (2) \( \tau_{a}(t) \): controller-to-actuator delay, and (3) \( \gamma(t) \): indicator of packet losses at actuator.

Research supported in part by NSF ECCS #1128281 and Intelligent Systems Center. Hao Xu and S. Jagannathan are with the Dept. of Electrical and Computer Engineering, Missouri University of Science and Technology, USA. ([hx@h7, sarangap]@mst.edu).
Next the following assumption is made similar as the literature in NCS [12]:

**Assumption 1:** a). Sensor is time-driven while the controller and actuator are event-driven [12]; b). Communication network is a wide area wireless network so that the two network-induced delays are considered independent and unknown whereas their probability distribution functions are considered known [12]; c). the sum of both the delay types is bounded [12] while the initial state of the nonlinear system is deterministic [12].

In this paper, a continuous-time nonlinear affine system of the form \( \dot{x} = f(x(t)) + g(x(t))u(t - \tau(t)), y(t) = Cx(t) \) is considered, where \( x \) and \( u \) denotes system states, output and input while \( f(\cdot) \) and \( g(\cdot) \) are smooth nonlinear functions of the state. When the network-induced random delays and packet losses of the wireless network are considered, the control input \( u(t) \) is delayed and can be lost at times due to packet losses. Therefore the original nonlinear affine system by incorporating the delay and packet loss effects can be expressed as

\[
\dot{x}(t) = f(x(t)) + \gamma(t)g(x(t))u(t - \tau(t)), \quad y(t) = Cx(t)
\]

where

\[
\gamma(t) = \begin{cases} \mathbf{I}^{ax} & \text{if control input is received by the actuator at time } t \ t - \tau(t) \text{is delayed control input} \end{cases}
\]

with \( \mathbf{I}^{ax} \) is identity matrix, \( u(t - \tau(t)) \) is delayed control input vector, \( x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, \gamma(t) \in \mathbb{R}^n, f(t) \in \mathbb{R}^{n \times n}, g(\cdot) \in \mathbb{R}^{n \times m} \) and \( C \in \mathbb{R}^{m \times n} \) which is invertible. From Assumption 1, sum of network-induced delays is bounded above, i.e.,

\[
\tau(t) = \tau_{eq}(t) + \tau_{eq}(t) < \tilde{\tau}, \quad \text{where } \tilde{\tau} \text{ represents the delay bound with } \tilde{T} \text{ being the sampling interval.}
\]

For wireless network-based NCSs, information is communicated in the form of packets. As a result, the remote controller has to convert the control inputs into packets and transmit them to the actuator through the wireless network. Then actuator applies the control inputs in response to a received control command packet. Consequently, the controller for NCSs is normally referred to as event-driven and the control input \( u(t) \) to the plant is considered piecewise constant [12] during each sampling interval, i.e.,

\[
u(t) = u_k, t \in [kT, (k + 1)T) \forall k \text{.}
\]

According to Assumption 1, there are at most \( \tilde{\tau} \) various 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 latest control input will be applied to the plant during any sampling interval \([kT, (k + 1)T) \forall k \) while the others are ignored. System states change at time instants \( kT, t_{i}^k, \quad i = 0, 1, \ldots, \tilde{\tau} \) and \( t_{i}^k < t_{i-1} \) where \( t_{i}^k = \tau_{eq}(t) - iT_{s} \) as illustrated in [12].

Since the controller is event-driven, (the controller updates the command signal based on the receipt of a 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 \([kT, (k + 1)T) \) yields

\[
x_{k+i} = Z_{t_{i-1}}(x_k, u_{k-1}, \ldots, u_{k-i}) + P_{x,t_{i-1}}(x_k, u_{k-1}, \ldots, u_{k-i})u_k, \quad y_k = Cx_k
\]

where \( x(kT) = x_k, y(kT) = y_k, \gamma((k-i)\tilde{T}) = \gamma_k \) and \( u((i-k)\tilde{T}) = u_{k-i}, i = 0, 1, 2, \ldots, \tilde{\tau} \) are pervious control inputs, and

\[
P_{x,t_{i-1}}(x_k, u_{k-1}, \ldots, u_{k-i}) = \gamma_k \int_{t_{i-1}}^{t_{i}} g(x(t))dt, Z_{t_{i-1}}(x_k, u_{k-1}, \ldots, u_{k-i}) = \gamma_k \int_{t_{i-1}}^{t_{i}} g(x(t))dt u_{k-1} + \ldots + \gamma_k \int_{t_{i-1}-i\tilde{T}}^{t_{i-1}} g(x(t))dt u_{k-i}.
\]

Using (2), define a new augment state variable

\[
z_k = [x_k^T u_{k-1}^T \ldots u_{k-i}^T] \in \mathbb{R}^{m+n} \text{ and a modified state vector consisting of the current output and input vectors as}
\]

\[
y_k = [y_k^T u_k^T] \in \mathbb{R}^{m+n}, \quad \text{where } u_{k-i}, i = 1, \ldots, \tilde{\tau} \text{ are previous control inputs. Now equation (2) can be represented as}
\]

\[
z_{k+1} = H(z_k) + L(z_k)u_k, \quad y_k = Cz_k
\]

where \( H(z_k) = [Z_{t_{i-1}}(x_k, u_{k-1}, \ldots, u_{k-i}) 0_i^{T} u_{k-i}^T] \) and \( L(z_k) = [P_{x,t_{i-1}}(x_k, u_{k-1}, \ldots, u_{k-i}) 0_i^{T} u_{k-i}^T] \), \( I_m \in \mathbb{R}^{m \times m} \) is the identity matrix 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 \( H(z_k) \in \mathbb{R}^{m+n} \) and \( L(z_k) \in \mathbb{R}^{m+n} \) are nonlinear matrix functions in terms of newly defined augmented state vector \( z_k \). Hence, the NCS dynamics (3) is still in nonlinear affine form in terms of the augmented state vector. The output matrix \( C_o \) is known and invertible since the output matrix \( C \) is considered known and invertible.

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

\[
y_k^o = C_o H(C_o^{-1} y_k^o) + C_o L(C_o^{-1} y_k^o)u_k = F(y_k^o) + G(y_k^o)u_k
\]

where \( F(y_k^o) = C_o H(C_o^{-1} y_k^o), G(y_k^o) = C_o L(C_o^{-1} y_k^o), \|F(y_k^o)\|_F \leq G_M, \) with \( \| \cdot \|_F \) denoting the Frobenius norm [11]. Here due to random delays and packet losses, \( F(y_k^o) \) and \( G(y_k^o) \) are 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 requiring adaptive techniques.

We derive the optimal adaptive controller to minimize the stochastic value function [12] as

\[
V_k = E \sum_{i=0}^{\tilde{\tau}} (x_i^T Qx_i + u_i^T R u_i) \quad k = 0, 1, 2, \ldots
\]
where \( Q \) and \( R \) are symmetric positive semi-definite and symmetric positive definite constant matrices respectively and \( E(\bullet) \) is the expectation operator (in this case the mean value) of \( \sum_{i=1}^{n} z_i^T Q z_i + u_i^T R u_i \) based on random networked-induced delays and packet losses at different time interval.

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

\[
V_k = E \left[ \sum_{i=1}^{n} \left( z_i^T Q z_i + u_i^T R u_i \right) \right] \quad k = 0, 1, 2, \ldots
\]

where \( Q_k = \begin{bmatrix} Q & 0 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{R}{d} \end{bmatrix} \) and \( R_z = \frac{1}{d} R \).

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

\[
V_k = E \left[ \sum_{i=1}^{n} \left( y_i^T Q_i y_i + u_i^T R_i u_i \right) \right] \quad k = 0, 1, 2, \ldots
\]

where \( Q_i = (C_k^{-1})^T Q_i C_k^{-1}, R_i = R_z \). Note the matrices \( Q_i \) and \( R_i \) are still symmetric positive semi-definite and symmetric positive definite respectively. Equation (7) can be also expressed as

\[
V_k = E \left[ y_k^T Q_k y_k + u_k^T R_k u_k + V_{k+1} \right]
\]

where \( V_k \) is the stochastic value function corresponding to the optimal control input \( u^*(y_k) \).

In the general nonlinear case, discrete-time HJB equation develops backward-in-time and cannot be solved exactly. This paper introduces a novel scheme in contrast to [7], which is based on time-based NDP scheme, and that can solve the optimal control for unknown NNCS when information on random delays and packet losses are not accurately known.

### III. STOCHASTIC OPTIMAL REGULATION

In this section, for solving the drawbacks of existing NDP-based schemes and to utilize the HJB equation forward-in-time, first a novel online identifier is introduced to relax the requirement on the partial knowledge of NNCS dynamics. Subsequently, a novel optimal control of NNCS is proposed by using critic and action NNs.

#### A. Online NN-Based Identifier for \( G(y_k) \)

In this part, a novel online NN-based identification is proposed to identify \( G(y_k) \). According to [9], NNCS (4) can be expressed by using following approximation representation on a compact set \( \Omega \) as

\[
y_k = F(y_{k-1}) + G(y_{k-1}) u_{k-1} = W_k^T \psi_C(y_{k-1}) U_{k-1} + \tilde{e}_{k-1,1}
\]

where \( W_k = [W_{k}^T \ W_{k}^T \ U_k^T \ \psi_C(y_{k-1})] \),

\[
U_{k-1} = [1, u_{k-1}^T]^T, \quad e_{k-1,1} = [e_{k-1,1} \ e_{k-1,1}], \quad e_{k-1,1} = e_{k-1,1} U_{k-1}, \quad \|e_{k-1,1}\| \leq \Psi \]

Using (10) and (11), the identification error is defined as

\[
e_k = y_k - \tilde{y}_k = y_k - \tilde{W}_k C \psi_C(y_k) U_k
\]

The identification error dynamics (12) are expressed as

\[
e_{k+1} = y_{k+1} - \tilde{y}_{k+1} = y_{k+1} - \tilde{W}_{k+1} C \psi_C(y_{k+1}) U_k
\]

where \( y_k = [y_{k-1} \ y_{k-1} \cdots y_{k-k} \cdots y_{k+k}] \), \( \Delta \psi_{k-1} = [\psi_C(y_{k-2}) \psi_C(y_{k-1}) \cdots \psi_C(y_{k-k}) \cdots \psi_C(y_{k+k})] \) and \( U_{k-1} = [U_{k-1} \ U_{k-2} \cdots U_{k-k}] \), \( 0 < k < k-1 \). Note equation (14) represents previous identification errors which are recalculated by using most recent NN-based weights \( \tilde{W}_{k+1} \).

Similar to (14), the auxiliary identification error dynamics are revealed to be

\[
\Sigma_{k+1} = y_{k+1} - \tilde{W}_{k+1} C \psi_C(y_{k+1}) U_k
\]

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

\[
\tilde{W}_{k+1} = U_k \Delta \psi_C(y_{k+1}) (\Delta \psi_C(y_{k+1}) U_k \Delta \psi_C(y_{k+1}))^{-1} (y_{k+1} - \alpha_{c} \Sigma_{k+1})
\]

where \( \alpha_{c} \) is the tuning parameter of the NN-based identifier satisfying \( 0 < \alpha_{c} < 1 \). Substituting (16) into (15)
\[ \Sigma_{y_1} = \alpha_c \Sigma_{y_k} \]  

**Remark 1:** We can define \( \beta_k = \psi_C(y_k^e)U_k \), and \( \beta_k \) has to persistently exiting [11] long enough for the online NN-based identifier to learn the NNCS dynamics \( G(y_k^o) \).

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

\[ e_{ck+1} = y_{ck+1} - \hat{y}_{ck+1} = \tilde{W}_{ck+1} \psi_C(y_{k+1}^e)U_{k+1} + \hat{\Sigma}_{ck} \]  

Using \( e_{ck+1} = \alpha_c e_{ck} \) from (17), we have

\[ \tilde{W}_{ck}^T \psi_C(y_k^e)U_k = \alpha_c (\tilde{W}_{ck}^T \psi_C(y_k^e)U_k - e_{ck}) + \alpha_c \hat{\Sigma}_{ck-1} - \hat{\Sigma}_{ck} \]  

Eventually, the boundedness of the identification error dynamics \( e_{ck} \) given by (12) and NN weights estimation error dynamics \( \tilde{W}_{ck} \) given by (19) will be demonstrated.

**Theorem 1 (Boundedness of the identifier).** Let the proposed NN-based identifier be defined as (11) and NN weights update law be given by (16). Under the assumption that \( \beta_k \) defined in Remark 1 satisfies the PE condition, there exists a positive constant \( \alpha_c \) satisfying \( 0 < \alpha_c < \min \{ 1, \Psi_{\text{min}}/2\Psi_{\text{max}} \} \) and computable positive constants \( B_{\text{pe}}, B_{\alpha} \), such that the identification error (12) and NN weights estimation error dynamics (19) are all uniformly ultimately bounded (UUB) [13] with ultimate bounds given by \( \| e_{ck} \| \leq B_e \) and \( \| \tilde{W}_{ck} \| \leq B_{\alpha c} \).

**Proof:** Combined with the overall stability.

**B. NN Approximation of the Value Function and Control Policy**

In [13], by using universal approximation property of NN, the stochastic value function (7) and control policy can be represented with NN and action NN as

\[ V(y_k^e) = \hat{W}_f \theta(y_k^e) + e_{\alpha f} \]  
\[ u^*(y_k^e) = \hat{W}_u \phi(y_k^e) + e_{\alpha u} \]  

where \( \hat{W}_f \) and \( \hat{W}_u \) represent the constant target NN weights, \( e_{\alpha f}, e_{\alpha u} \) are the approximation errors for critic NN and action NN respectively, and \( \theta(\bullet) \) and \( \phi(\bullet) \) are the vector activation functions for two NNs, respectively. The upper bounds for the two target NNs weights are defined as \( \| \hat{W}_f \| \leq W_{Yf} \) and \( \| \hat{W}_u \| \leq W_{Yu} \) where \( W_{Yf}, W_{Yu} \) are positive constants [7], and the approximation errors are also considered bounded as \( \| e_{\alpha f} \| \leq e_{\alpha f} \) and \( \| e_{\alpha u} \| \leq e_{\alpha u} \) where \( e_{\alpha f}, e_{\alpha u} \) are also positive constants [7] respectively. Additionally, the gradient of approximation error is assumed to be bounded as

\[ \| \partial e_{\alpha f}/\partial y_k^e \| \leq e_{\alpha f} \]  

with \( e_{\alpha f}, e_{\alpha u} \) being a positive constant [10].

The critic NN and action NN approximation of (20) can be expressed as [13]

\[ \hat{V}(y_k^e) = \hat{W}_f \theta(y_k^e), \hat{u}(y_k^e) = \hat{W}_u \phi(y_k^e) \]  

where \( \hat{W}_f \) and \( \hat{W}_u \) are the approximation of the target weights \( W_f \) and \( W_u \), respectively. In this work, the activation functions \( \mathcal{G}(\bullet), \phi(\bullet) \) are selected to be a basis function set and linearly independent [10]. Since it is required that \( V(y_k^e = 0) = 0 \) and \( u(y_k^e = 0) = 0 \), the basis functions \( \mathcal{G}(\bullet), \phi(\bullet) \) are chosen such that \( \mathcal{G}(y_k^e = 0) = 0, \phi(y_k^e = 0) = 0 \).

Substituting (21) into equation (8), it can be rewritten as

\[ W_f^T \hat{\Delta} \theta(y_{k+1}^e) + r(y_{k+1}^e, u_k) = \Delta e_{\alpha f} \]  

where \( r(y_k^e, u_k) = E(y_k^e Y_{k+1}^e u_k + u_k \hat{T} R_j y_k) \), \( \Delta \theta(y_k^e) = \theta(y_k^e) - \theta(y_k^e) \) and \( \Delta e_{\alpha f} = e_{\alpha f} - e_{\alpha f+1} \). However, when implementing the estimated value function (21), equation (22) does not hold. Therefore, using delayed values for convenience, the residual error or cost-to-go error with (24) can be expressed as

\[ e_{\alpha f} = r(y_k^e, u_k) + \hat{W}_f \Delta \theta(y_k^e) \]  

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

\[ \hat{W}_{f_{k+1}} = \hat{W}_f - \alpha_c \Delta \theta(y_k^e) \]  

**Remark 2:** It is important to note that the stochastic value function (8) and critic NN approximation (21) 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, PE condition is met by introducing noise.

As a final step in the critic NN design, define the weight estimation error as \( \hat{W}_{f_k} = W_f - \hat{W}_f \). Since \( r^T(y_k^e, u_k) = -\Delta \theta(y_{k+1}^e) W_f + \Delta e_{\alpha f} \) in equation (24), the dynamics of the critic NN weights estimation error can be rewritten as

\[ \hat{W}_{f_{k+1}} = \hat{W}_{f_k} - \alpha_c \Delta \theta(y_k^e)(\Delta \theta(y_k^e) \hat{W}_{f_k} + \Delta e_{\alpha f}) \]  

\[ \frac{\Delta \theta(y_k^e)}{\Delta \theta(y_k^e) \hat{W}_{f_k} + \Delta e_{\alpha f}} \]  

Now we need to find the control policy via action NN which minimizes the approximated value function (21). First, the action NN estimation errors are defined to be the difference between the optimal controls applied to NNCS (4) and the control signal that minimizes the estimated value function (21) with identified NNCS dynamics \( \hat{G}(y_k^e) \), which can be expressed as

\[ e_{\alpha u} = \hat{W}_u \phi(y_k^e) + \frac{1}{2} R_{\gamma}^T \hat{G}(y_k^e) \frac{\partial \theta(y_{k+1}^e)}{\partial y_{k+1}^e} \hat{W}_f \]  

Next, the update law for action NN weights is defined as

\[ \hat{W}_{u_{k+1}} = \hat{W}_u - \alpha_u \phi(y_k^e) \frac{\partial \theta(y_k^e)}{\partial y_k^e} + 1 e_{\alpha u} \]  

where \( 0 < \alpha_u < 1 \) is a positive constant. By selecting the control policy \( u_k \) to minimize the desired value function (20), the following equation results

\[ \hat{W}_f^T \phi(y_k^e) + e_{\alpha f} + \frac{1}{2} R_{\gamma}^T \hat{G}(y_k^e) \frac{\partial \theta(y_{k+1}^e)}{\partial y_{k+1}^e} \hat{W}_f + \frac{\partial e_{\alpha f}}{\partial y_{k+1}^e} = 0 \]
Substituting (28) into (26), the action NN estimation error dynamics can be rewritten as
\[
e_{ek} = -\tilde{W}^T_{ek} \phi(y^e_k) - \frac{1}{2} R^{-1} \tilde{G}^T(y^e_k) \frac{\partial \phi^T(y^e_k)}{\partial y^e_k} \tilde{W}_{ek} + \frac{1}{2} R^{-1} \tilde{G}^T(y^e_k) \frac{\partial \phi^T(y^e_k)}{\partial y^e_k} \tilde{W}_{ek}^T - e_{ek}
\]
where \(\tilde{G}(y^e_k) = G(y^e_k) - \hat{G}(y^e_k), e_{ek} = e_{ak} + \frac{1}{2} R^{-1} \tilde{G}^T(y^e_k) \frac{\partial \phi^T(y^e_k)}{\partial y^e_k} \tilde{W}_{ek}^T - e_{ek} \), satisfying \(\|e_{ek}\| \leq e_{ek} \) with \(e_{ek} \) being a positive constant, and
\[
\frac{\partial \phi^T(y^e_k)}{\partial y^e_k} \leq e_{ek}
\]

The action NN weight estimation error dynamics can be represented as
\[
\tilde{W}_{ek1} = \tilde{W}_{ek} + \alpha_e \phi^T(y^e_k) \phi(y^e_k) + 1 e_{ek}^2
\]

**C. Closed-loop Stability**

**Proof:** Consider the Lyapunov function candidate

\[
L = L_{DN} + L_{NN} + L_{CN} + L_{AN} + L_{BN}
\]

where \(L_{DN} = (y^e_k)^T y^e_k, L_{NN}, L_{CN}, L_{AN}\) and \(L_{BN}\) are defined
\[
L_{NN} = tr(\tilde{W}_{ek1}^T \Omega \tilde{W}_{ek1}), L_{CN} = tr(\tilde{W}_{ek1}^T \Omega \tilde{W}_{ek1} )
\]

\[
L_{AN} = (tr(\tilde{W}_{ek1}^T \Omega \tilde{W}_{ek1} )^2, L_{BN} = (tr(\tilde{W}_{ek1}^T \Omega \tilde{W}_{ek1} )^2
\]

with \(\Omega = 24G^2 \phi^2(\phi^2 + 1), \Lambda = 288G^2 \phi^2(\phi^2 + 1) \times (\phi^2 + 12 \phi^2)
\]

\[
\gamma = (85 \sqrt{\frac{\Xi(\phi^2 + 1)^2}{\phi^2 + 12 \phi^2}}) I + O(2 \psi_{NN}^2 + 6 \phi^2(\phi^2 + 1) + 6 \phi^2(\phi^2 + 1) I)
\]

\(I\) is positive definite matrices, \(\Xi\) is identity matrix, \(\Xi\) is defined as \(\lambda_{\min}(R) G^2 M^2, \psi_{NN}^2(\phi^2 + 1)\) and \(\lambda_{\min}(R) \) is the maximum singular value of \(R\). The first difference of (31) is given by

\[
\Delta L = \Delta L_{DN} + \Delta L_{NN} + \Delta L_{CN} + \Delta L_{AN} + \Delta L_{BN}
\]

Considering first difference \(\Delta L_{DN} = (y^e_k)^T y^e_k - (y^e_k)^T y^e_k\), using the NNCS dynamics (4), and applying the Cauchy-Schwarz inequality reveals that the first difference becomes

\[
\Delta L_{DN} = \frac{\left| (y^e_k)^T + G(y^e_k) u_k - G(y^e_k) w_k - G(y^e_k) w_k + G(y^e_k) w_k - G(y^e_k) w_k + G(y^e_k) w_k - G(y^e_k) w_k \right|}{2 \phi^2(M)^2} \leq \psi_{NN}^2(\phi^2 + 1) I
\]

or

\[
\Delta L_{NN} = \frac{\left| (y^e_k)^T + G(y^e_k) u_k - G(y^e_k) w_k - G(y^e_k) w_k \right|}{2 \phi^2(M)^2} \leq \psi_{NN}^2(\phi^2 + 1) I
\]

Next, first different \(L_{NN}\) can be expressed as

\[
\Delta L_{NN} = \frac{\left| (y^e_k)^T + G(y^e_k) u_k - G(y^e_k) w_k \right|}{2 \phi^2(M)^2} \leq \psi_{NN}^2(\phi^2 + 1) I
\]
where $0 < \phi_{\text{min}} < \phi(Y^*_k)$. The condition is ensured by the PE condition described in Remarks 1 and 2, $\|x_{k+1}\| = \frac{\alpha_r}{(\phi(Y^*_k) - \phi(Y^*_n) + 1)} < \Delta e_{CM}$, which is a bounded positive constant.

Next, the difference $L_{AN}, L_{BN}$ can be expressed as

$$ \Delta_l = -\frac{(\alpha_r - 2\alpha_f - \frac{1}{12})\Delta \phi_{2u}}{(\Delta \phi_{2u} + 1)} + $$

$$ + \frac{3(\Delta \phi_{2u})^2}{(\Delta \phi_{2u} + 1)} \left(\frac{\Delta X_{CM}^2}{(\Delta \phi_{2u} + 1)^2}\right)^2 + $$

$$ + \frac{4}{(\Delta \phi_{2u} + 1)} \left(\frac{\Delta X_{CM}^2}{(\Delta \phi_{2u} + 1)^2}\right)^2 $$

(35)

$$ \Delta_l_N = -\frac{(\alpha_r - 2\alpha_f - \frac{1}{12})\Delta \phi_{2u}}{(\Delta \phi_{2u} + 1)} + $$

$$ + \frac{3(\Delta \phi_{2u})^2}{(\Delta \phi_{2u} + 1)} \left(\frac{\Delta X_{CM}^2}{(\Delta \phi_{2u} + 1)^2}\right)^2 + $$

$$ + \frac{4}{(\Delta \phi_{2u} + 1)} \left(\frac{\Delta X_{CM}^2}{(\Delta \phi_{2u} + 1)^2}\right)^2 $$

(36)

Next, using (33), (34), (35) and (36) to form $\Delta_l$ as

$$ \Delta_l \leq (1 - 2k^2) \|x_{k+1}\|^2 \|x_{k+1}\| - \frac{288\beta \alpha_r}{(\Delta \phi_{2u} + 1)} \|x_{k+1}\|^2 - 4\eta(1 - 4\alpha_r \frac{\psi_{\text{max}}^2}{\psi_{\text{min}}^2})^2 $$

$$ - \frac{216(\alpha_r - 2\alpha_f - \frac{1}{12})\Delta \phi_{2u}}{(\Delta \phi_{2u} + 1)} \left(\frac{\Delta X_{CM}^2}{(\Delta \phi_{2u} + 1)^2}\right)^2 + $$

$$ + \frac{3(\Delta \phi_{2u})^2}{(\Delta \phi_{2u} + 1)} \left(\frac{\Delta X_{CM}^2}{(\Delta \phi_{2u} + 1)^2}\right)^2 + $$

$$ - \frac{8\psi_{\text{max}}^2}{\psi_{\text{min}}^2}(\Delta \phi_{2u} + 1) $$

(37)

where, $\eta = \psi_{\text{max}}^2 + 4\frac{\psi_{\text{max}}^2}{\psi_{\text{min}}^2}$ and

$$ \rho = \frac{\Delta \phi_{2u}}{(\Delta \phi_{2u} + 1)} $$

(38)

Therefore, $\Delta_l$ is less than zero when the following inequalities hold

$$ \|x_{k+1}\| > \frac{\Delta \phi_{2u}}{(1 - \alpha_r)} $$

$$ \|x_{k+1}\| > \frac{\Delta \phi_{2u} \psi_{\text{max}}^2}{(1 - \alpha_r) \psi_{\text{min}}^2} $$

(39)

IV. CONCLUSIONS

In this work, an online time-based approximate dynamic programming technique for NNCS was proposed using three NNs to solve the stochastic optimal regulation of NNCS with unknown dynamics in presence of random delays and packet losses. The NN identifier relaxed the requirement of input gain matrix for NNCS. The history of past cost-to-go values relaxed the need for value and policy iterations while rendering a truly forward-in-time scheme that can be implemented in practical NNCS.

The initial admissible control policy ensured that NNCS is stable while NN identifier learns the input gain matrix, the critic NN approximates the stochastic value function $V(y^*_k)$, 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.

REFERENCES


