STOCHASTIC OPTIMAL DESIGN OF NONLINEAR NETWORKED CONTROL SYSTEM USING INPUT/OUTPUT MEASUREMENTS

ABSTRACT
Stochastic optimal control design for the nonlinear networked control system (NNCS) with unknown 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 not known a priori. Neuro dynamic programming (NDP) techniques, based on value and policy iterations are not applicable for real-time control due to iterative approach and time-based NDP techniques are desirable. Therefore, in this paper, a novel NNCS representation incorporating the 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 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. 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) in the mean while the approximated control input converges close to its target value with time. Simulation results are included to show the effectiveness of the proposed scheme.

1. 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 reduced system wiring, ease of system diagnosis and maintenance, and increased 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 [4] 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. Therefore, using the stochastic optimal control theory, Nilsson [6] proposed the optimal and suboptimal controller design for NCS with random delays. Although these optimal [6] and suboptimal controller designs [6] 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 dynamic programming (NDP) technique, on the other hand, proposed by Werbos [11] 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 [8] 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 policy iterations [8] are not useful for real-time control since a significant number of iterations may be needed within any sampling interval. Also, in some cases [9], a model may be needed to iterate the value and policies. Finally, existing state-feedback based NDP schemes [8-9] are not applicable for NNCS since the unknown network imperfections such as delays and packet losses can cause instability [3] 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 are not used. Instead, 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.

2. NNCS BACKGROUND

The basic structure of NNCS considered in this paper is shown in Figure 1 where the feedback control loop is closed over a wireless network. Due to wireless network, two types of network-induced and one type of packet losses are included in this structure: (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.

![Diagram of Nonlinear Networked Control System (NNCS)](image)

Next the following assumption is made similar as the literature in NCS [18]:

**Assumption 1:**

a. Sensor is time-driven while the controller and actuator are event-driven [19].

b. Communication network is a wide area wireless network so that the two network-induced delays are considered independent, ergodic and unknown whereas their probability distribution functions are considered known [18].

c. The sum of the both delay types is bounded [18] while the initial state of the nonlinear system is deterministic [18].

In this paper, a continuous-time nonlinear affine system of the form \( \dot{x} = f(x) + g(x)u \), \( y = Cx \) is considered, where \( x, y \) and \( u \) denotes system states, output and input. 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

\[
\dot{x}(t) = f(x(t)) + \gamma(t) g(x(t)) u(t - \tau(t)), \quad y(t) = Cx(t) \tag{1}
\]

where

\[
\gamma(t) = \begin{cases} 
I^{\text{nox}} & \text{if control input is received by the actuator at time } t \\
0^{\text{nox}} & \text{if control input is lost at time } t
\end{cases}
\]

\( I^{\text{nox}} \) is identity matrix, \( u(t - \tau(t)) \) is a delayed control input, \( \tau(t) = \tau_c(t) + \tau_a(t) \leq \delta T \), where \( \delta \) represents the delay bound with \( T \) being the sampling interval.

For wireless network-based NNCS, information is communicated in the form of packets [26]. 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 NNCS is normally referred to as event-driven and the control input \( u(t) \) to the plant is considered piecewise constant [10] during each sampling interval, i.e.

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

According to Assumption 1, there are at most \( \delta \) 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 \([kT, (k + 1)T) \) while the others are ignored.

Since the controller is event-driven [19], the controller updates control command 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

\[
Z(x_k, u_{k+1}, \ldots, u_{k-\delta}) = x_k + \int_{T_k}^{(k+1)T} f(x(t)) dt + \gamma_k \int_{T_k}^{(k+1)T} g(x(t)) dt u_{k-\delta} + \ldots + \gamma_k \int_{T_k}^{(k+1)T} g(x(t)) dt u_{k-1} + \gamma_k \int_{T_k}^{(k+1)T} g(x(t)) dt u_k + y_k = Cx_k \tag{2}
\]

where \( x(kT_k) = x_k, y(kT_k) = y_k, \gamma((k-1)T_k) = \gamma_{k-1} \), and

\[
u((k-1)T_k) = u_{k-1} \quad i = 0, 1, 2, \ldots, \delta \quad \text{are pensive control inputs}
\]

Using (2), define a new augment state variable

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

\[
y_k = \begin{bmatrix} y_k^T & u_{k-1}^T & \ldots & u_{k-\delta}^T \end{bmatrix}^T \in R^{n + \delta m} , \quad \text{where } \quad u_{k-\delta}, i = 1, \ldots, \delta \quad \text{are previous control inputs}
\]

Now equation (2) can be represented as

\[
z_{k+1} = H(z_k) + L(z_k) u_k, \quad y_k = C z_k \tag{3}
\]

where \( I_m \in R^{n \times m} \) is the identity matrix,

\[
H(z_k) = \begin{bmatrix} 0 \cdots 0 \\ u_k \vdots \\ u_{k-\delta} \end{bmatrix}, \quad L(z_k) = \begin{bmatrix} P(x_k, u_{k+1}, \ldots, u_{k-\delta}) \\ I_m \end{bmatrix}
\]

and

\[
C = \begin{bmatrix} 0 \cdots 0 \\ 0 \vdots \\ 0 \cdots I_m \end{bmatrix}
\]

It is important to note that \( H(z_k) \in R^{n \times \delta m} \) and \( L(z_k) \in R^{n \times \delta m} \) are nonlinear matrix functions in terms of

---

2
newly defined augmented state vector $z_k$. Hence, the NNCS dynamics (3) is still a nonlinear affine system in terms of the augmented state vector. The output matrix $C_u$ 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

$$y_{k+1} = C_v H(C_o^{-1} y_k) + C_v L(C_o^{-1} y_k) u_k = F(y_k) + G(y_k) u_k$$

where $F(y_k) = C_v H(C_o^{-1} y_k)$ and $G(y_k) = C_v L(C_o^{-1} y_k)$, $\|G(y_k)\|_p \leq G_M$, with $\|\cdot\|_p$ denoting the Frobenius norm [20]. Here due to random delays and packet losses, $F(y_k)$ and $G(y_k)$ are real-valued functions and $F(y_k) + G(y_k)$ 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 [25] as

$$V_k = E \left[ \sum_{i=1}^{n} (x_i^T Q x_i + u_i^T R u_i) \right]$$

where $Q$ and $R$ are symmetric positive semi-definite and symmetric positive definite constant matrices respectively and $E(\cdot)$ is the expectation operator (in this case the mean value) of

$$\sum_{i=1}^{n} (x_i^T Q x_i + u_i^T R u_i)$$

based on the random networked-induced delays and packet losses at various time interval.

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

$$V_k = E \left[ \sum_{i=1}^{n} (z_i^T Q z_i + u_i^T R u_i) \right]$$

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

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

$$V_k = E \left[ \sum_{i=1}^{n} (y_i^T \tilde{w}_k^{-1} C_o^{-1} y_i + u_i^T R u_i) \right]$$

$$= E \left[ \sum_{i=1}^{n} (\tilde{w}_k^{-1} y_i^T Q z_i + u_i^T R u_i) \right]$$

where $Q_z = \left( C_o^{-1} \right)^T Q C_o^{-1}$, $R_z = R_z$. Note the matrices $Q_z$ and $R$ are still symmetric positive semi-definite and symmetric positive definite respectively.

In the general nonlinear case, discrete-time HJB equation develops backward-in-time and cannot be solved exactly. This paper introduced a novel scheme in contrast to [14,23], which is based on time-based NDP scheme, and that can solve the near optimal control for unknown NNCS when information on random delays and packet losses are not accurately known.

### 3. STOCHASTIC OPTIMAL REGULATION

#### 3.1. Online NN-Based Identification of $G(y_k)$

In this part, a novel online NN-based identification is proposed to identify $G(y_k)$. According to [20], NNCS (4) can be expressed by using following approximation representation on a compact set $\Omega$ as

$$y_{k+1} = F(y_k) + G(y_k) u_k$$

$$= \left[ \begin{array}{c} W_{f-k}^T \\ W_{g-k}^T \end{array} \right] \left[ \begin{array}{c} \tilde{y}_{k+1} \\ \tilde{u}_{k+1} \end{array} \right] + \left[ \begin{array}{c} e_{f-k} \\ e_{g-k} \end{array} \right]$$

where $W_{f-k} = W_{f-k} + \tilde{y}_{k+1}$, $W_{g-k} = W_{g-k} + \tilde{y}_{k+1}$, and $\tilde{e}_{g-k} = \tilde{e}_{g-k} + \tilde{u}_{k+1}$, with $\|\tilde{y}_{k+1}\| \leq \Psi_y$ and $\|\tilde{u}_{k+1}\| \leq \Psi_u$ are the bounds while the estimation error satisfies $\|\tilde{e}_{g-k}\| < \varepsilon_{CM}$, $\forall k$. Since the NN activation functions $\theta_l(\cdot), \theta_r(\cdot)$, and $\psi(\cdot)$ are known, NNCS dynamics $G(y_k)$ can be identified, when NN-based identifier weights $W_{f-k}$ are learned. 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 the controller design and proof in Theorem 3.

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

$$\hat{y}_k = W_{f-k} \psi(y_k)$$

Using (8) and (9), the identification error is defined as

$$e_k = y_k - \hat{y}_k = y_k - W_{f-k} \psi(y_k)$$

The identification error dynamics (10) are expressed as

$$e_{k+1} = e_k - W_{f-k} \psi(y_k)$$

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

$$\Sigma_{k+1} = Y_k - \hat{Y}_k$$

where

$$Y_k = \left[ Y_{k-1} \cdots Y_{k-(l-1)} \right]$$

and

$$\hat{Y}_k = \left[ \hat{Y}_{k-1} \cdots \hat{Y}_{k-(l-1)} \right]$$

It is desired to tune the NN identifier weights $W_{f-k}$ such that the identification error $e_k$ converges to zero asymptotically, i.e.
\( k \to \infty, e_{sk} \to 0 \). Hence, the update law for NN weights can be defined as
\[
\dot{\hat{W}}_{ck,1} = \mathcal{U}_k \Delta \psi_{ck} (\partial \psi_{ck} - 1 \Delta \psi_{ck}) \left[ (y^0_t - \alpha_C \Sigma_{sk} ) \right]
\]
(14)
where \( \alpha_C \) is the tuning parameter of the NN-based identifier satisfying 0 < \( \alpha_C < 1 \). Substituting (14) into (13)
\[
\Sigma_{sk} = \alpha_C \Sigma_{ck}
\]
(15)
Remark 2: We can define \( \beta_k = \psi_C (\dot{y}_k) U_k \), and \( \beta_k \) has to persistently exiting [20] long enough for the online NN-based identifier to learn the NNCS dynamics \( G(y^0_t) \).

Next, NN-based identifier weight estimation error is defined as \( \hat{W}_{ck} = W_C - \hat{W}_{ck} \), and recalling (11), the identification error dynamics can be rewritten as
\[
e_k = y_{sk} - \hat{y}_{sk} = W_C \psi_C \left( y_{sk} U_k + \varepsilon_{ck} - \hat{W}_{ck} \psi_{ck} \right) U_k
\]
(16)
Using \( e_{sk} = \alpha_C e_{ck} \) from (15), we have
\[
\hat{W}_{ck} = \alpha_C \hat{W}_{ck} \psi_{ck} \left( y_{sk} - \varepsilon_{ck} \right) + \alpha_C \varepsilon_{ck} - \hat{W}_{ck}
\]
(17)
Eventually, the boundedness of the identification error dynamics \( e_{ck} \) given by (10) and NN weights estimation error dynamics \( \hat{W}_{ck} \) given by (17) will be demonstrated.

**Theorem 1 (Boundedness of the identifier).** Let the proposed NN-based identifier be defined as (9) and NN weights update law be given by (14). Under the assumption that \( \beta_k \) defined in Remark 2 satisfy PE condition, there exists a positive constant \( \alpha_C \) satisfying 0 < \( \alpha_C < \min \left\{ \frac{1}{\Psi_{\min}}, \sqrt{2} \Psi_{\max} \right\} \) and computable positive constants \( B_{wc}, B_{ey} \), such that the identification error (10) and NN weights estimation errors \( \hat{W}_{ck} \) (17) are all uniformly ultimately bounded (UUB) in the mean [20] with ultimate bounds given by
\[
\left\| e_{ck} \right\| \leq B_{ey} \text{ and } \left\| \hat{W}_{ck} \right\| \leq B_{wc}
\]

**Proof:** Omitted due to limited space.

### 3.2. 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 critic and action ANNs as
\[
V(y^0_t) = W_C \phi(y^0_t) + e_{vk}
\]
(18)
and
\[
u(t, y^0_t) = W_C \delta(y^0_t) + e_{ak}
\]
(19)
respectively, where \( W_C \) and \( W_r \) represent the constant target NN weights, \( e_{vk} \) and \( e_{ak} \) are the approximation errors for critic and action ANNs respectively, and \( \phi(\bullet) \) and \( \delta(\bullet) \) are the vector activation functions for two NNs, respectively. The upper bounds for the two target ANNs weights are defined as \( \left\| W_C \right\| \leq W_{vm} \) and \( \left\| W_r \right\| \leq W_{am} \) where \( W_{vm}, W_{am} \) are positive constants [13], and the approximation errors are also considered bounded as \( \left\| e_{vk} \right\| \leq e_{vm} \) and \( \left\| e_{ak} \right\| \leq e_{am} \) where \( e_{vm}, e_{am} \) are also positive constants [13] respectively. Additionally, the gradient of approximation error is assumed to be bounded as \( \left\| e_{vk} \right\| \Delta \phi(y^0_t) \leq e_{vm} \) with \( e_{vm} \) being a positive constant [10,13].

The critic and action NNs approximation of (18) and (19) can be expressed as [13]
\[
\dot{V}(y^0_t) = \hat{W}_{vk} \Delta y^0_t \phi(y^0_t)
\]
(20)
and
\[
u(y^0_t) = \hat{W}_{ak} \phi(y^0_t)
\]
(21)
where \( \hat{W}_{vk} \) and \( \hat{W}_{ak} \) are the approximation of the target weights \( W_v \) and \( W_a \), respectively. In this work, the activation functions \( \phi(\bullet), \phi(\bullet) \) are selected to be a basis function set and linearly independent [13]. Since it is required that \( V(y^0_t) = 0 \) and \( u(y^0_t) = 0 \), the basis functions \( \phi(\bullet), \phi(\bullet) \) are chosen such that \( \phi(y^0_t) = 0 \) and \( \phi(y^0_t) = 0 \), respectively.

Substituting (21) into equation (8), it can be rewritten as
\[
W_v T\left( \phi(y^0_t) - \phi(y^0_t) \right) + 2 E_{\tau} \left( y^0_t Q, y^0_t + u^T R, u^T \right) = e_{vk} - e_{vk-1}
\]
(22)
In other words,
\[
T \frac{\Delta y^0_t}{t} = T e_{vk} + e_{vk-1}
\]
(22)
where \( r(y^0_t, u_t) = E_{\tau} \left( y^0_t Q, y^0_t + u^T R, u^T \right) \). \( \Delta \phi(y^0_t) = \phi(y^0_t) - \phi(y^0_t) \) and \( \Delta e_{vk} = e_{vk} - e_{vk-1} \). However, when implementing the estimated value function (20), equation (22) does not hold. Therefore, using delayed values for convenience, the residual error or cost-to-go error with (22) can be expressed as
\[
e_{vk} = \frac{T}{\tau^T} \left( y^0_t Q, y^0_t + u^T R, u^T \right) V(y^0_t) - V(y^0_t)
\]
(23)
\[
= \frac{T}{\tau^T} \phi(y^0_t) + \frac{\Delta \phi(y^0_t)}{\Delta \phi(y^0_t)} + \frac{\Delta g(y^0_t)}{\Delta g(y^0_t)}
\]
Based on gradient descent algorithm, the update law of critic ANNs is given by
\[
\hat{W}_{vk+1} = \hat{W}_{vk} - \alpha_v \frac{\Delta \phi(y^0_t)}{\Delta \phi(y^0_t) + 1} e_{vk}^T
\]
(24)
\[
= \hat{W}_{vk} - \alpha_v \frac{\Delta \phi(y^0_t)}{\Delta \phi(y^0_t) + 1} \frac{\Delta g(y^0_t)}{\Delta g(y^0_t)} + \frac{\Delta g(y^0_t)}{\Delta g(y^0_t) \frac{\Delta g(y^0_t)}{\Delta g(y^0_t)}}
\]
Remark 3: It is important to note that the stochastic value function (8) and critic NN approximation (20) all become zero only when \( y^0_t = 0 \). 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 \( \hat{W}_{vk} = W_v - \hat{W}_{vk} \). Since \( r^T(y^0_t, u^t) =
\]
\[-\Delta \theta^T \left( y_{k+1} \right) W_v + \Delta \epsilon_v^T \text{ in equation (24)}, \] the dynamics of the critic NN weights estimation error can be rewritten as
\[
\tilde{W}_{y_{k+1}} = \tilde{W}_{y_k} - \alpha \left[ \frac{\partial \theta^T \left( y_{k+1} \right) \Delta \theta^T \left( y_{k+1} \right) \tilde{W}_{y_k} + \Delta \epsilon_v^T}{\Delta \theta^T \left( y_{k+1} \right) \Delta \theta^T \left( y_{k+1} \right) + 1} \right]
\] (25)

Now we need to find the control policy via action NN (21) which minimizes the approximated value function (20). First, the action NN estimation errors are defined to be the difference between the optimal controls (21) applied to NNCS (4) and the control signal that minimizes the estimated value function (20) with identified NNCS dynamics \( \hat{G}(y_k) \), which can be expressed as
\[
e_{ak} = \hat{W}_{ak} \phi \left( y_k \right) + \frac{1}{2} R^a \gamma G^T \left( y_k \right) \frac{\partial \theta^T \left( y_{k+1} \right) \tilde{W}_{y_k}}{\partial y_{k+1}} + e_{ak}
\] Next, the update law for action NN weights is defined as
\[
\hat{W}_{ak+1} = \hat{W}_{ak} - \alpha \left[ \frac{\partial \theta^T \left( y_k \right) \phi \left( y_k \right) + 1}{\partial \theta^T \left( y_k \right) \phi \left( y_k \right) + 1} \right] e_{ak}
\] (26)

where \( 0 < \alpha_1 < 1 \) is a positive constant. By selecting the control policy \( u_k \) to minimize the desired value function (18), the following equation results
\[
W_{ay} \phi \left( y_k \right) + e_{ak} + \frac{1}{2} R^a \gamma G^T \left( y_k \right) \frac{\partial \theta^T \left( y_{k+1} \right) W_v}{\partial y_{k+1}} = 0
\] (28)

Substituting (28) into (26), the action NN estimation error dynamics can be rewritten as
\[
e_{ak} = -\hat{W}_{ak} \phi \left( y_k \right) - \frac{1}{2} R^a \gamma G^T \left( y_k \right) \frac{\partial \theta^T \left( y_{k+1} \right) \tilde{W}_{y_k}}{\partial y_{k+1}} - \frac{1}{2} R^a \gamma G^T \left( y_k \right) \frac{\partial \theta^T \left( y_{k+1} \right) \tilde{W}_{y_k}}{\partial y_{k+1}} + e_{ak}
\] (29)

where \( \bar{G}(y_k) = G(y_k) - \hat{G}(y_k) \), \( e_{ak} = \hat{W}_{ak} \phi \left( y_k \right) + \frac{1}{2} R^a \gamma G^T \left( y_k \right) \frac{\partial \theta^T \left( y_{k+1} \right) \tilde{W}_{y_k}}{\partial y_{k+1}} \)

satisfying \( \| e_{ak} \| \leq e_{ak} \) with \( e_{ak} \) being a positive constant, and
\[
\| e_{ak} \| \leq e_{ak}.
\]

The action NN weight estimation error dynamics can be represented as
\[
\tilde{W}_{ak+1} = \tilde{W}_{ak} + \alpha \left[ \frac{\partial \theta^T \left( y_k \right) \phi \left( y_k \right) + 1}{\partial \theta^T \left( y_k \right) \phi \left( y_k \right) + 1} \right] e_{ak}
\] (30)

\[
= \tilde{W}_{ak} - \alpha \left[ \frac{\partial \theta^T \left( y_k \right) \phi \left( y_k \right) + 1}{\partial \theta^T \left( y_k \right) \phi \left( y_k \right) + 1} \right] e_{ak} + \tilde{W}_{ak} \phi \left( y_k \right) + \frac{1}{2} R^a \gamma G^T \left( y_k \right) \frac{\partial \theta^T \left( y_{k+1} \right) \tilde{W}_{y_k}}{\partial y_{k+1}} + e_{ak}
\]

3.3. Closed-loop stability

In this section, the time-based NDP for NNCS with uncertain system dynamics and unknown network imperfections has been proposed.

For the closed-loop stability and convergence proof, the initial system outputs are considered to reside in a compact set \( \Omega \in \mathbb{R}^n \) due to the initial admissible control input \( u_0 \left( y_k \right) \). Also, in compact set \( \Omega \), the critic NN basis function and its gradient as well as the activation function of the action NN are considered bounded with \( \| \theta(y_k) \| \leq \theta_M \), \( \| \theta(y_k) \| / \theta_M \) \leq \( \phi_M \), and \( \| \phi(y_k) \| \leq \phi_M \), respectively. Further, sufficient conditions for the three NNs tuning parameters, \( \alpha_c, \alpha_v, \) and \( \alpha_u \), are derived to guarantee that all future output states never leave the compact set.

**Theorem 2:** (Convergence of the Optimal Control Signal).

Let \( u_0 \left( y_k \right) \) be any initial stabilizing control policy for the NNCS described in (4) when \( 0 < k^* / 1.2 \). Let the NN weight tuning for the identifier, critic and the action NN be provided by (14), (24) and (27), respectively. Then, there exists positive constant \( \alpha_c, \alpha_v, \alpha_u \) satisfying \( 0 < \alpha_c \leq \frac{1}{2} \gamma \left( \frac{1}{2} \right) / \| \psi \|_M \) and \( 0 < \alpha_v \leq \frac{1}{1.2} \gamma (1) / \| \psi \|_M \), and positive constants \( b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7 \) such that the system output vector \( y_k \), NN identification error \( e_{ak} \), weight estimation errors \( \tilde{W}_{ak} \), critic and action NNs weight estimation errors \( \tilde{W}_{ak} \), and \( \tilde{W}_{ak} \), respectively, are all UUB in the mean for all \( k \geq k_0 + T \) with ultimate bounds given by \( \| y_k \| \leq b_5 \left( b_5 \right) \left( b_5 \right) \leq b_7 \), \( \| \tilde{W}_{ak} \| \leq b_9 \), and \( \| \tilde{W}_{ak} \| \leq b_9 \). Further, \( \| y_k \| - u^* (y_k) \| \leq \delta_u \) for a small positive constant \( \delta_u \).

**Proof:** Omitted due to limited space.

4. SIMULATION AND RESULTS

In this section, stochastic optimal control of NNCS with unknown dynamics in presence of unknown random delays and packet losses is evaluated. The continuous-time version of original nonlinear affine system is given by
\[
\dot{x} = f(x) + g(x)u, \quad x \in \mathbb{R}^2, \quad y = Cx
\] (31)

where \( f(x) = \left[ -0.5x_1 + x_2 \right] \), \( g(x) = \left[ 0 \cos(2x_1) + 2 \right] \), and \( C = \left[ 0 \right] \). The network parameters of the NNCS are selected as [12]: a) The sampling time: \( T_s = 100ms \); b) The bound of delay is set as two, i.e. b = 2; c) The mean random delay values are \( E(\tau_m) = 80ms, E(\tau) = 150ms \). Packet losses follow Bernoulli distribution with \( p = 0.3 \).

First, the effect of random delays and packet losses for NNCS is studied. The initial state is taken as \( x_0 = [5 \ -3]^T \). The initial static control \( u_k = [-2 \ -5]x_k \), which maintains the original nonlinear affine system (31) stable, cannot maintain the NNCS stable in presence of random delays and packet losses as shown in Figure 6.

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 $y_k^a$ is generated as $y_k^a = [y_k^o, u_{k-1}, u_{k-2}] \in \mathbb{R}^{k+2}$, and the initial stabilizing policy for proposed algorithm was selected as $u_0(y_k^o) = [-2 - 5 -1 -1]$ while the activation functions for NN-based identifier were generated as $\text{tanh}(y_1^o, y_2^o, \ldots, y_{k-1}^o, (y_1^o)^T, (y_2^o)^T, \ldots, (y_{k-1}^o)^T)$, critic NN activation function were selected as the sigmoid of six order polynomial $\{y_1^o, y_2^o, \ldots, y_{k-1}^o, (y_1^o)^T, (y_2^o)^T, \ldots, (y_{k-1}^o)^T\}$, 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 $\alpha_y = 0.005$ 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 second, for the first 10 seconds exploration noise with mean zero and variance 0.06 was added to the system in order to ensure the PE condition (See Remarks 2 and 3).

Fig. 6. Effect of delays and packet losses
State Regulation Errors with Proposed Optimal control of NNCS with unknown dynamics

Fig. 7. Performance of stochastic optimal controller with unknown network imperfections.

The performance of proposed stochastic optimal controller is evaluated. This stochastic optimal controller can force the NNCS state regulation errors converge to zero even when the NNCS dynamics are unknown as shown in Figure 7. From these results, the proposed stochastic optimal control scheme to the NNCS with uncertain dynamics will have the same performance as that of an optimal controller for NNCS when the system dynamics, random delays and packet losses are known.

5. CONCLUSION
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 $V(y_k^o)$, 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.

6. REFERENCES