Received March 1, 2018, accepted April 9, 2018, date of publication May 15, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2833210

Neural-Network Aided Dynamic Control for Delivering Media Streams in Selfish Wireless Networks With Unknown Node-Selfishness

JINGLEI LI[1], XINBO GAO[1], (Senior Member, IEEE), QINGHAI YANG[1], (Member, IEEE), WEN GAO[2], AND KYUNG SUP KWAK[3], (Member, IEEE)

[1]State Key Laboratory of ISN, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
[2]College of Electrical and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China
[3]Department of Information and Communication Engineering, Inha University, Incheon 402-751, South Korea

Corresponding author: Kyung Sup Kwak (kskwak@inha.ac.kr)

This work was supported in part by the National Science Foundation for Young Scientists of China under Grant 61701371, in part by the China Postdoctoral Science Foundation under Grant 2017M613073, in part by the NSF of China under Grant 61471287, and in part by the NRF of Korea-Grant through the Korean Government under Grant NRF-2017R1A2B2012337.

ABSTRACT In this paper, we investigate the effect of selfish behaviors on the media stream delivery in selfish wireless networks with selfish relay nodes (RN). The RN's selfish behavior of forwarding media streams, depicted as its degree of node-selfishness (DeNS), is affected by both its available resources and the received incentives, while an incentive mechanism is employed to depress the selfish behaviors of the RNs having limited available resources. Under the RN's node-selfishness, the optimal incentives are controlled by the sources delivering media streams to maximize the finite-horizon utility, namely, the tradeoff between the path reliability and the incentive cost for delivering media streams during a finite time period. With the unknown node-selfishness owing to the dynamic information about the RNs' available resources and their received incentives being private, we provide a dynamic node-selfishness model to formulate the dynamic of the RN's DeNS, and further design an online neural-networks aided approximation scheme to identify the node-selfishness dynamics and to determine the optimal incentives. The convergence of this scheme is analyzed by Lyapunov techniques. Simulation results demonstrate the effectiveness of delivering media streams.

INDEX TERMS Neural networks, selfish wireless networks, node-selfishness dynamics, delivering media streams.

I. INTRODUCTION

A drastic growth of the media stream delivery has been witnessed [1], and the media streams of some nodes (namely the sources) are delivered with the aid of relay nodes (RNs) in wireless networks. With the proliferation of smart devices, the RNs have autonomic functions and would prefer to forward media streams selfishly rather than altruistically in wireless networks. The wireless network consisting of nodes exhibiting a selfish behavior is referred to as a selfish wireless network (SeWN). In SeWNs, the selfish behavior of RNs resulting from its energy resources may degrade the path reliability for the end-to-end (E2E) delivery of media streams, but the source generating media streams employs incentive mechanisms to control the RNs' selfish behaviors for delivering media streams. Accordingly, the RN's selfish behaviors, i.e., its node-selfishness, dynamically change with both its available energy resources and the received incentives, and their relationship may be complex owing to the RNs possessing different behavior characteristics [10] and/or they being in different contexts [12], [13]. Since the amount of the RN's available energy resources and the effective sensitivity of its received incentive are its private information, the RN's node-selfishness dynamic is unable to be known by other nodes. Hence, for delivering media streams, the source should firstly determine the node-selfishness dynamics of the RNs within the path. Meanwhile, with the time requirement of media streams [3], the source should guarantee the reliability of delivering media streams through the paths during a finite-horizon duration. Therefore, under the unknown node-selfishness and with the finite-horizon-duration constraint,

the source needs to effectively and dynamically control the media stream delivery by providing some appropriate incentives at network layer.

Some dynamically optimal schemes for delivering media streams have been investigated in many literatures [4]–[9]. An optimal rate allocation and admission control scheme was proposed in [5] for adaptive video streaming in wireless networks with user dynamics. For effectively delivering the multi-source video on-demand streaming, a joint routing and rate allocation algorithm was proposed in [4] over wireless mesh networks. The system metrics of delivering packets on delay, symbol error probability and packet loss probability were balanced in wireless multimedia relay networks [9]. A quality-of-user-experience (QoE)-driven cross-layer optimization scheme was developed in [6] for dynamic adaptive streaming transmission of scalable videos in wireless broadband access networks. In [7], a bit-rate switching mechanism was determined QoE metrics for video streaming service in wireless networks. A routing optimization of wireless vehicular ad-hoc networks was designed in [8] for achieving a better tradeoff between the transmission power consumption and the E2E reliability. Although these aforementioned schemes have optimized the media stream delivery from the perspective of multi-layer parameters, their feasibilities disregard the impact of the node-selfishness in distributed wireless networks.

The node-selfishness analysis has been investigated in many literatures [10]–[16]. Three energy-based node-selfishness models were elicited in [10] from the psychological behavior of human beings in mobile ad-hoc networks, while the variation of the node's residual energy with its node-selfishness was modeled in [11] as the semi-Markov process. In addition, the incentive mechanisms based on credibility and virtual currency were employed in [12] and [13] to stimulate selfish nodes for forwarding packets. To motivate nodes for cooperation in ad hoc networks, a hybrid technique of reputation-based and incentive-based mechanisms was addressed in [14]. These literatures just analyze the effect of the available resource or the incentive mechanism. For analyzing the synthetical effects of these factors, a distributed framework of the node-selfishness management was constructed in [15] to manage the RNs' node-selfishness. A dynamic incentive incentive model was proposed in [16] based on Markov decision process to determine the optimal incentives. Nevertheless, the aforementioned literatures have a precondition that the node-selfishness model is firstly known, thus it is not adaptive for the wireless networks with unknown node-selfishness. Although the offline NN-based scheme was employed in [17] to estimate the RN's node-selfishness in the case of the unknown node-selfishness dynamics, the requirement of obtaining a large amount of training data was the precondition of its feasibility.

In this paper, we investigate the effect of selfish behaviors on the media stream delivery in SeWNs with selfish RNs under the unknown node-selfishness. The RN's behavior of forwarding the packets of media streams is affected by both its available resources and the received incentives. Accordingly, we define the RN's degree of node-selfishness (DeNS) and design its node-selfishness dynamic to its selfish behaviors. Additionally, a finite-horizon utility is defined to maximize the path reliability and to minimize the incentive cost for delivering media streams during a finite time period. Under the known RN's node-selfishness dynamic, by using the neural-network (NN) techniques, the source of delivering media streams controls some optimal incentives to maximize its finite-horizon utility.

The main contributions of this paper are outlined as follows:

- A dynamic node-selfishness model is designed to formulate the dynamic of the RN's DeNS with respect to both its available resources and the received incentives. And this RN's dynamic model includes an intrinsic dynamic and an extrinsic dynamic for separating the dynamic impacts of its available resources and its received incentives, which is propitious to obtain the optimal incentives.
- An online NN-aided approximation scheme (ONAS) is conceived to determine the optimal incentives received from the source for the path reliability of delivering media streams during a finite time period. It identifies the dynamic node-selfishness model, which is uncertain for the source, and approximates the finite-horizon utility function and the optimal incentives, which is hard to be solved by traditional methods owing to the formulations related with the RNs' future DeNSs and the incentive decisions during the finite time period.
- The convergence of the ONAS is proved to be achievable by Lyapunov techniques.

The remainder of the paper is organized as follows. The system model of the media stream delivery in SeWNs is introduced in Section II. Section III depicts the problem of the optimal media stream delivery. In Section IV, the ONAS is employed for delivering media streams under the unknown node-selfishness dynamics of all RNs. Simulation results are provided in Section V, and Section VI concludes this paper.

II. SYSTEM MODEL

In this section, we introduce the network model including some selfish RNs and the RN's node-selfishness. Finally, we give the path reliability and the incentive cost of delivering media streams.

A. NETWORK MODEL

Media streams are continuous flows generated at sources and continue a finite time period [3], and this finite time period is divided into K time frames. The media streams are delivered between the source-destination pairs with the aid of some selfish RNs in SeWNs. The path of delivering media streams includes selfish RNs, and its reliability degrades owing to their selfish behaviors resulting from their limited residual energy. For raising the path reliability, a virtual-currency

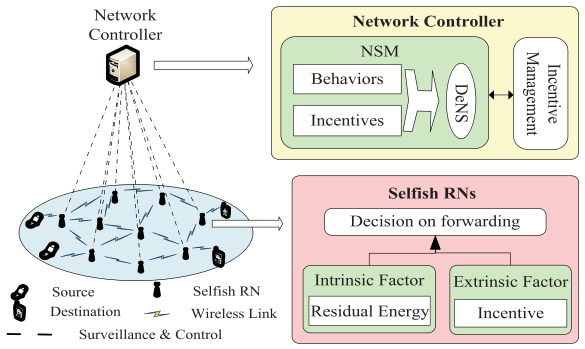

FIGURE 1. A topology example of the SeWN consisting of selfish RNs forwarding media streams and the network controller extracting the RNs' NSI.

incentive mechanism is employed to stimulate these selfish RNs for forwarding media streams. While delivering media streams, in terms of RNs' selfish behaviors, the sources select the most reliable and/or shortest one from the paths obtained by a route protocol, e.g., dynamic source routing (DSR) protocol [2]. The set of the RN in the selected path R is denoted by $\mathcal{R} = \{R_1, \cdots, R_i, \cdots, R_N\}$ with R_i being the selected RN and N being the number of RNs. At the beginning of each time frame k ($k \in \{1, 2, \cdots, K\}$), the sources provide some incentives for selfish RNs cascaded in the selected path R to depress their selfish behaviors and enhance the path reliability of delivering media streams. In SeWNs, the network controller manages the incentive resources and extracts the RNs' node-selfishness information (NSI) by surveilling their selfish behaviors, and the selfish RNs decide their forwarding behaviors in terms of their residual energy and the incentives received from the network controller, shown in Fig. 1.

B. RN's NSI

In SeWNs, the RN's residual energy is the indispensable element of determining the forwarding willingness, and the virtual-currency incentive stimulates it to forward media streams. The RN's residual energy and its received incentive, referred to as its intrinsic and extrinsic factors, respectively, affect its behavior of forwarding media streams. Here, the RN's DeNS is defined to reflect the RN's selfish behavior, related to its intrinsic and extrinsic factors, as follows.

Definition 1 (DeNS): The RN's DeNS s is defined as the degree reflecting the effect of its residual energy and its received incentive on its behavior of forwarding packets. The DeNS s varies from 0 (altruistic) to 1 (completely selfish), i.e., $s \in [0, 1]$.

The RN of abundant energy has the high forwarding willingness, and the value of its DeNS gets closes to 0. The RN of low residual energy may cherish its energy resource and has the low forwarding willingness, so the value of its DeNS gets closes to 1. In addition, the incentive enforces forwarding media streams on RNs, so the value of its DeNS decreases to 0. The high incentive largely enhances the forwarding willing of the RN, and thus the decreasing level of its DeNS in case of the high incentive is larger than that in case of the low incentive. Accordingly, the variations of both the RN's residual energy and its received incentive cause the variation of its DeNS. When building the RN's node-selfishness dynamics, its DeNS inherently varies with the variation of its residual energy, but the incentive is employed in order to control its DeNS.

C. SOURCE's UTILITY OF DELIVERING MEDIA STREAMS

Since the RNs' selfish behaviors degrade the path reliability of delivering media streams, the source should utilize some incentives to stimulate these RNs for forwarding media streams and depress their selfish behaviors. Accordingly, we configure the source's utility of delivering media streams in terms of both the path reliability and the incentive cost.

The path reliability of delivering media streams is mainly related with the RNs' selfish behaviors, i.e., their DeNSs. As the DeNS of each RN R_i ($\forall R_i \in \mathcal{R}$) increases, the path reliability of delivering media streams through the path R decreases. Hence, the path reliability of the path R is defined as the product of the reliable degrees of all selected RNs during frame k, expressed as [18]

$$P(\mathbf{S}_k) = \prod_{R_i \in \mathcal{R}} (1 - s_{i,k}), \tag{1}$$

where $\mathbf{S}_k = [s_{1,k}, \cdots, s_{N,k}]^T$ is the vector of all RNs' DeNSs during frame k with $s_{i,k}$ being the DeNS of RN R_i.

Meanwhile, due to the cascaded structure of the path R, the received incentives of the RNs within the path R are correlated. And then, we formulate the incentive cost of delivering packets through this path during frame k in a simple quadratic form as

$$C(\mathbf{u}(\mathbf{S}_k)) = \mathbf{u}^T(\mathbf{S}_k)\mathbf{Q}\mathbf{u}(\mathbf{S}_k), \tag{2}$$

where $\mathbf{u}(\mathbf{S}_k) = [u(s_{1,k}), \cdots, u(s_{i,k}), \cdots, u(s_{N,k})]^T$ is the vector of the incentives received with $u(s_{i,k})$ being the received incentive for RN R_i during frame k, $\mathbf{Q} \in \mathbb{R}^{N \times N}$ is the correlation matrix of the incentives depending on the localities of the RNs within the path R.

While delivering a media stream, the source should maintain the reliability of the selected path during K time frames. During each frame k, the source provides some incentives to all RNs within the path R for decreasing their DeNSs, which raises up the path reliability of delivering media streams through this path. Hence, the finite-horizon utility of the source is to balance the path reliability and the incentive cost from frame k to frame $K-1$ and to maximizes the path reliability with regard to its incentive cost during the terminal frame K, expressed as

$$V(\mathbf{S}_k, \mathbf{u}(\mathbf{S}_k)) = P(\mathbf{S}_K) + \sum_{j=k}^{K-1} \left(P(\mathbf{S}_j) - \pi C(\mathbf{u}(\mathbf{S}_j))\right), \tag{3}$$

where π ($\pi > 0$) is a factor, like a price parameter in [19], to balance the path reliability and the incentive cost. Here, for the sake of reading, we summarize some important notations in Table 1.

TABLE 1. List of Important Notations

Notation	Description
R	The path selected by sources for delivering media streams
R_i	The i-th RN within path R
$\mathcal{R}$	The set of the RNs in the selected path R
N	The number of the RNs within path R.
K	The number of time frames which the lasted time period of a media stream
k	The k-th time frame during the lasted time period of a media stream
$\mathbf{S}_k$	The vector of the DeNSs of all RNs within path R during frame k
$s_{i,k}$	The DeNS of the i-th RN within path R during frame k
$\mathbf{u}(\mathbf{S}_k)$	The vector of the incentives received by all RNs under the DeNSs $\mathcal{S}_k$
$u(s_{i,k})$	The received incentive for RN i during frame k
$\mathbf{Q}$	The $N \times N$ correlation matrix of the incentives
$P(\mathbf{S}_k)$	The path reliability of the path R during frame k
$C(\mathbf{u}(\mathbf{S}_k))$	The incentive cost of delivering packets through path R during frame k
$V(\mathbf{S}_k, \mathbf{u}(\mathbf{S}_k))$	The finite-horizon utility of delivering packets from frame k to frame K
π	A factor to balance the path reliability and the incentive cost
$\mathbf{f}(\mathbf{S}_k)$	The vector of the intrinsic dynamics of RNs during frame k
$\mathbf{g}(\mathbf{S}_k)$	The vector of the incentive-gain rules of RNs during frame k
$\mathbf{G}(\mathbf{S}_k)$	The diagonal matrix whose diagonal entries are the vector entries of $\mathbf{g}(\mathbf{S}_k)$
$\mathbf{W}_f$ $\mathbf{W}_g$	The NN weights of the node-selfishness dynamics
$\mathbf{W}_c$ $\mathbf{W}_I$	The NN weights of the finite-horizon utility and the optimal incentives
θ_f θ_g	The NN activation functions of the node-selfishness dynamics
θ_c θ_I	The NN activation functions of the finite-horizon utility and the optimal incentives
$\hat{\mathbf{W}}_f$ $\hat{\mathbf{W}}_g$	The estimated NN weights of the node-selfishness dynamics
$\hat{\mathbf{W}}_c$ $\hat{\mathbf{W}}_I$	The estimated NN weights of the finite-horizon utility and the optimal incentives

III. PROBLEM FORMULATION OF OPTIMAL MEDIA STREAM DELIVERY

Before delivering media streams with the aid of the selfish RNs cascaded in the selected path R, the source obtains their NSI and adjusts the incentives to stimulate their forwarding behaviors for maintaining the reliability of this path during K frames. For each frame k ($1 \leq k < K$), the source needs to maximize its finite-horizon utility $V(\mathbf{S}_k, \mathbf{u}(\mathbf{S}_k))$, expressed as

$$\begin{aligned} &V^*(\mathbf{S}_k, \mathbf{u}(\mathbf{S}_k)) \\ &= \max_{\mathbf{u}(\mathbf{S}_k)} P(\mathbf{S}_K) + \sum_{j=k}^{K-1} \left(P(\mathbf{S}_j) - \pi C(\mathbf{u}(\mathbf{S}_j))\right) \\ &= \max_{\mathbf{u}(\mathbf{S}_k)} P(\mathbf{S}_k) - \pi C(\mathbf{u}(\mathbf{S}_k)) + V^*(\mathbf{S}_{k+1}, \mathbf{u}(\mathbf{S}_{k+1})). \quad (4) \end{aligned}$$

For frame K, the source just maximizes the reliability of this path without regard to its incentive cost. By getting the derivation of Eq. (4) and setting it to zero, we obtain the optimal incentives paid by the source, expressed as

$$\mathbf{u}^*(\mathbf{S}_k) = \frac{1}{2\pi} \mathbf{Q}^{-1} \Psi(\mathbf{S}_k) \frac{\partial V^*(\mathbf{S}_{k+1}, \mathbf{u}(\mathbf{S}_{k+1}))}{\partial \mathbf{S}_{k+1}}, \quad (5)$$

where $\Psi(\mathbf{S_k}) = \frac{\partial \mathbf{S_{k+1}}}{\partial \mathbf{u}(\mathbf{S}_k)}$ reflects the variation of the DeNS $\mathbf{S}_{k+1}$ with the incentive $\mathbf{u}(\mathbf{S}_k)$.

In our considered networks, the network controller extracts the RNs' NSI by observing their behaviors of forwarding media streams, and offers some incentives to the RNs in terms of the sources' demands [15], [27]. At the beginning of each frame k ($k \in \{1, \cdots, K\}$), the source obtains the RNs' current NSI $\mathbf{S}_k$ and future NSI $\mathbf{S}_{k'}(k' \in \{k+1, \cdots, K\})$ from the network controller, and then maximizes its finite-horizon utility (cf. Eq. (4)), and finally determines the optimal incentives (cf. Eq. (5)). This controller offers these optimal incentives to the corresponding RNs for stimulating their forwarding behaviors. Not until frame $k = K$, the source continually updates the RNs' NSI, its finite-horizon utility and optimal incentives for effectively delivering media streams. Fig. 2 shows the flowchart of maximizing the source's finite-horizon utility and obtaining the optimal incentives for delivering media streams during K frames,

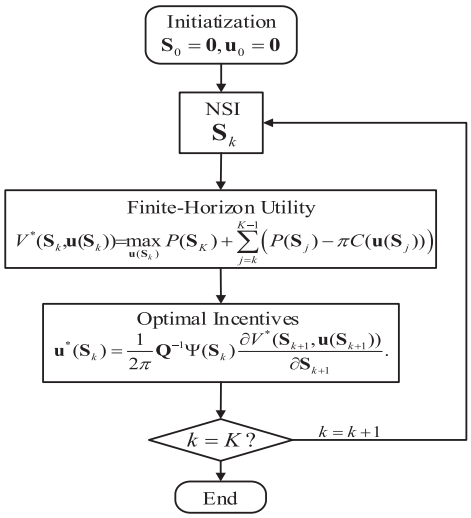

FIGURE 2. Flowchart of maximizing the source's finite-horizon utility and determining the optimal incentives during K frames.

However, in the aforementioned flowchart, there are two problems: how for network controller to obtain the RNs' future NSI and how for the sources to compute the values of

$\frac{\partial V^*(\mathbf{S}_{k+1},\mathbf{u}(\mathbf{S}_{k+1}))}{\partial \mathbf{S}_{k+1}}$ in Eq. (5) while determining the RNs' optimal incentives $\mathbf{u}^*(\mathbf{S}_k)$. Due to the effects of the RNs' selfish behaviors on both the finite-horizon utility and the incentive cost, the source should know beforehand the RNs' current NSI $\mathbf{R}_k$ and future NSI $\mathbf{R}_{k'}$ ($\forall k' \in \{k+1, \cdots, K\}$). Although the network controller has had the function of extracting the RNs' current NSI in terms of their behaviors, as shown in Fig. 4, it is unavailable for the network controller to directly know the RNs' future NSI. Additionally, since the current optimal incentive cost $\mathbf{u}^*(\mathbf{S}_k)$ is related with the derivation of the future finite-horizon utility $\frac{\partial V^*(\mathbf{S}_{k+1},\mathbf{u}(\mathbf{S}_{k+1}))}{\partial \mathbf{S}_{k+1}}$, the source should obtain its future value $\frac{\partial V^*}{\partial \mathbf{S}_{k+1}}$ while computing the value $\mathbf{u}^*(\mathbf{S}_k)$.

To overcome these issues, in the following sections, we design the ONAS, i.e., learning their dynamic node-selfishness model, the maximum finite-horizon utility, and the optimal incentive cost by using NNs [20], in order to effectively deliver media streams from frame k to frame K.

IV. ONLINE NN-AIDED APPROXIMATION SCHEME UNDER UNKNOWN NODE-SELFISHNESS DYNAMICS

In this section, the ONAS is designed for delivering media streams while the source is completely unaware of the RNs' node-selfishness. Firstly, we design a dynamic node-selfishness model in terms of the RN's intrinsic and extrinsic factors. And then, we describe the ONAS under unknown node-selfishness dynamics. Finally, we analyze its convergence.

A. NODE-SELFISHNESS DYNAMICS

From *Definition* 1, the RN's DeNS is related with its residual energy and the incentive provided by the source. The RN's DeNS varies with its residual energy, which is referred to as its intrinsic dynamic. Meanwhile, the RN's DeNS is controlled by the source's incentive to depress its selfish behaviors, which is referred to as its extrinsic dynamic. Accordingly, we design the node-selfishness dynamic of RN R_i by using the plant model of the control system [21] as

$$s_{i,k+1} = f(s_{i,k}) + g(s_{i,k})u(s_{i,k}), \tag{6}$$

where $s_{i,k+1} \in [0, 1]$ is the DeNS of RN R_i during frame $k+1$, $f(s_{i,k}) \in [0, 1]$ is the intrinsic dynamic under the RN's DeNS $s_{i,k}$ depending on the variation of its residual energy, e.g., the energy consumption of its forwarding behavior (cf. [10]), and $g(s_{i,k})u(s_{i,k}) \in [-1, 0]$ is the extrinsic dynamic under the RN's DeNS $s_{i,k}$ with $g(s_{i,k})$ being its incentive-gain rule depending on its incentive sensitivity (cf. [12]) and the control input $u(s_{i,k})$ being the received incentive. If RN R_i receives no incentive from the source during frame k, i.e., $u(s_{i,k}) = 0$, the DeNS of RN R_i is just depended on its intrinsic factor during frame $k+1$, i.e., $s_{i,k+1} = f(s_{i,k})$. When $u(s_{i,k}) \neq 0$, the DeNS of RN R_i is controlled by the source's incentive, besides of the effect of its residual energy.

According to Eq. (6), the node-selfishness dynamics, i.e., the dynamic DeNSs of all RNs within the selected path R, are expressed as

$$\mathbf{S}_{k+1} = \mathbf{f}(\mathbf{S}_k) + \mathbf{G}(\mathbf{S}_k)\mathbf{u}(\mathbf{S}_k) \tag{7}$$

$$= \begin{bmatrix} f(s_{1,k}) \\ f(s_{2,k}) \\ \vdots \\ f(s_{N,k}) \end{bmatrix} + \begin{bmatrix} g(s_{1,k}) & 0 & \cdots & 0 \\ 0 & g(s_{2,k}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g(s_{N,k}) \end{bmatrix} \begin{bmatrix} u(s_{1,k}) \\ u(s_{2,k}) \\ \vdots \\ u(s_{N,k}) \end{bmatrix}, \tag{8}$$

where $\mathbf{S}_{k+1}$ is the vector of the DeNSs of all RNs within the path R during frame $k+1$, $\mathbf{f}(\mathbf{S}_k)$ represents the vector of the intrinsic dynamics of these RNs during frame k. $\mathbf{G}(\mathbf{S}_k)$ is the diagonal matrix whose diagonal entries are the vector entries of the incentive-gain rules $\mathbf{g}(\mathbf{S}_k)$, which consist of the incentive-gain rule $g(s_{i,k})$ of RN R_i ($\forall R_i \in \mathcal{R}$) during frame k, and $||\mathbf{G}_k|| \leq G_M$. Additionally, $\mathbf{G}^T(\mathbf{S}_k) = \frac{\partial \mathbf{S}_{k+1}}{\partial \mathbf{u}(\mathbf{S}_k)} = \Psi(\mathbf{S}_k)$.

Since the residual-energy information of these RNs is their private information and the incentive-gain rules are their own characteristics, it is hard for the source to know the closed-form expressions of the intrinsic dynamics $\mathbf{f}(\mathbf{S}_k)$ and the incentive-gain rules $\mathbf{g}(\mathbf{S}_k)$. In the ONAS, we employ the NNs to determine the intrinsic dynamics $\mathbf{f}(\mathbf{S}_k)$ and the incentive-gain rules $\mathbf{g}(\mathbf{S}_k)$ of the RNs within its selected path R for delivering media streams during frame k.

B. ONLINE NN-AIDED APPROXIMATION SCHEME

The ONAS includes three parts: the NN-aided identification of the dynamic node-selfishness model, the NN-aided approximations of the finite-horizon utility $V^*(\mathbf{S}_k, \mathbf{u}(\mathbf{S}_k))$ and the optimal incentive $\mathbf{u}^*(\mathbf{S}_k)$. Meanwhile, we also describe the flowchart of this scheme.

1) NN-AIDED IDENTIFICATION OF NODE-SELFISHNESS DYNAMICS

An NN-aided identification is proposed to estimate $\mathbf{f}(\mathbf{S}_k)$ and $\mathbf{g}(\mathbf{S}_k)$, which are expressed by using NNs

$$\mathbf{f}(\mathbf{S}_k) = \mathbf{W}_f^T \theta_f(\mathbf{S}_k) + \varepsilon_{f,k}, \tag{9}$$

$$\mathbf{g}(\mathbf{S}_k) = \mathbf{W}_g^T \theta_g(\mathbf{S}_k) + \varepsilon_{g,k}, \tag{10}$$

where $\mathbf{W}_f \in \mathbb{R}^{L\times N}$ and $\mathbf{W}_g = [\mathbf{W}_{g,1}, \mathbf{W}_{g,2}, \cdots, \mathbf{W}_{g,N}] \in \mathbb{R}^{L\times N}$ are the NN weights of the node-selfishness dynamics with L being the number of hidden neurons, $\theta_f(\mathbf{S}_k) \in \mathbb{R}^{L\times 1}$ and $\theta_g(\mathbf{S}_k) \in \mathbb{R}^{L\times 1}$ are the NN activation functions of the node-selfishness dynamics, $\varepsilon_{f,k} \in \mathbb{R}^{N\times 1}$ and $\varepsilon_{g,k} = [\varepsilon_{g,k}^1, \varepsilon_{g,k}^2, \cdots, \varepsilon_{g,k}^N]^T \in \mathbb{R}^{N\times 1}$ are the NN approximation errors. Accordingly, the diagonal matrix $\mathbf{G}(\mathbf{S}_k)$ in Eq. (7) is written as

$$\mathbf{G}(\mathbf{S}_k) = \begin{bmatrix} \mathbf{W}_{g,1}^T & \mathbf{0}_{1\times L} & \cdots & \mathbf{0}_{1\times L} \\ \mathbf{0}_{1\times L} & \mathbf{W}_{g,2}^T & \cdots & \mathbf{0}_{1\times L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times L} & \mathbf{0}_{1\times L} & \cdots & \mathbf{W}_{g,N}^T \end{bmatrix}$$

$$\times \begin{bmatrix} \theta_g(\mathbf{S}_k) & \mathbf{0}_{L\times 1} & \cdots & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{L\times 1} & \theta_g(\mathbf{S}_k) & \cdots & \mathbf{0}_{L\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{L\times 1} & \mathbf{0}_{L\times 1} & \cdots & \theta_g(\mathbf{S}_k)) \end{bmatrix}$$
$$+ \begin{bmatrix} \varepsilon_{g,k}^1 & 0 & \cdots & 0 \\ 0 & \varepsilon_{g,k}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varepsilon_{g,k}^N \end{bmatrix}$$
$$= \mathbf{W}_G^T \theta_G(\mathbf{S}_k) + \varepsilon_{G,k} \tag{11}$$

where $\mathbf{W}_G \in \mathbb{R}^{NL\times N}$ and $\theta_G(\mathbf{S}_k) \in \mathbb{R}^{NL\times N}$ are the NN weights and activation functions of $\mathbf{G}(\mathbf{S}_k)$, respectively, $\varepsilon_{G,k} \in \mathbb{R}^{N\times N}$ is the NN approximation errors. By substituting Eqs. (9) and (11), the node-selfishness dynamics of Eq. (7) are rewritten as

$$\begin{aligned} \mathbf{S}_{k+1} &= (\mathbf{W}_f^T \theta_f(\mathbf{S}_k) + \varepsilon_{f,k}) + (\mathbf{W}_G^T \theta_G(\mathbf{S}_k) + \varepsilon_{G,k})\mathbf{u}(\mathbf{S}_k) \\ &= \begin{bmatrix} \mathbf{W}_f^T & \mathbf{W}_G^T \end{bmatrix} \begin{bmatrix} \theta_f(\mathbf{S}_k) & \mathbf{0}_{L\times N} \\ \mathbf{0}_{NL\times 1} & \theta_G(\mathbf{S}_k) \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{u}(\mathbf{S}_k) \end{bmatrix} \\ &\quad + \begin{bmatrix} \varepsilon_{f,k} & \varepsilon_{G,k} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{u}(\mathbf{S}_k) \end{bmatrix} \\ &= \mathbf{W}_s^T \Theta_s(\mathbf{S}_k) U(\mathbf{S}_k) + \bar{\varepsilon}_{s,k}, \end{aligned} \tag{12}$$

where $\mathbf{W}_s \in \mathbb{R}^{(N+1)L\times N}$ and $\Theta_s(\mathbf{S}_k) \in \mathbb{R}^{(N+1)L\times(N+1)}$ are the NN weights and activation functions of the dynamic DeNSs of all RNs during frame k, $U(\mathbf{S}_k) \in \mathbb{R}^{(N+1)\times 1}$, $\bar{\varepsilon}_{s,k} \in \mathbb{R}^{N\times 1}$, with $||\Theta_s(\mathbf{S}_k)|| \le \Theta_M$, $0 < \Phi_{\min} \le ||\Theta_s(\mathbf{S}_k)U(\mathbf{S}_k)|| \le \Phi_M$, and $||\bar{\varepsilon}_{s,k}|| < \bar{\varepsilon}_M$, $\forall k$. Accordingly, the RNs' DeNSs $\mathbf{S}_{k+1}$ during frame $k+1$ can be estimated as

$$\hat{\mathbf{S}}_{k+1} = \hat{\mathbf{W}}_{s,k}^T \Theta_s(\mathbf{S}_k) U(\mathbf{S}_k), \tag{13}$$

where $\hat{\mathbf{W}}_{s,k}$ is the estimated value of the NN weight $\mathbf{W}_s$ during frame k.

From Eqs. (12) and (13), the identification error of the node-selfishness dynamics during frame $k+1$ is expressed as $\mathbf{e}_{k+1} = \mathbf{S}_{k+1} - \hat{\mathbf{S}}_{k+1}$. Meanwhile, we should update the estimated NN-aided identification weights $\hat{\mathbf{W}}_s$ for determining the RNs' future DeNSs $\mathbf{S}_{k+1}$ $\forall k$. The update law for the corresponding NN weights can be expressed as

$$\begin{aligned} \hat{\mathbf{W}}_{s,k+1} = (\Theta_s(\mathbf{S}_k)U(\mathbf{S}_k)U^T(\mathbf{S}_k)\Theta_s^T(\mathbf{S}_k))^{-1} \\ \Theta_s(\mathbf{S}_k)U(\mathbf{S}_k)(\mathbf{S}_{k+1} - \alpha_s \mathbf{e}_k)^T, \end{aligned} \tag{14}$$

where α_s is the tuning parameter of the NN-aided identification satisfying $0 < \alpha_s < 1$ and $\mathbf{e}_{k+1} = \alpha_s \mathbf{e}_k$.

2) NN-AIDED APPROXIMATION OF FINITE-HORIZON UTILITY

By using NNs, the finite-horizon utility function is expressed as

$$V(\mathbf{S}_k) = \mathbf{W}_c^T \theta_c(\mathbf{S}_k) + \varepsilon_{c,k}, \tag{15}$$

where $\mathbf{W}_c \in \mathbb{R}^{L\times 1}$ are the NN weights of the finite-horizon utility, $\theta_c(\mathbf{S}_k) \in \mathbb{R}^{L\times 1}$ are the NN activation functions of the finite-horizon utility with $\theta_{cm} \le ||\theta_c(\mathbf{S}_k)|| \le \theta_{cM}$ and its gradient $\nabla\theta_c(\mathbf{S}_k)$ are bounded as $||\nabla\theta_c(\mathbf{S}_k)|| \le \nabla\theta_{cM}$, $\varepsilon_{c,k} \in \mathbb{R}$ is the NN approximation error of the finite-horizon utility. Using the approximation property of NNs [22], the approximation of this finite-horizon utility is expressed as

$$\hat{V}_k(\mathbf{S}_k) = \hat{\mathbf{W}}_{c,k}^T \theta_c(\mathbf{S}_k), \tag{16}$$

where $\hat{\mathbf{W}}_{c,k} \in \mathbb{R}^{L\times 1}$ are the estimated values of the NN weights $\mathbf{W}_c$ during frame k. Nevertheless, there exists an error between the approximated finite-horizon utility of Eq. (16) and the optimal finite-horizon utility of Eq. (4), and this error consists of two parts, expressed as $\mathbf{e}_{c,k} = \mathbf{e}_{c,k}^1 + \mathbf{e}_{c,k}^2$, where $\mathbf{e}_{c,k}^1$ and $\mathbf{e}_{c,k}^1$ are referred to as Bellman error and terminal constraint error, respectively. The Bellman error is the approximate error associated with Eq. (16) during frame k $(1 \le k \le K-1)$ [20], written as

$$\begin{aligned} \mathbf{e}_{c,k}^1 &= \hat{V}_k(\mathbf{S}_{k+1}) - \hat{V}_k(\mathbf{S}_k) + P(\mathbf{S}_k) - C(\mathbf{u}(\mathbf{S}_k)) \\ &= P(\mathbf{S}_k) - C(\mathbf{u}(\mathbf{S}_k)) + \hat{\mathbf{W}}_{c,k}^T \Delta\theta_c(\mathbf{S}_k), \end{aligned} \tag{17}$$

where $\Delta\theta_c(\mathbf{S}_k) = \theta_c(\mathbf{S}_{k+1}) - \theta_c(\mathbf{S}_k)$. Meanwhile, the terminal constraint error is the other approximate error during terminal frame K, written as

$$\mathbf{e}_{c,k}^2 = P(\mathbf{S}_k) - \hat{\mathbf{W}}_{c,K}^T \theta_c(\hat{\mathbf{S}}_K), \tag{18}$$

Based on gradient descent, the update law for the NN weights of the finite-horizon utility is expressed as

$$\hat{\mathbf{W}}_{c,k+1} = \hat{\mathbf{W}}_{c,k} - \alpha_c \frac{\bar{\theta}_c(\mathbf{S}_k)\mathbf{e}_{c,k}}{1 + \bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}, \tag{19}$$

where $\bar{\theta}_c(\mathbf{S}_k) = \Delta\theta_c(\mathbf{S}_k) - \theta_c(\hat{\mathbf{S}}_K)$ with $\bar{\theta}_{cm} \le ||\bar{\theta}_c(\mathbf{S}_k)|| \le \bar{\theta}_{cM}$, and α_c is a tuning parameter with $0 < \alpha_c < 1$.

3) NN-AIDED APPROXIMATION OF OPTIMAL INCENTIVES

The optimal incentives paid by the source in Eq. (5) have the NN-based representation, expressed as

$$\mathbf{u}^*(\mathbf{S}_k) = \mathbf{W}_I^T \theta_I(\mathbf{S}_k) + \varepsilon_{I,k}, \tag{20}$$

where $\mathbf{W}_I \in \mathbb{R}^{L\times N}$ are the NN weights of the optimal incentives, $\theta_I(\mathbf{S}_k) \in \mathbb{R}^{L\times 1}$ are the NN activation functions of the optimal incentives with $||\theta_I(\mathbf{S}_k)|| \le \theta_{IM}$, $\varepsilon_{I,k} \in \mathbb{R}^{L\times 1}$ are the NN approximation errors of the optimal incentives. By using an NN, the optimal incentives are approximated as

$$\hat{\mathbf{u}}(\mathbf{S}_k) = \hat{\mathbf{W}}_{I,k}^T \theta_I(\mathbf{S}_k), \tag{21}$$

where $\hat{\mathbf{W}}_{I,k} \in \mathbb{R}^{L\times N}$ are the estimated values of the NN weights $\mathbf{W}_I$ during frame k.

Under the NN-aided approximations of the node-selfishness dynamics and finite-horizon utility, the approximated optimal incentives $\hat{\mathbf{u}}(\mathbf{S}_k)$ applied by Eq. (21) may be different from the other approximated optimal incentive of Eq. (5) $\hat{\mathbf{u}}^*(\mathbf{S}_k)$ obtained by maximizing the approximated finite-horizon utility of Eq. (16). The NN-aided approximation errors of the optimal incentives $\tilde{\mathbf{u}}^*(\mathbf{S}_k)$ are defined as the

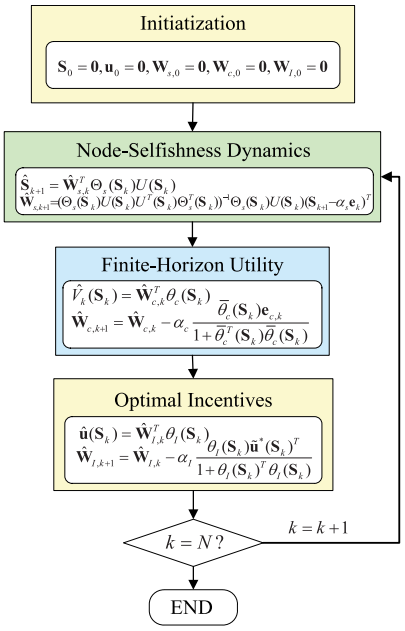

FIGURE 3. Flowchart of the online NN-aided approximation scheme.

corresponding difference value between $\hat{\mathbf{u}}(\mathbf{S}_k)$ and $\hat{\mathbf{u}}^*(\mathbf{S}_k)$, expressed as

$$\tilde{\mathbf{u}}^*(\mathbf{S}_k)=\hat{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k)-\frac{1}{2}Q^{-1}\hat{\mathbf{G}}^T(\mathbf{S}_k)\frac{\partial(\hat{\mathbf{W}}_{c,k}^T\theta_c(\mathbf{S}_{k+1}))}{\partial\mathbf{S}_{k+1}}, \tag{22}$$

with $\hat{\mathbf{G}}(\mathbf{S}_k)$ being the estimated value of the incentive gain $\mathbf{G}(\mathbf{S}_k)$. Furthermore, based on gradient descent, the update law for the NN weights of the optimal incentive is expressed as

$$\hat{\mathbf{W}}_{I,k+1}=\hat{\mathbf{W}}_{I,k}-\alpha_I\frac{\theta_I(\mathbf{S}_k)\tilde{\mathbf{u}}^*(\mathbf{S}_k)^T}{1+\theta_I(\mathbf{S}_k)^T\theta_I(\mathbf{S}_k)}, \tag{23}$$

where α_I $(0<\alpha_I<1)$ is a positive tuning parameter.

4) FLOWCHART OF ONLINE NN-AIDED APPROXIMATION SCHEME

We describe the flowchart of the ONAS for optimally delivering media streams, as shown in Fig. 3. Initially, the network controller sets $\mathbf{S}_0=0$ and $\mathbf{W}_{s,0}=0$, and the source sets $\mathbf{u}_0=0$, $\mathbf{W}_{c,0}=0$, and $\mathbf{W}_{I,0}=0$. At the beginning of each frame k ($k\in\{1,\cdots,K\}$), the network controller estimates the RNs' current and future NSI by using Eq. (13) and update the NN weights by using Eq. (14). And then, the source estimates its finite-horizon utility by using Eq. (16) and updates the NN weights by using Eq. (19). Finally, the source estimates the optimal incentives by using Eq. (21) and updates the NN weights by using Eq. (23). These optimal incentives will be provided by the network controller to stimulate the RNs' forwarding behaviors in terms of the sources' demands. Not until frame $k=K$, the network controller continually estimates the RNs' NSI, and the source continually estimates its finite-horizon utility and optimal incentives for effectively delivering media streams.

C. CONVERGENCE OF ONLINE NN-AIDED APPROXIMATION SCHEME

By using the Lyapunov technique in control theory [23], sufficient conditions for the NN tuning parameters α_s, α_c and α_I are derived to ensure the convergence results of this scheme, which is demonstrated by Theorem 1.

Theorem 1 (Boundedness for ONAS): For the ONAS, if several positive constants α_s, α_c, α_I satisfy $0<\alpha_s<\frac{1}{2}$, $0<\alpha_c<\frac{\bar{\theta}_{cm}^2}{4(1+\theta_{cM}^2)}$ and $0<\alpha_I<\frac{3}{7}$, the RNs' DeNSs $\mathbf{S}_k$, the NN identification error $\mathbf{e}_k$, weight estimation errors $\tilde{\mathbf{W}}_{s,k}$, $\tilde{\mathbf{W}}_{c,k}$ and $\tilde{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k)$ are all uniformly ultimately bounded, given by $||\mathbf{e}_k||\le\sqrt{\frac{\varepsilon_{TM}}{1-\alpha_s^2}}$, $||\tilde{\mathbf{W}}_{s,k}||\le\min\left\{\sqrt{\frac{2\varepsilon_{TM}}{\Phi_{\min}^2}},\sqrt[4]{\frac{\varepsilon_{TM}}{\Phi_{\min}^4}}\right\}$, $||\tilde{\mathbf{W}}_{c,k}||\le\min\left\{\sqrt{\frac{\varepsilon_{TM}}{\Pi_4}},\sqrt[4]{\frac{\varepsilon_{TM}}{\Pi_5}}\right\}$, $||\mathbf{S}_k||\le\sqrt{\frac{2G_M^2(1+\theta_{IM}^2)\varepsilon_{TM}}{\alpha_c^2(1-2\phi^*)}}$, and $||\tilde{\mathbf{W}}_{I,k}\theta_I(\mathbf{S}_k)||\le\sqrt{\frac{2(\theta_{IM}^2+1)\varepsilon_{TM}}{\alpha_I(3-7\alpha_I)}}$ with $\Pi_4=\alpha_c(\frac{\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-4\alpha_c)$, $\Pi_5=\alpha_c(\frac{\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-3\alpha_c)(1-\alpha_c(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-3\alpha_c))$ and ε_{TM} being an error in Eq. (43).

Proof: Refer to Appendix A. ■

The results of *Theorem 1* show the convergence of the ONAS. This scheme effectively tracks the node-selfishness dynamics, the finite-horizon utility and the optimal incentives, and finally achieves the optimal media stream delivery under the unknown node-selfishness dynamics.

V. SIMULATION RESULTS

In this section, simulation results are provided to demonstrate the effectiveness of the ONAS for delivering media streams under the unknown node-selfishness dynamics.

A. PARAMETER SETTING

In our simulated SeWN scenario, there are two completely independent RNs, marked by 'RN One' and 'RN Two', which are cascaded between one source-destination pair. Due to the effects of the RN's resource consumption and its received incentive on its forwarding behavior, e.g., their linear relationships of [10] and [24], we set the node-selfishness dynamics of these two RNs with $\mathbf{f}(\mathbf{S})=\begin{bmatrix}0.8s_1+0.2\\0.6s_2+0.4\end{bmatrix}$ and $\mathbf{G}(\mathbf{S})=\begin{bmatrix}-0.5&0\\0&-0.2\end{bmatrix}$. Meanwhile, the media streams of the source have to be completely delivered during 100 milliseconds, thus we set the terminal frame $K=100$. While delivering these media streams, these node-selfishness dynamics and their initial DeNSs are unknown to the source, thus we set their initial DeNSs as $\mathbf{S}_0=[0.5\ 0.5]^T$. For the sake of simple analysis, we set the correlation matrix of the incentives as $Q=\mathbf{I}$ with $\mathbf{I}\in\mathbb{R}^{2\times2}$ being an identity matrix, thus the finite-horizon utility $V(\mathbf{S}_k)=\prod_{i=1}^2(1-s_{i,K})+\sum_{j=k}^K(\prod_{i=1}^2(1-s_{i,k})-\pi\sum_{i=1}^2(u(s_{i,k}))^2)$ with the discount factor $\pi=0.01$.

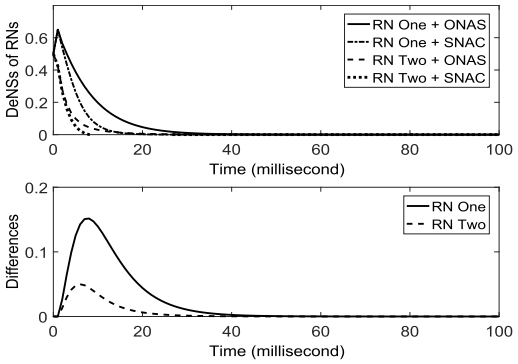

FIGURE 4. The DeNS variations of two RNs under the ONAS and the SNAC, and the corresponding differences between NN and SNAC.

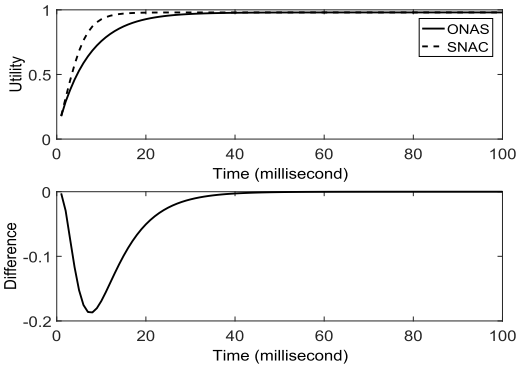

FIGURE 5. The utilities of delivering media streams under the ONAS and the SNAC, their corresponding difference.

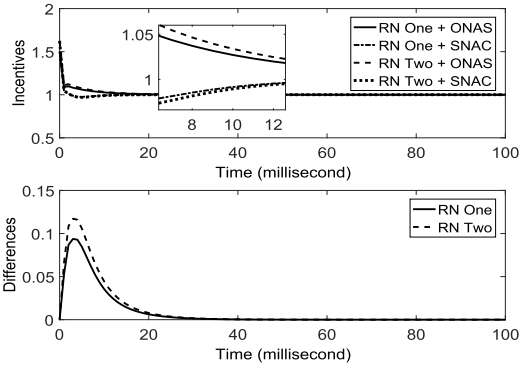

FIGURE 6. The variations of the optimal incentives under the ONAS and the SNAC, and their corresponding differences.

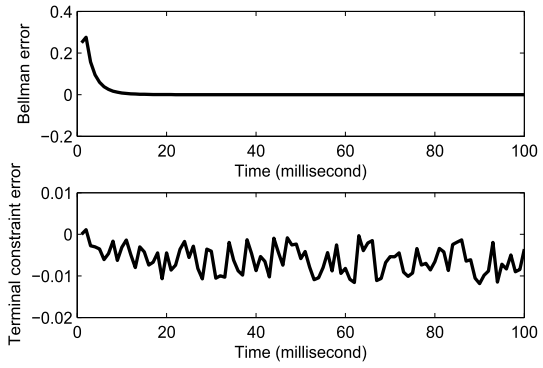

FIGURE 7. The Bellman error and terminal constraint error for estimating the finite-horizon utility.

For the ONAS, the corresponding activation functions are constructed from the expansion of the even polynomial $\sum_{\beta=1}^{M/2}(\sum_{i=1}^{2} s_{i,k})^{2\beta}$ with M being the order of approximation [25], resulting in 7 neurons and the normalized time-to-go $(K-k)/K$. And the corresponding tuning parameters are designed as $\alpha_s = 0.4$, $\alpha_c = 0.1$, and $\alpha_I = 0.01$. When the node-selfishness dynamics of these RNs are known, it is easy to determine the finite-horizon utility and incentives by using the single network adaptive critics (SNAC) of [26]. Nevertheless, since the source does not know the node-selfishness dynamics of these RNs, we employ the ONAS for estimating their DeNSs, the finite-horizon utility, and the optimal incentives. From the following simulated results, we show the effectively estimating performance of this scheme.

B. ONLINE NN-AIDED APPROXIMATION SCHEME

Fig. 4 shows the DeNS variations of these two RNs under the ONAS and the SNAC, and also depicts the RNs' corresponding difference values between them. The DeNSs of these two RNs gradually vary with time elapsion and finally tend to zeros, which means that the incentives from the source gradually enforce the RNs' altruistic behaviors of forwarding media streams. Nevertheless, during the first half of K time frames, the DeNSs of these RNs obtained by using the ONAS is larger than that by using the SNAC, owing to the bias of the estimated node-selfishness dynamics of these RNs. During the second half of K time frames, the RNs' DeNSs under the ONAS are approximately equal to that under the SNAC, which means that the ONAS has traced the RNs' DeNSs to the SNAC. Meanwhile, the differences under these two methods are distinctly shown in the bottom subfigure of Fig. 4.

Fig. 5 depicts the variations of the finite-horizon utilities of delivering media streams under these two schemes, and their corresponding differences. The utility of delivering media streams gradually change with time and finally tend to a certain value, approximately equal to 1, which means that the source has optimal path reliability and minimum incentive cost for delivering media streams. Nevertheless, since there exist the biases of the estimated node-selfishness dynamics of these RNs and the estimated utility, the delivery utility obtained by using the ONAS is smaller than that by using the SNAC during the first half of K time frames. During the second half of K time frames, the delivery utility under the ONAS is approximately equal to that under the SNAC. The corresponding differences under these two schemes are also shown in the bottom subfigure of Fig. 5.

Likewise, Fig. 6 shows the incentive variations of these two RNs under the ONAS and the SNAC, and their corresponding differences. Since the source's incentives gradually and steadily control the RNs' altruistic behaviors of forwarding media streams, the incentives of these two RNs vary with time elapsion and finally tend to a value. Meanwhile,

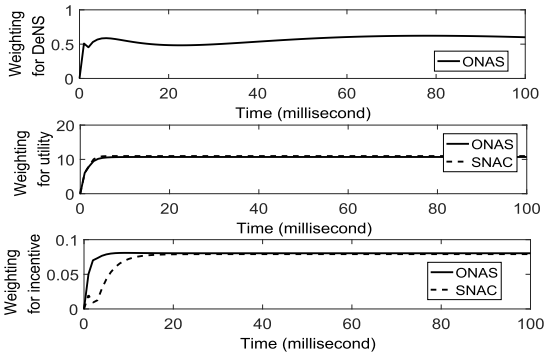

FIGURE 8. The weighting value of NNs under the ONAS and the SNAC.

there exist some differences between the incentives of these RNs obtained by using the ONAS and that by using the SNAC, which are also shown in the bottom subfigure of Fig. 6.

For analyzing the stability of the ONAS, Figs. 7 and 8 show the errors of estimating the utility and the NN weighting values between the ONAS and the SNAC. In Fig. 7, the Bellman error of Eq. (17) tends to 0 with time elapsion and the terminal constraint error of Eq. (18) is controlled in a little range. With the decreasing of the sum of these two errors, the estimated utility by using NNs gradually converges to the maximum utility. Meanwhile, owing to the large neuron number for three NNs, the Frobenius norm of the NN weighting values are shown in Fig. 8. We clearly observe from these three subfigures that their actual NN weights converge to three certain values.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we have investigated the media stream delivery through one path in SeWNs with the unknown node-selfishness dynamics. The node-selfishness dynamics have been designed to formulate the dynamics of the RN's DeNS with respect to its own resources and the received incentives. Under the unknown node-selfishness dynamics, the ONAS has been conceived to identify the dynamic node-selfishness model and also to approximate the finite-horizon utility and the optimal incentives. The convergence of the ONAS achieves with Lyapunov techniques.

For the future works, our interesting points are as follows. On the one hand, although we mainly focus on the effect of the node-selfishness on the media stream delivery in this paper, the delivery utility can be extend by introducing more impact factors, e.g., wireless channel model and the real time constraints of media streams. On the other hand, the node-selfishness dynamic proposed in this paper is just a linear and simple control model. How to extend this model for adapting to some complex scenarios, e.g., heterogenous networks where selfish nodes possess different characters on impact factors.

APPENDIX

A. PROOF OF THEOREM 1

For analyzing the convergence of the ONAS, we should analyze the convergence of updating the NN weights of estimating the node-selfishness dynamics, the finite-horizon utility and the optimal incentives. Firstly, we set the weight estimation error of the NN-aided identification as $\tilde{\mathbf{W}}_{s,k} = \mathbf{W}_s - \hat{\mathbf{W}}_{s,k}$, and obtain

$$\tilde{\mathbf{W}}_{s,k+1}\Theta_s(\mathbf{S}_k)U(\mathbf{S}_k) = \alpha_s\tilde{\mathbf{W}}_{s,k}\Theta_s(\mathbf{S}_{k-1})U(\mathbf{S}_{k-1}) + \alpha_s\bar{\varepsilon}_{s,k-1} - \bar{\varepsilon}_{s,k}. \quad (24)$$

Secondly, for analyzing the error dynamics, the NN weight error of the finite-horizon utility is expressed as $\tilde{\mathbf{W}}_{c,k} = \mathbf{W}_c - \hat{\mathbf{W}}_{c,k}$. From Eq. (3), we have

$$\begin{aligned} &P(\mathbf{S}_k) - C(\mathbf{u}(\mathbf{S}_k)) + V(\mathbf{S}_{k+1}) - V(\mathbf{S}_k) \\ &\quad = P(\mathbf{S}_k) - C(\mathbf{u}(\mathbf{S}_k)) + \mathbf{W}_c^T\theta_c(\mathbf{S}_{k+1}) \\ &\qquad + \varepsilon_{c,k+1} - \mathbf{W}_c^T\theta_c(\mathbf{S}_k) - \varepsilon_{c,k} \\ &\quad = P(\mathbf{S}_k) - C(\mathbf{u}(\mathbf{S}_k)) + \mathbf{W}_c^T\Delta\theta_c(\mathbf{S}_k) + \Delta\varepsilon_{c,k} = 0, \end{aligned} \quad (25)$$

where $\Delta\varepsilon_{c,k} = \varepsilon_{c,k+1} - -\varepsilon_{c,k}$. By substituting Eq. (25) into Eq. (17), we have

$$\begin{aligned} \mathbf{e}_{c,k}^1 &= -\mathbf{W}_c^T\Delta\theta_c(\mathbf{S}_k) - \Delta\varepsilon_{c,k} + \hat{\mathbf{W}}_{c,k}^T\Delta\theta_c(\mathbf{S}_k) \\ &= -\tilde{\mathbf{W}}_c^T\Delta\theta_c(\mathbf{S}_k) - \Delta\varepsilon_{c,k} \end{aligned} \quad (26)$$

From Eq. (18), the terminal constraint error is written as

$$\begin{aligned} \mathbf{e}_{c,k}^2 &= \mathbf{W}_c^T\theta_c(\mathbf{S}_K) + \varepsilon_{c,K} - \hat{\mathbf{W}}_{c,k}^T\theta_c(\hat{\mathbf{S}}_K) \\ &= \tilde{\mathbf{W}}_{c,k}^T\theta_c(\hat{\mathbf{S}}_K) + \mathbf{W}_{c,k}^T\tilde{\theta}_c(\mathbf{S}_K) + \varepsilon_{c,K} \end{aligned} \quad (27)$$

where $\tilde{\theta}_c(\mathbf{S}_K) = \theta_c(\mathbf{S}_K) - \theta_c(\hat{\mathbf{S}}_K)$. And then, by substituting Eqs. (26) and (27), we have

$$\mathbf{e}_{c,k} = -\tilde{\mathbf{W}}_{c,k}^T\bar{\theta}_c(\mathbf{S}_k) + \mathbf{W}_{c,k}^T\tilde{\theta}_c(\mathbf{S}_K) + \tilde{\varepsilon}_{c,k}, \quad (28)$$

where $\tilde{\varepsilon}_{c,k} = \varepsilon_{c,K} - \Delta\varepsilon_{c,k}$. Substituting Eq. (28) into Eq. (19) yields

$$\begin{aligned} \tilde{\mathbf{W}}_{c,k+1} = \tilde{\mathbf{W}}_{c,k} &- \alpha_c\frac{\bar{\theta}_c(\mathbf{S}_k)\bar{\theta}_c^T(\mathbf{S}_k)\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)} \\ &+\alpha_c\frac{\bar{\theta}_c(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_{c,k} + \tilde{\varepsilon}_{c,k})}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}, \end{aligned} \quad (29)$$

which will be used to the boundedness of the ONAS.

Finally, for the NN-aided approximation of the optimal incentive, the corresponding NN weight error is expressed as $\tilde{\mathbf{W}}_{I,k} = \mathbf{W}_I - \hat{\mathbf{W}}_{I,k}$. According to Eqs. (5), (20) and (22), it yields

$$\begin{aligned} \tilde{\mathbf{u}}^*(\mathbf{S}_k) = -\tilde{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k) &- \frac{1}{2}Q^{-1}\hat{\mathbf{G}}_k^T(\mathbf{S}_k)\frac{\partial(\hat{\mathbf{W}}_{c,k}^T\theta_c(\mathbf{S}_{k+1}))}{\partial\mathbf{S}_{k+1}} \\ &-\varepsilon_{I,k} + \frac{1}{2}Q^{-1}\mathbf{G}^T(\mathbf{S}_k)\frac{\partial(\mathbf{W}_c^T\theta_c(\mathbf{S}_{k+1}))}{\partial\mathbf{S}_{k+1}}. \end{aligned} \quad (30)$$

By substituting Eq. (30) into Eq. (23), the estimated NN weight error of the optimal incentive is formulated as

$$\begin{aligned}\tilde{\mathbf{W}}_{I,k+1} &= \mathbf{W}_I - \hat{\mathbf{W}}_{I,k+1}\\ &= \tilde{\mathbf{W}}_{I,k} + \alpha_I \frac{\theta_I(\mathbf{S}_k)\tilde{\mathbf{u}}(\mathbf{S}_k)^T}{1+\theta_I(\mathbf{S}_k)^T\theta_I(\mathbf{S}_k)}\\ &= \tilde{\mathbf{W}}_{I,k} - \alpha_I \frac{\theta_I(\mathbf{S}_k)}{1+\theta_I(\mathbf{S}_k)^T\theta_I(\mathbf{S}_k)}\\ &\quad\times\left[\frac{1}{2}Q^{-1}\hat{\mathbf{G}}_k^T(\mathbf{S}_k)\hat{\mathbf{W}}_{c,k}^T\nabla\theta_c(\mathbf{S}_{k+1}) + \tilde{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k)\right.\\ &\quad\left.+\varepsilon_{I,k} - \frac{1}{2}Q^{-1}\mathbf{G}^T(\mathbf{S}_k)\mathbf{W}_c^T\nabla\theta_c(\mathbf{S}_{k+1})\right]^T \quad (31)\end{aligned}$$

with $\nabla\theta_c(\mathbf{S}_{k+1}) = \frac{\partial\theta_c(\mathbf{S}_{k+1})}{\partial\mathbf{S}_{k+1}}$.

For the convergence of the ONAS, we employ the Lyapunov technique in control theory [23] to provide the sufficient conditions of the NN tuning parameters. Consider the Lyapunov function candidate

$$L = \frac{\alpha_I^2\Lambda}{2G_M^2(1+\theta_{IM}^2)}L_s + L_c + \Lambda L_I + L_A + L_B + L_C, \quad (32)$$

where $L_s = \mathbf{e}_k^T\mathbf{e}_k + \Phi_{\min}^2\mathbf{tr}(\tilde{\mathbf{W}}_{s,k}^T\tilde{\mathbf{W}}_{s,k}) + ||\tilde{\mathbf{W}}_{s,k}^T\Theta_s(\mathbf{S}_{k-1})U(\mathbf{S}_{k-1})||^2$, $L_c = \tilde{\mathbf{W}}_{c,k}^T\tilde{\mathbf{W}}_{c,k}$, $L_I = \mathbf{tr}\{\tilde{\mathbf{W}}_{I,k}^T\tilde{\mathbf{W}}_{I,k}\}$, $L_A = \mathbf{S}_k^T\mathbf{S}_k$, $L_B = (\tilde{\mathbf{W}}_{c,k}^T\tilde{\mathbf{W}}_{c,k})^2$, $L_C = \Phi_{\min}^2\mathbf{tr}(\tilde{\mathbf{W}}_{s,k}^T\tilde{\mathbf{W}}_{s,k})^2$, $\Lambda = \min\{\frac{\alpha_c^2}{2\alpha_I\Pi_1}, \frac{\Phi_{\min}^2}{4\alpha_I\Pi_2\theta_M^2}, \frac{\alpha_c(\theta_m^2-3\alpha_c(1+\theta_M^2))}{\alpha_I(1+\theta_M^2)}, \frac{\Phi_{\min}^4}{2\alpha_I\Pi_3\theta_M^2}\}$ with $\Pi_1 = \max\{\frac{(G_M\nabla\theta_{cM}Q)^2(8+9\alpha_I)}{2}, \frac{3\alpha_I^2(G_M\nabla\theta_{cM}Q)^2}{(6\alpha_I+16)}\}$, $\Pi_2 = \max\{\frac{(W_{cM}\nabla\theta_{cM}Q)^2(8+\alpha_I)}{2}, \frac{3\alpha_I^2(W_{cM}\nabla\theta_{cM}Q)^2}{2(3\alpha_I+8)}\}$ and $\Pi_3 = \frac{3(\nabla\theta_{cM}Q)^2(3+8\alpha_I)}{8}$. And then, the terms in Eq. (32) will be analyzed individually.

Firstly, the first-order difference of L_s is given by

$$\begin{aligned}\Delta L_s &= \mathbf{e}_{k+1}^T\mathbf{e}_{k+1} - \mathbf{e}_k^T\mathbf{e}_k + \Phi_{\min}^2\\ &\quad\times\left[\mathbf{tr}(\tilde{\mathbf{W}}_{s,k+1}^T\tilde{\mathbf{W}}_{s,k+1}) - \mathbf{tr}(\tilde{\mathbf{W}}_{s,k}^T\tilde{\mathbf{W}}_{s,k})\right]\\ &\quad+||\tilde{\mathbf{W}}_{s,k+1}^T\Theta_s(\mathbf{S}_k)U(\mathbf{S}_k)||^2\\ &\quad-||\tilde{\mathbf{W}}_{s,k}^T\Theta_s(\mathbf{S}_{k-1})U(\mathbf{S}_{k-1})||^2 \quad (33)\end{aligned}$$

and using Cauchy-Schwartz inequality together with Eqs. (14) and (24) yields

$$\begin{aligned}\Delta L_s \le &-(1-\alpha_s^2)||\mathbf{e}_k||^2 - \Phi_{\min}^2||\tilde{W}_{s,k}^T||^2 + 4||\alpha_s\bar{\varepsilon}_{s,k+1} - \bar{\varepsilon}_k||^2\\ &-(1-4\alpha_s^2)||\tilde{\mathbf{W}}_{s,k+1}^T\Theta_s(\mathbf{S}_k)U(\mathbf{S}_k)||^2. \quad (34)\end{aligned}$$

We set $4||\alpha_s\bar{\varepsilon}_{s,k+1} - \bar{\varepsilon}_{s,k}|| \le \Delta\bar{\varepsilon}_{sM}$ with $\Delta\bar{\varepsilon}_{sM}$ being a constant, which is guaranteed by the bound of $\bar{\varepsilon}_{s,k}$.

Secondly, the first-order difference of L_c is given by

$$\Delta L_c = \tilde{\mathbf{W}}_{c,k+1}^T\tilde{\mathbf{W}}_{c,k+1} - \tilde{\mathbf{W}}_{c,k}^T\tilde{\mathbf{W}}_{c,k}. \quad (35)$$

And then, substituting Eq. (19) into Eq. (35) yields

$$\begin{aligned}\Delta L_c = &\left(\tilde{\mathbf{W}}_{c,k} - \alpha_c\frac{\bar{\theta}_c(\mathbf{S}_k)\bar{\theta}_c^T(\mathbf{S}_k)\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right.\\ &\left.+\alpha_c\frac{\bar{\theta}_c(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right)^T\\ &\times\left(\tilde{\mathbf{W}}_{c,k} - \alpha_c\frac{\bar{\theta}_c(\mathbf{S}_k)\bar{\theta}_c^T(\mathbf{S}_k)\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right.\\ &\left.+\alpha_c\frac{\bar{\theta}_c(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right) - \tilde{\mathbf{W}}_{c,k}^T\tilde{\mathbf{W}}_{c,k}\\ = &-2\alpha_c\left(\frac{\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)\tilde{\mathbf{W}}_{c,k}^T\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right.\\ &\left.-\frac{\bar{\theta}_c^T(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})^T\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right)\\ &+\alpha_c^2\left\|\frac{\bar{\theta}_c(\mathbf{S}_k)\bar{\theta}_c^T(\mathbf{S}_k)\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)} - \frac{\bar{\theta}_c(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\right\|^2.\end{aligned}$$

By using Cauchy-Schwartz inequality, we further obtain

$$\begin{aligned}\Delta L_c \le &-2\alpha_c\frac{\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}||\tilde{\mathbf{W}}_{c,k}||^2\\ &+2\alpha_c\frac{\bar{\theta}_c^T(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})^T\tilde{\mathbf{W}}_{c,k}}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}\\ &+2\alpha_c^2\frac{\left\|\bar{\theta}_c(\mathbf{S}_k)\bar{\theta}_c^T(\mathbf{S}_k)\tilde{\mathbf{W}}_{c,k}\right\|^2}{\left(1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)\right)^2}\\ &+2\alpha_c^2\frac{\left\|\bar{\theta}_c(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})\right\|^2}{\left(1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)\right)^2}. \quad (36)\end{aligned}$$

Recall that the time span of interest is finite and $\bar{\theta}_c(\mathbf{S}_k)$ is a smooth function, then it is bounded by $0 < \bar{\theta}_{cm} \le ||\bar{\theta}_{c,k}|| < 3\theta_{cM}$. Then, by separating the term $\bar{\theta}_c(\mathbf{S}_k)(\tilde{\theta}_c^T(\mathbf{S}_K)\mathbf{W}_c+\tilde{\varepsilon}_{c,k})$ with $||\tilde{\theta}_c^T(\mathbf{S}_K)|| \le 2\theta_{cM}$, $||\mathbf{W}_c|| \le W_{cM}$ and $||\tilde{\varepsilon}_{c,k}|| \le 3\varepsilon_{cM}$ together with $\frac{\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)}{1+\bar{\theta}_c^T(\mathbf{S}_k)\bar{\theta}_c(\mathbf{S}_k)} \le 1$, we obtain

$$\begin{aligned}\Delta L_c \le &-2\alpha_c\frac{\bar{\theta}_{cm}^2}{1+9\theta_{cM}^2}||\tilde{\mathbf{W}}_{c,k}||^2\\ &+\alpha_c^2||\tilde{\mathbf{W}}_{c,k}||^2 + 8\alpha_c\frac{\theta_{cM}^2W_{cM}^2}{1+\bar{\theta}_{cm}^2} + 18\varepsilon_{cM}^2\\ &+2\alpha_c^2||\tilde{\mathbf{W}}_{c,k}||^2 + 2\alpha_c^2(8\theta_{cM}^2W_{cM}^2 + 18\varepsilon_{cM}^2)\\ = &-\alpha_c\left(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2} - 3\alpha_c\right)||\tilde{\mathbf{W}}_{c,k}||^2 + \bar{\varepsilon}_{cM}^2, \quad (37)\end{aligned}$$

where $\bar{\varepsilon}_{cM}^2 = \alpha_c(1+2\alpha_c)(8\theta_{cM}^2W_{cM}^2 + 18\alpha_c\varepsilon_{cM}^2) + 18\varepsilon_{cM}^2$.

Thirdly, the first-order difference of ΔL_I is expressed as

$$\Delta L_I = \mathbf{tr}\{\tilde{\mathbf{W}}_{I,k+1}^T\tilde{\mathbf{W}}_{I,k+1}\} - \mathbf{tr}\{\tilde{\mathbf{W}}_{I,k}^T\tilde{\mathbf{W}}_{I,k}\}. \quad (38)$$

By substituting Eq. (31) into Eq. (38), the upper bound of ΔL_I yields

$$\begin{aligned}\Delta L_I \le &-\frac{\alpha_I(3-3\alpha_I)}{2(\theta_{IM}^2+1)}||\tilde{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k)||^2\\ &+\frac{\alpha_I}{2}(G_M\nabla\theta_{cM}Q)^2(8+9\alpha_I)||\tilde{\mathbf{W}}_{c,k}||^2\end{aligned}$$

$$
\begin{aligned}
&+\frac{\alpha_I}{2}(W_{cM}\nabla\theta_{cM}Q)^2(8+\alpha_I)||\tilde{\mathbf{G}}_k||^2\\
&+\frac{3\alpha_I}{8}(\nabla\theta_{cM}Q)^2(8+3\alpha_I)||\tilde{\mathbf{W}}_{c,k}||^4\\
&+\frac{3\alpha_I(G_M\nabla\theta_{cM}Q)^2}{2(8+3\alpha_I)}||\tilde{\mathbf{W}}_{c,k}||^2\\
&+\frac{3\alpha_I(G_M\nabla\theta_{cM}Q)^2}{2(8+3\alpha_I)}\alpha_I^2W_{cM}^2||\tilde{\mathbf{G}}_k||^2\\
&+\tilde{\varepsilon}_{I_1M}^2+\tilde{\varepsilon}_{I_2M}^2+\tilde{\varepsilon}_{I_3M}^2+\tilde{\varepsilon}_{I_4M}^2\\
\le&-\frac{\alpha_I(3-3\alpha_I)}{2(\theta_{IM}^2+1)}||\tilde{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k)||^2+2\alpha_I\Pi_1||\tilde{\mathbf{W}}_{c,k}||^2\\
&+2\alpha_I\Pi_2||\tilde{\mathbf{G}}_k||^2+\alpha_I\Pi_3||\tilde{\mathbf{W}}_{c,k}||^4+\alpha_I\Pi_3||\tilde{\mathbf{G}}_k||^4\\
&+\tilde{\varepsilon}_{I_1M}^2+\tilde{\varepsilon}_{I_2M}^2+\tilde{\varepsilon}_{I_3M}^2+\tilde{\varepsilon}_{I_4M}^2, \qquad (39)
\end{aligned}
$$

where $\tilde{\varepsilon}_{I_1M}^2=\frac{8\alpha_I(1+\alpha_I)^3}{(1+\theta_{Im}^2)\bar{\varepsilon}_{IM}^2}$, $\tilde{\varepsilon}_{I_2M}^2=\frac{2\alpha_I^3\bar{\varepsilon}_{IM}^2}{(1+\theta_{Im}^2)(8+9\alpha_I)}$, $\tilde{\varepsilon}_{I_3M}^2=\frac{2\alpha_I^3\bar{\varepsilon}_{IM}^2}{(1+\theta_{Im}^2)(8+\alpha_I)}$ and $\tilde{\varepsilon}_{I_4M}^2=\frac{6\alpha_I^3\bar{\varepsilon}_{IM}}{4(1+\theta_{Im}^2)(8+3\alpha_I)}+\frac{3\alpha_I^3W_{cM}G_M\nabla\theta_{cM}Q}{4(1+\theta_{Im}^2)(8+3\alpha_I)}$.

According to the dynamic DeNS in Eq. (7), we obtain $||\mathbf{f}(\mathbf{S}_k)+\mathbf{G}(\mathbf{S}_k)\mathbf{u}^*(\mathbf{S}_k)||^2\le\phi^*||\mathbf{S}_k||^2$ with $0\le\phi^*\le1/2$ is a constant. The first-order difference ΔL_A can be expressed as

$$
\begin{aligned}
\Delta L_A&=\mathbf{S}_{k+1}^T\mathbf{S}_{k+1}-\mathbf{S}_k^T\mathbf{S}_k\\
&=||\mathbf{f}(\mathbf{S}_k)+\mathbf{G}(\mathbf{S}_k)\hat{\mathbf{u}}(\mathbf{S}_k)||^2-||\mathbf{S}_k||^2\\
&\le 2||\mathbf{f}(\mathbf{S}_k)+\mathbf{G}(\mathbf{S}_k)\mathbf{u}^*(\mathbf{S}_k)||^2-||\mathbf{S}_k||^2\\
&\quad+2||\mathbf{G}(\mathbf{S}_k)\mathbf{u}^*(\mathbf{S}_k)-\mathbf{G}(\mathbf{S}_k)\hat{\mathbf{u}}(\mathbf{S}_k)||^2\\
&\le-(1-2\phi^*)||\mathbf{S}_k||^2+4G_M^2||\tilde{\mathbf{W}}_{I,k}\theta_I(\mathbf{S}_k)||^2+4\varepsilon_{I,M}. \qquad (40)
\end{aligned}
$$

Furthermore, by using Eq. (40), we have

$$
\begin{aligned}
\Delta L_B&=(\tilde{\mathbf{W}}_{c,k+1}^T\tilde{\mathbf{W}}_{c,k+1})^2-(\tilde{\mathbf{W}}_{c,k}^T\tilde{\mathbf{W}}_{c,k})^2\\
&\le\left(\left(2-\alpha_c\left(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-3\alpha_c\right)\right)||\tilde{\mathbf{W}}_{c,k}||^2+\bar{\varepsilon}_{cM}^2\right)\\
&\quad\times\left(-\alpha_c\left(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-3\alpha_c\right)||\tilde{\mathbf{W}}_{c,k}||^2+\bar{\varepsilon}_{cM}^2\right)\\
&\le-\alpha_c\left(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-\alpha_c\right)\left(\frac{3}{2}-\alpha_c\left(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-\alpha_c\right)\right)||\tilde{\mathbf{W}}_{c,k}||^2\\
&\quad+\left(\frac{3}{\alpha_c\left(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-\alpha_c\right)}\right)\bar{\varepsilon}_{cM}^2. \qquad (41)
\end{aligned}
$$

Recalling the NN weights estimation error dynamics (24) and applying Cauchy-Swartz inequality, we have

$$
\begin{aligned}
\Delta L_C&=(\Phi_{\min}^2\mathbf{tr}(\tilde{\mathbf{W}}_{s,k+1}^T\tilde{\mathbf{W}}_{s,k+1}))^2-(\Phi_{\min}^2\mathbf{tr}(\tilde{\mathbf{W}}_{s,k}^T\tilde{\mathbf{W}}_{s,k}))^2\\
&\quad+||\tilde{\mathbf{W}}_{s,k+1}^T\Theta_s(\mathbf{S}_k)U(\mathbf{S}_k)||^4\\
&\quad-||\tilde{\mathbf{W}}_{s,k}^T\Theta_s(\mathbf{S}_{k-1})U(\mathbf{S}_{k-1})||^4\\
&\le-\Phi_{\min}^4||\tilde{\mathbf{W}}_{s,k}||^4\\
&\quad-(1-16\alpha_s^4)||\tilde{\mathbf{W}}_{s,k}^T\Theta_s(\mathbf{S}_{k-1})U(\mathbf{S}_{k-1})||^4+\Delta\bar{\varepsilon}_{sM}^2\\
&\le-\Phi_{\min}^4||\tilde{\mathbf{W}}_{s,k}||^4+\Delta\bar{\varepsilon}_{sM}^2. \qquad (42)
\end{aligned}
$$

Finally, by using all the terms of Eq. (32), we get the upper bound of ΔL as

$$
\begin{aligned}
\Delta L&=\frac{\alpha_I^2\Lambda}{2G_M^2(1+\theta_{IM}^2)}\Delta L_s+\Delta L_c+\Lambda\Delta L_I\\
&\quad+\Delta L_A+\Delta L_B+\Delta L_C\\
&\le-\frac{\alpha_c^2(1-2\phi^*)}{2G_M^2(1+\theta_{IM}^2)}||\mathbf{S}_k||^2\\
&\quad-(1-\alpha_s^2)||\mathbf{e}_k||^2-\frac{1}{2}\Phi_{\min}^2||\tilde{\mathbf{W}}_{s,k}^T||^2\\
&\quad-\Pi_4||\tilde{\mathbf{W}}_{c,k}||^2-\frac{\alpha_I(3-7\alpha_I)}{2(\theta_{IM}^2+1)}||\tilde{\mathbf{W}}_{I,k}^T\theta_I(\mathbf{S}_k)||^2\\
&\quad-\Pi_5||\tilde{\mathbf{W}}_{c,k}||^4+\Phi_{\min}^4||\tilde{\mathbf{W}}_{s,k}||^4+\varepsilon_{TM}, \qquad (43)
\end{aligned}
$$

where $\Pi_4=\alpha_c(\frac{\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-4\alpha_c)$, $\Pi_5=\alpha_c(\frac{\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-3\alpha_c)(1-\alpha_c(\frac{2\bar{\theta}_{cm}^2}{1+\theta_{cM}^2}-3\alpha_c))$ and $\varepsilon_{TM}=\frac{3}{\Theta_{\min}^2}\Delta\bar{\varepsilon}_{sM}^2+\frac{3}{\Theta_{\min}^2}\bar{\varepsilon}_{cM}^4+\Lambda(\tilde{\varepsilon}_{I_1M}^2+\tilde{\varepsilon}_{I_2M}^2+\tilde{\varepsilon}_{I_3M}^2+\tilde{\varepsilon}_{I_4M}^2)+\frac{\Delta\bar{\varepsilon}_{sM}^2+\bar{\varepsilon}_{cM}^2+2\alpha_I^2\bar{\varepsilon}_{IM}^2}{1+\theta_{IM}^2}$. Therefore, ΔL is less than zero when the following inequalities hold

$$||\mathbf{e}_k||>\sqrt{\frac{\varepsilon_{TM}}{1-\alpha_s^2}}=B_e \qquad (44)$$

or

$$||\tilde{\mathbf{W}}_{s,k}||>\min\left\{\sqrt{\frac{2\varepsilon_{TM}}{\Phi_{\min}^2}},\sqrt[4]{\frac{\varepsilon_{TM}}{\Phi_{\min}^4}}\right\}=B_s \qquad (45)$$

or

$$||\tilde{\mathbf{W}}_{c,k}||>\min\left\{\sqrt{\frac{\varepsilon_{TM}}{\Pi_4}},\sqrt[4]{\frac{\varepsilon_{TM}}{\Pi_5}}\right\}=B_c \qquad (46)$$

or

$$||\mathbf{S}_k||>\sqrt{\frac{2G_M^2(1+\theta_{IM}^2)\varepsilon_{TM}}{\alpha_c^2(1-2\phi^*)}}=B_A \qquad (47)$$

or

$$||\tilde{\mathbf{W}}_{I,k}\theta_I(\mathbf{S}_k)||>\sqrt{\frac{2(\theta_{IM}^2+1)\varepsilon_{TM}}{\alpha_I(3-7\alpha_I)}}=B_I. \qquad (48)$$

Meanwhile, for guaranteeing that the parameters B_e, B_s, B_c, B_A and B_I are positive, we obtain the ranges of α_s, α_c and α_I.

REFERENCES

[1] Y. Xiao and H. Li, "Voice and video transmissions with global data parameter control for the IEEE 802.11e enhance distributed channel access," *IEEE Trans. Parallel Distrib. Syst.*, vol. 15, no. 11, pp. 1041–1053, Nov. 2004.

[2] N. Zhou, H. Wu, and A. A. Abouzeid, "The impact of traffic patterns on the overhead of reactive routing protocols," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 547–560, Mar. 2005.

[3] Y. Chen, B. Zhang, Y. Liu, and W. Zhu, "Measurement and modeling of video watching time in a large-scale Internet video-on-demand system," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2087–2098, Dec. 2013.

[4] N.-S. Vo, T. Q. Duong, H.-J. Zepernick, and M. Fiedler, "A cross-layer optimized scheme and its application in mobile multimedia networks with QoS provision," *IEEE Syst. J.*, vol. 10, no. 2, pp. 817–830, Jun. 2016.

[5] V. Joseph, S. Borst, and M. I. Reiman, "Optimal rate allocation for video streaming in wireless networks with user dynamics," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 820–835, Apr. 2016.

[6] M. Zhao, X. Gong, J. Liang, W. Wang, X. Que, and S. Cheng, "QoE-driven cross-layer optimization for wireless dynamic adaptive streaming of scalable videos over HTTP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 451–465, Mar. 2015.
[7] Y. Xu, Y. Zhou, and D.-M. Chiu, "Analytical QoE models for bit-rate switching in dynamic adaptive streaming systems," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2734–2748, Dec. 2014.
[8] Z. Ding and K. K. Leung, "Cross-layer routing using cooperative transmission in vehicular ad-hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 571–581, Mar. 2011.
[9] M. Egan, P. L. Yeoh, M. Elkashlan, and I. B. Collings, "A new cross-layer user scheduler for wireless multimedia relay networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 301–311, Jan. 2013.
[10] E. Ataie and A. Movaghar, "Performance evaluation of mobile ad hoc networks in the presence of energy-based selfishness," in *Proc. 3rd Int. Conf. Broadband Commun., Netw., Syst.*, Oct. 2006, pp. 1–6.
[11] F. Xing and W. Wang, "On the survivability of wireless ad hoc networks with node misbehaviors and failures," *IEEE Trans. Dependable Secure Comput.*, vol. 7, no. 3, pp. 284–299, Jul. 2010.
[12] J. Konorski and S. Szott, "Credibility of threats to jam anonymous traffic remapping attacks in ad hoc WLANs," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 624–627, Mar. 2017.
[13] K. Wang, L. Yuan, T. Mizayaki, Y. Sun, and S. Guo, "Antieavesdropping with selfish jamming in wireless networks: A Bertrand game approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6268–6279, Jul. 2017.
[14] R. Kaushik and J. Singhai, "Enhanced node cooperation technique for outwitting selfish nodes in an ad hoc network," *IET Netw.*, vol. 4, no. 2, pp. 148–157, Mar. 2015.
[15] J. Li, Q. Yang, P. Gong, and K. S. Kwak, "End-to-end multiservice delivery in selfish wireless networks under distributed node-selfishness management," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1132–1142, Mar. 2016.
[16] Z. Yang, H. Tian, S. Fan, and G. Chen, "Dynamic incentive design in content dissemination process through D2D communication," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1799–1802, Aug. 2017.
[17] J. Li, Q. Yang, and K. S. Kwak, "Neural-network based optimal dynamic control of delivering packets in selfish wireless networks," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2246–2249, Dec. 2015.
[18] A. E. Zonouz, L. Xing, V. M. Vokkarane, and Y. L. Sun, "Reliability-oriented single-path routing protocols in wireless sensor networks," *IEEE Sensors J.*, vol. 14, no. 11, pp. 4059–4068, Nov. 2014.
[19] K. Zhu, D. Niyato, P. Wang, and Z. Han, "Dynamic spectrum leasing and service selection in spectrum secondary market of cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1136–1145, Mar. 2012.
[20] Q. Zhao, H. Xu, and S. Jagannathan, "Neural network-based finite-horizon optimal control of uncertain affine nonlinear discrete-time systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 3, pp. 486–499, Mar. 2015.
[21] H. Xu, Q. Zhao, and S. Jagannathan, "Finite-horizon near-optimal output feedback neural network control of quantized nonlinear discrete-time systems with input constraint," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1776–1788, Aug. 2015.
[22] J. Sarangapani, *Neural Network Control of Nonlinear Discrete-Time Systems*. Boca Raton, FL, USA: CRC Press, 2006.
[23] D. Swaroop, J. K. Hedrick, P. P. Yip, and J. C. Gerdes, "Dynamic surface control for a class of nonlinear systems," *IEEE Trans. Autom. Control*, vol. 45, no. 10, pp. 1893–1899, Oct. 2000.
[24] D. Niyato and E. Hossain, "Competitive pricing in heterogeneous wireless access networks: Issues and approaches," *IEEE Netw.*, vol. 22, no. 6, pp. 4–11, Nov. 2008.
[25] 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. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1118–1129, Jul. 2012.
[26] A. Heydari and S. N. Balakrishnan, "Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 1, pp. 145–157, Jan. 2013.
[27] J. Li, Q. Yang, and K. S. Kwak, "stimulating multi-service forwarding under node-selfishness information in selfish wireless networks," *IEICE Trans. Commun.*, vol. E99B, no. 7, pp. 1426–1434, 2016.

JINGLEI LI received the B.S. degree in electronic information engineering from The PLA Information Engineering University in 2008 and the M.S. and Ph.D. degrees in communication and information systems from Xidian University in 2011 and 2016, respectively. He is currently with Xidian University. His research interests include wireless network connectivity and node-selfishness management.

XINBO GAO (M'02–SM'07) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. He was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan, from 1997 to 1998. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Professor of pattern recognition and intelligent systems and the Director of the State Key Laboratory of Integrated Services Networks, Xidian University. He has authored five books and around 150 technical articles in refereed journals and proceedings, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, and *Pattern Recognition* in his areas of expertise. His current research interests include computational intelligence, machine learning, computer vision, pattern recognition, and wireless communications.

QINGHAI YANG (M'08) received the B.S. degree in communication engineering from the Shandong University of Technology, China, in 1998, the M.S. degree in information and communication systems from Xidian University, China, in 2001, and Ph.D. degree in communication engineering from Inha University, South Korea, in 2007 with university-president award. From 2007 to 2008, he was a Research Fellow with UWB-ITRC, South Korea. Since 2008, he has been with Xidian University, China. His current research interest lies in the fields of autonomic communication, content delivery networks, and LTE-A techniques.

WEN GAO received the B.S. degree in electronic information engineering from the Henan University of Technology in 2011 and the Ph.D. degree in cryptography from Xidian University in 2017. She is currently with the Shaanxi University of Science and Technology. Her research interests include lattice-based cryptography and quantum computation and quantum attack.

KYUNG SUP KWAK (M'81) received the B.S. degree from Inha University, Inchon, South Korea, in 1977, the M.S. degree from the University of Southern California in 1981, and the Ph.D. degree from the University of California at San Diego in 1988, under the Inha University Fellowship and the Korea Electric Association Abroad Scholarship Grants. From 1988 to 1989, he was a Member of Technical Staff at Hughes Network Systems, San Diego, CA, USA. From 1989 to 1990, he was with the IBM Network Analysis Center, Research Triangle Park, NC, USA. Since 1990, he has been with the School of Information and Communication, Inha University, as a Professor. He had been the Chairman of the School of Electrical and Computer Engineering from 1999 to 2000 and the Dean of the Graduate School of Information Technology and Telecommunications, Inha University, from 2001 to 2002. He is currently the Director of the Advanced IT Research Center, Inha University, the UWB Wireless Communications Research Center, and the Key Government IT Research Center, South Korea. Since 1994, he had been serving as a member of Board of Directors. From 2000 to 2002, he had been the Vice President for the Korean Institute of Communication Sciences (KICS). He has been the KISC's President in 2006. In 1993, he received Engineering College Young Investigator Achievement Award from Inha University, and a Distinguished Service Medal from the Institute of Electronics Engineers of Korea. In 1996 and 1999, he received Distinguished Service Medals from the KICS. He was a recipient of the Inha University Engineering Paper Award and the LG Paper Award in 1998, and Motorola Paper Award in 2000. His research interests include multiple access communication systems, mobile communication systems, UWB radio systems and ad-hoc networks, and high-performance wireless Internet. He is a member of IEICE, KICS, and KIEE.