www.ietdl.org

Published in IET Communications
Received on 19th February 2012
Revised on 20th August 2012
doi: 10.1049/iet-com.2012.0101

ISSN 1751-8628

Delay controlled proportional fair scheduling in Rayleigh fading wireless channel

J.-H. Jeon J.-T. Lim

Department of Electrical Engineering, KAIST 373-1 Guseong-dong, Yuseong-gu Daejeon 305-701, Korea
E-mail: jtlim@stcon.kaist.ac.kr

Abstract: In this study, the authors consider a delay control problem with proportional fair scheduling in time-varying wireless channels. For proportional fairness of transmission rates among users, a congestion control scheme is proposed for guaranteeing both throughput and delay of the wireless network. Specifically, a scheduling weight is introduced to control queue lengths without degrading the throughput of the network because the delay of users depends on their queue length. Using the asymptotic approximation technique, the throughput and delay performance of the proposed scheme are analysed and showed the multiuser diversity effect analytically in the Rayleigh-fading wireless channels. Finally, simulation results show that the proposed scheme improves delay without degrading throughput.

1 Introduction

Proportional fair scheduling (PFS) [1] is known as the best way to guarantee throughput based on each user's condition without degrading fairness among users. Although it is firstly proposed for wireline networks to guarantee the throughput among flows, which have different routing paths [1] by the network utility maximisation (NUM), it is still useful to guarantee the throughput of each mobile station (MS) in downlink transmission, which suffers different characteristics of wireless channels. Each MSs channel characteristic is reflected to the statistical history of previously transmitted packets with an implementation scheme of PFS in [2]. Generally, although wireless channel capacity is degraded because of the fading characteristics of wireless channels, there are more chances to transmit packets in the high signal-to-noise ratio (SNR) condition when the number of MSs is growing. Specifically, in a flat-fading environment with sufficiently large coherence time, the throughput of PFS in time-varying wireless channels is improved by exploiting multiuser diversity [2]. The requirement for exploiting multiuser diversity is an availability of a channel state information (CSI) of all MSs in a base station (BS) without distortion.

In a view of NUM, the proportional fairness among users is achieved with the distributed implemented algorithm. Only the required information in a source is the sum of the traversed links implicit cost value [1]. Also, the required information in the link is the sum of the injection rates and the link capacity. However, the result of [1] does not consider the scheduling part of the NUM framework, which is more important because of the time-varying characteristics of the wireless capacity. The scheduling in time-varying wireless channels for delay performance is represented by throughput-optimal scheduling [3–5], which is a subproblem of NUM. If an input rate vector is inside the capacity region [3, 6] of wireless channels, the scheduling to maximise a queue length weighted sum of the rates is stabilising the queue length of each user. Tassiulas and Ephremides [3] have shown that the capacity region of the throughput-optimal scheduling with on–off channels is improved while stabilising queue lengths. By controlling the input arrival rate of each user for NUM in downlink transmission, Eryilmaz and Srikant [4] have shown that the throughput-optimal scheduling results in the same performance as PFS. Also, they have claimed that the dual subproblem of NUM is implicitly solved with the queue update equation and analyse the system performance both in the continuous-fluid model and the original model. Similar results for multihop networks are shown in [5]. Recently, Weeraddana *et al.* [7] present the resource allocation algorithm for multicommodity multichannel wireless networks. Gatsis *et al.* [8] consider the optimal wireless network design in the presence of fading and develop the algorithm to solve the physical layer resource allocation problem.

In addition to the throughput performance of PFS in time-varying wireless channels, the delay performance is more important for quality of service of the wireless networks. However, since the queue length is used for an indicator of network congestion and scheduling urgency, the delay cannot be easily controlled. Moreover, the tradeoff exists between the proportional fairness and the queue length. Although the results using a virtual queue are proposed in [9, 10] and the result using a shadow queue is proposed in [11], the throughput is more degraded while controlling delay. Recently, Huang and Neely [12] have presented the improved delay reduction result, which is comparable with the result of Neely [9]. Also, Khodaian and Khalaj [13] have presented the optimal random access control scheme in the delay constrained NUM problem and Jaramillo *et al.* [14] consider the max-weight scheduling problem in time-varying wireless channel with heterogeneous delay constraints.

In this paper, we propose a scheme to achieve controlled delay for PFS in the Rayleigh fading wireless channels and analyse both the throughput and delay performance. To control the delay of each user, we propose a congestion control scheme for guaranteeing both the fair throughput and the controlled delay. Then, we design the scheduling weight of the throughput-optimal scheduling in [4] so that the parameter dependency on queue length in [4] is eliminated. Moreover, we analyse the average performance using the asymptotic approximation [15] and show the multiuser diversity effect analytically. With the obtained closed form of throughput and delay, we easily evaluate the performance of both the proposed scheme and the conventional PFS. Finally, we show that the proposed scheme controls delay without degrading the network throughput compared with the previous results through the simulation.

The paper is organised as follows: in Section 2, we introduce the system description. In Section 3, we review the conventional PFS and propose a new congestion control scheme with the scheduling weight. In Section 4, we analyse the performance of the proposed scheme by using asymptotic approximation and show the asymptotic stability of the steady state. In Section 5, we compare our proposed scheme with the existing results. Finally, we present the conclusion in Section 6.

2 System description

We consider a downlink transmission from a BS to MSs through time-division multiple access (TDMA) with slotted time $n = 0, 1, 2, \ldots$. We assume that only one user is associated with one MS. Also, we only consider the elastic traffic as shown in [1, 4], whose rate can be controlled according to the network congestion condition. We assume that every MSs CSI is available in the BS at the beginning of each time slot without distortion because we do not consider the feedback reduction scheme in the multiuser diversity system [16]. Thus, the SNR values of all MSs are assumed to be fixed within a time slot and are available at the BS. Although the relation between wireless channel capacity, C, and SNR is described from Shannon's capacity theorem such as $C = W\log_2(1 + \mathrm{SNR})$ generally, we assume that the channel capacity is a linear function of SNR throughout this paper as shown in [17–19]. Note that the Shannon's capacity is approximated to a linear function of SNR at small values of SNR. Thus, we adopt the channel capacity equation in [17, 18] as $C = W' \cdot \mathrm{SNR}$ with $W' = W/2q\log_2 10$, where W is the bandwidth of the network, and q is the required bit-error-rate (BER) index, that is, $q = 4$ if the required BER is 10^{-4}. Finally, the symbols used in this paper are summarised in Table 1.

Table 1 List of symbols used

Symbol	Description
$q_i(n)$	queue length of user i at time n
$w_i(n)$	weight of user i for throughput-optimal scheduling at time n
$x_i(n)$	input arrival rate of user i at time n
$I_i(n)$	scheduling indicator of user i at time n
$C_i(n)$	channel capacity of user i at time n
$A_i(n)$	controlled input arrival process of elastic traffic during $[n, n+1)$
N	number of users
K	parameter for adjusting tradeoff between PF and delay
β_i	desired delay of user i
θ_i, γ_i, α_i	control parameters
ε_i	minimum value of $w_i(n)$
p_i	optimisation variable for throughput-optimal scheduling
ζ	channel capacity variable
$f_i(\zeta)$	pdf of user is channel capacity
λ_i	parameter of exponential distribution for channel capacity
$\bar{q}_i(n)$, $\bar{w}_i(n)$, $\bar{x}_i(n)$	average state variables of $q_i(n)$, $w_i(n)$, $x_i(n)$
$\bar{q}_{is}$, $\bar{w}_{is}$, $\bar{x}_{is}$	average steady states of $\bar{q}_i(n)$, $\bar{w}_i(n)$, $\bar{x}_i(n)$
y_i	defined as $\lambda_i/\bar{w}_{is}$
ρ_i	average throughput of user i
$\bar{d}_i$	average delay of user i

3 Delay-controlled PFS scheme with scheduling weight

In order to control the delay with PFS, we adopt the structure in [4]. Since both the throughput-optimal scheduling scheme and the congestion control scheme in [4] use the queue length as the implicit cost, it is not easy to control the each user's delay. We first review the queue length-based scheme in [4], then propose the delay controlled PFS scheme, which eliminates limitation of the delay control in [4].

3.1 Review of conventional PFS scheme

At the BS, the queue length of user i, q_i, is updated as follows

$$q_i(n + 1) = [q_i(n) + A_i(n) - C_i(n)I_i(n)]^+ \tag{1}$$

where $A_i(n)$ is the controlled input arrival process of the elastic traffic during $[n, n + 1)$, $C_i(n)$ is the channel capacity of user i at time n and $I_i(n)$ is the scheduling indicator at time n. Note that $[\cdot]^+ = \max(\cdot, 0)$. The result in [4] achieves PFS using both the queue length-based throughput-optimal scheduling for $I_i(n)$ and the queue length-based congestion controller for $A_i(n)$. First, the throughput-optimal scheduling is described with optimisation variables, $\vec{p} = [p_1, \ldots, p_N]$ where $p_i \in [0,1]$, at the BS in the TDMA environment as follows

$$\begin{aligned} \vec{I}(n) &= \arg\max_{\vec{p}\geq 0} \sum_{i=1}^{N} w_i(n)C_i(n)p_i \\ \text{s.t.} &\quad \sum_{i=1}^{N} p_i \leq 1 \end{aligned} \tag{2}$$

where $w_i(n)$ is the weight of user i for the throughput-optimal scheduling at time n, N is the number of users and $\vec{I}(n) = [I_1(n), \ldots, I_N(n)]$. The conventional throughput-optimal scheduling for proportional fairness [4, 5] uses the queue length as the weight directly, that is, $w_i(n) = q_i(n)$. From the scheduling in (2), we know that a user with the highest value of $w_i(n)C_i(n)$ is always selected to transmit at time n, that is, $I_i(n) = 1$, if $i = i^*$, and $I_i(n) = 0$, if $i \neq i^*$ where $i^* = \arg\max_i w_i(n)C_i(n)$.

The queue length-based congestion controller is designed with controlling the average arrival rate $E[A_i(n)|q_i(n)]$ as $K/q_i(n)$. Then, with the throughput-optimal scheduling, the queues of all users are stabilised while achieving proportional fairness. Note that the parameter K is used for adjusting tradeoff between the proportional fairness and the delay performance. If we want the better proportional fairness, that is, larger K, each user suffers from the larger amount of delay from the authors [4, 5]. However, because the delay is affected by the channel characteristic as shown in (1) and (2), the delay of each user is hard to control to the desired value by adjusting K.

3.2 Proposed scheme with scheduling weight

We newly design the weight of the throughput-optimal scheduling, $w_i(n)$, without using the queue length directly. Note that since the queue length of both the throughput-optimal scheduling and the congestion controller in [4] come from the same cost of the optimisation problem, the optimal input arrival rate for proportional fairness now changes to $K/w_i(n)$. Then, using Little's formula [20] and the optimal arrival rate, $K/w_i(n)$, we propose a new congestion control scheme to control the delay of each user. Moreover, we present a weight update scheme using the optimal arrival rate.

First, we design the congestion control scheme to satisfy both the delay of each user and the input arrival rate as follows

$$x_i(n+1) = \left[x_i(n) + \theta_i\left\{q_i(n) - \beta_i\frac{K}{w_i(n)}\right\} + \gamma_i\left\{x_i(n) - \frac{K}{w_i(n)}\right\}\right]^+ \tag{3}$$

where $x_i(n)$ is an input arrival rate of user i at time n, β_i is a desired delay of user i, and γ_i and θ_i are control parameters. Note that $x_i(n)$ is the same as $A_i(n)$. The second term of the right-hand side (RHS) of (3) is affected by the queue length of each user. Since the optimal input arrival rate for proportional fairness is described by $K/w_i(n)$, we set the desired queue length as $\beta_i(K/w_i(n))$ from Little's formula [20]. Then, if the queue length is larger than the desired queue length, $\beta_i(K/w_i(n))$, the input arrival rate should be decreased, that is, θ_i should be negative. Moreover, the third term of the RHS of (3) controls the input arrival rate directly. If the arrival rate is larger than the optimal input arrival rate, $K/w_i(n)$, we lower the input arrival rate,that is, γ_i should be also negative. Secondly, since we do not use directly the queue length as the weight of the throughput-optimal scheduling any more, we design the weight, $w_i(n)$, as follows

$$w_i(n+1) = \max\left[\varepsilon_i, w_i(n) + \alpha_i\left\{x_i(n) - \frac{K}{w_i(n)}\right\}\right] \tag{4}$$

where α_i is a control parameter and ε_i is a very small positive parameter to prevent the term, $K/w_i(n)$, from going to infinity. Note that the parameter, α_i, should be negative because the optimal input arrival rate, $K/w_i(n)$, permanently increases when the weight decreases with positive α_i.

With the congestion control of (3) and the proposed weight update of (4), we adjust the queue length for the delay, the scheduling weight for the throughput-optimal scheduling, and the input arrival rate for proportional fairness. Then, we expect that the state variables such as $q_i(n)$, $x_i(n)$ and $w_i(n)$ from (1) to (4) go to the desired values from the stability analysis in [21]. We will confirm that the average throughput of the proposed scheme is the same as the throughput of the conventional PFS [4] through performance analysis. Also, we will confirm that the average delay is controlled to the desired value.

4 Performance analysis and asymptotic stability

Since only one user has a chance to transmit packets at time n from (2), it is hard to determine the steady state of (2)–(4). Thus, from the asymptotic approximation [15], we analyse the performance of the proposed scheme in the average sense. Since the scheduling indicator, $I_i(n)$, is determined according to the channel capacity, $C_i(n)$, and the state variables such as $q_i(n)$, $w_i(n)$, $x_i(n)$, we obtain $E[C_i(n)I_i(n)|\vec{q}(n), \vec{w}(n), \vec{x}(n)]$ using the probability density function (pdf) of the channel capacity, $C_i(n)$ where $\vec{q}(n) = [q_1(n), \ldots, q_N(n)]$, $\vec{w}(n) = [w_1(n), \ldots, w_N(n)]$ and $\vec{x}(n) = [x_1(n), \ldots, x_N(n)]$.

Consider two users indexed by i and j. If the channel capacity of user i is ζ and the scheduling weight vector is $\vec{w}(n)$, the conditional probability of the total weight of user i, which is a product of $w_i(n)$ and $C_i(n)$, larger than the total weight of user j is described as follows

$$P[w_i(n)C_i(n) \geq w_j(n)C_j(n)|C_i(n) = \zeta, \vec{w}(n)] = P\left[C_j(n) \leq \frac{w_i(n)}{w_j(n)}\zeta \middle| \vec{w}(n)\right] \tag{5}$$

As the channel state at time n is independently distributed among users, the conditional probability of user i chosen is a product of probabilities of (5) for all $j \neq i$. Also, because $I_i(n)$ is only related to $\vec{w}(n)$ from (2), the conditional probability is only associated with $\vec{w}(n)$. Thus, the conditional probability of user i chosen is described as follows

$$\begin{aligned} P[I_i(n) = 1|C_i(n) = \zeta, \vec{q}(n), \vec{w}(n), \vec{x}(n)] &= P[I_i(n) = 1|C_i(n) = \zeta, \vec{w}(n)] \\ &= P[w_i(n)C_i(n) \geq w_j(n)C_j(n), \ \forall j \neq i|C_i(n) = \zeta, \vec{w}(n)] \\ &= \prod_{j \neq i}^{N} P[w_i(n)C_i(n) > w_j(n)C_j(n)|C_i(n) = \zeta, \vec{w}(n)] \end{aligned} \tag{6}$$

Then, we show the following expectation

$$\begin{aligned} &E[C_i(n)I_i(n)|\vec{q}(n), \vec{w}(n), \vec{x}(n)] \\ &= \int_0^\infty \zeta P[I_i(n) = 1|C_i(n) = \zeta, \vec{q}(n), \vec{w}(n), \vec{x}(n)]f_i(\zeta)\,d\zeta \\ &= \int_0^\infty \zeta \prod_{j \neq i}^{N} P[w_i(n)C_i(n) > w_j(n)C_j(n)|C_i(n) = \zeta, \vec{w}(n)]f_i(\zeta)\,d\zeta \end{aligned} \tag{7}$$

where $f_i(\zeta)$ is the pdf of user is channel capacity.

doi: 10.1049/iet-com.2012.0101

As we mentioned in Section 2, the channel capacity of user i is represented by the product of W' and the SNR of user i, that is, $C_i(n) = W' \cdot \mathrm{SNR}_i(n)$. Also, since we consider the Rayleigh-fading channels, the channel capacity of user i has an exponential distribution with parameter λ_i. Note that $E[C_i(n)] = W' \cdot E[\mathrm{SNR}_i(n)] = 1/\lambda_i$. Then, we describe (5) as

$$P\left[C_j(n) \le \frac{w_i(n)}{w_j(n)} \zeta \middle| \vec{\boldsymbol{w}}(n) \right] = \int_0^{((w_i(n))/(w_j(n)))\zeta} \lambda_j \exp(-\lambda_j \psi)\,\mathrm{d}\psi = 1 - \exp\left(-\lambda_j \frac{w_i(n)}{w_j(n)} \zeta\right) \tag{8}$$

Hence, we obtain the expectation from (7) as follows

$$\begin{aligned} &E[C_i(n) I_i(n) | \vec{\boldsymbol{q}}(n),\ \vec{\boldsymbol{w}}(n),\ \vec{\boldsymbol{x}}(n)] \\ &= \int_0^{\infty} \zeta \prod_{j \ne i}^{N} \left[1 - \exp\left(-\lambda_j \frac{w_i(n)}{w_j(n)} \zeta\right) \right] \lambda_i \exp(-\lambda_i \zeta)\,\mathrm{d}\zeta \\ &= \frac{\lambda_i}{(w_i(n))^2} \left[\frac{1}{(\lambda_i/w_i(n))^2} - \sum_{j \ne i}^{N} \frac{1}{(\lambda_j/(w_j(n)) + \lambda_i/(w_i(n)))^2} \right. \\ &\quad + \sum_{j \ne i}^{N} \sum_{l \ne i,j}^{N} \frac{1}{(\lambda_j/(w_j(n)) + \lambda_l/(w_l(n)) + \lambda_i/(w_i(n)))^2} \\ &\quad \left. + \cdots + \frac{(-1)^{N+1}}{(\lambda_1/(w_1(n)) + \cdots + \lambda_N/(w_N(n)))^2} \right] \\ &:= h_i(w_1(n),\ \ldots,\ w_N(n)) \end{aligned} \tag{9}$$

Note that the detailed derivation of (9) is shown in the appendix. Finally, we obtain the average approximated dynamics using the asymptotic approximation in [15] as follows

$$\bar{q}_i(n+1) = [\bar{q}_i(n) + \bar{x}_i(n) - h_i(\bar{w}_1(n),\ \ldots,\ \bar{w}_N(n))]^+ \tag{10}$$

$$\bar{x}_i(n+1) = \left[\bar{x}_i(n) + \gamma_i \left\{ \bar{x}_i(n) - \frac{K}{\bar{w}_i(n)} \right\} + \theta_i \left\{ \bar{q}_i(n) - \beta_i \frac{K}{\bar{w}_i(n)} \right\} \right]^+ \tag{11}$$

$$\bar{w}_i(n+1) = \max\left[\varepsilon_i, \bar{w}_i(n) + \alpha_i \left\{ \bar{x}_i(n) - \frac{K}{\bar{w}_i(n)} \right\} \right] \tag{12}$$

where $\bar{q}_i(n)$, $\bar{w}_i(n)$ and $\bar{x}_i(n)$ are the average state variables of $q_i(n)$, $w_i(n)$ and $x_i(n)$, respectively. Hence, we obtain the steady state of the average state variables from (10) to (11) with ignoring saturation non-linearity as follows

$$\bar{q}_{\mathrm{is}} = \beta_i \frac{K}{\bar{w}_{\mathrm{is}}}, \quad \bar{x}_{\mathrm{is}} = \frac{K}{\bar{w}_{\mathrm{is}}} \tag{13}$$

$$\frac{K}{\bar{w}_{\mathrm{is}}} = h_i(\bar{w}_{1\mathrm{s}}, \ldots, \bar{w}_{N\mathrm{s}}) \tag{14}$$

If we obtain $\bar{w}_{\mathrm{is}}$ from (14), we obtain the average steady states of the queue length and the input arrival rate from (13). Define $y_i := \lambda_i/\bar{w}_{\mathrm{is}}$. Then, we obtain N number of non-linear equations as follows (see (15))

Note that we know that y_i is positive because $\bar{w}_i$ is lower bounded by the very small positive parameter ε_i from (12). Using the symbolic toolbox of MATLAB, we obtain an unique solution of (15) as $y_1 = y_2 = \ldots = y_N$, when $y_i > 0$ for all i. Then, from (15), we obtain (see (16))

where

$$\binom{N-1}{j} = \frac{(N-1)!}{j!(N-1-j)!}$$

Finally, we obtain the average steady state, $\bar{w}_{\mathrm{is}}$, as follows

$$\bar{w}_{\mathrm{is}} = K\lambda_i \left[\sum_{j=1}^{N} (-1)^{j+1} \frac{\binom{N-1}{j-1}}{j^2} \right]^{-1} \tag{17}$$

$$\begin{aligned} K &= y_1 \left[\frac{1}{y_1^2} - \sum_{j \ne 1}^{N} \frac{1}{(y_1 + y_j)^2} + \sum_{j \ne 1}^{N} \sum_{l \ne 1,j}^{N} \frac{1}{(y_1 + y_j + y_l)^2} + \cdots + \frac{(-1)^{N+1}}{(y_1 + \cdots + y_N)^2} \right] \\ &\vdots \\ K &= y_N \left[\frac{1}{y_N^2} - \sum_{j \ne N}^{N} \frac{1}{(y_N + y_j)^2} + \sum_{j \ne N}^{N} \sum_{l \ne N,j}^{N} \frac{1}{(y_N + y_j + y_l)^2} + \cdots + \frac{(-1)^{N+1}}{(y_1 + \cdots + y_N)^2} \right] \end{aligned} \tag{15}$$

$$\begin{aligned} K &= \frac{1}{y_i} \left[1 - \sum_{j \ne i}^{N} \frac{1}{2^2} + \sum_{j \ne i}^{N} \sum_{l \ne i,j}^{N} \frac{1}{3^2} + \cdots + (-1)^{N+1} \frac{1}{N^2} \right] \\ &= \frac{1}{y_i} \left[\frac{\binom{N-1}{0}}{1^2} - \frac{\binom{N-1}{1}}{2^2} + \frac{\binom{N-1}{2}}{3^2} + \cdots + (-1)^{N+1} \frac{\binom{N-1}{N-1}}{N^2} \right] = \frac{1}{y_i} \sum_{j=1}^{N} (-1)^{j+1} \frac{\binom{N-1}{j-1}}{j^2} \end{aligned} \tag{16}$$

4.1 Proportional fairness with exploiting multiuser diversity gain

If we set the weight of the throughput-optimal scheduling as the queue length for the proportional fairness shown in [4], we obtain the average steady state of the queue length from (1) to (2) as follows

$$\frac{K}{\bar{q}_{is}} = h_i(\bar{q}_{1s}, \ldots, \bar{q}_{Ns}) \tag{18}$$

We know that the solution of (18) is completely the same as the solution of (14), then the proportional fairness of the proposed scheme is guaranteed. Moreover, we obtain the average throughput of user i, ρ_i, from (17) as follows

$$\rho_i = \frac{K}{\bar{w}_{is}} = h_i(\bar{w}_{1s}, \ldots, \bar{w}_{Ns}) = \frac{1}{\lambda_i}\sum_{j=1}^{N}(-1)^{j+1}\frac{\binom{N-1}{j-1}}{j^2} \tag{19}$$

The average throughput of user i is composed of the average capacity with time fraction, $1/\lambda_i$, and the gain

$$\sum_{j=1}^{N}(-1)^{j+1}\frac{\binom{N-1}{j-1}}{j^2}$$

When the number of users increases, the throughput of user i decreases because the gain decreases. From each user's throughput in (19), the total network throughput, ρ, is obtained as follows

$$\rho = \sum_{i=1}^{N}\rho_i = \left(\sum_{i=1}^{N}\frac{1}{\lambda_i}\right)\left[\sum_{j=1}^{N}(-1)^{j+1}\frac{\binom{N-1}{j-1}}{j^2}\right] \tag{20}$$

Note that the total network throughput is improved compared with the throughput of the conventional TDMA scheduling without multiuser diversity [18], that is, $(1/N)\sum_{i=1}^{N} 1/\lambda_i$.

4.2 Delay control

From Little's formula [20], the average delay is described as the queue length divided by the input arrival rate. Thus, from (13), the average delay of user i, $\bar{d}_i$, is obtained as follows

$$\bar{d}_i = \frac{\bar{q}_{is}}{\bar{x}_{is}} = \beta_i \tag{21}$$

Note that we can adjust the delay parameter, β_i, according to the delay requirement of user i. Since the delay from (21) is not a function of the number of users, the delay is preserved when the number of users increases. Also, both the congestion control of (3) with the weight update of (4) do not require the information about the number of users.

4.3 Asymptotic stability

Now, we investigate the asymptotic stability at the steady state. Let $\boldsymbol{Z}(n)$ be the state vector represented by

$$\boldsymbol{Z}(n) := [\bar{q}_1(n) - \bar{q}_{1s}\ \bar{x}_1(n) - \bar{x}_{1s},\ \ldots,\ \bar{q}_N(n) - \bar{q}_{Ns}\ \bar{x}_N(n) - \bar{x}_{Ns}\bar{w}_1(n) - \bar{w}_{1s},\ \ldots,\ \bar{w}_N(n) - \bar{w}_{Ns}]^{\mathrm{T}} \tag{22}$$

Linearisation of the system around its steady state results in the following linear system

$$\boldsymbol{Z}(n+1) = \boldsymbol{A}\boldsymbol{Z}(n) \tag{23}$$

with

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \cdots & 0 & \boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{A}_N & \boldsymbol{B}_{N1} & \cdots & \boldsymbol{B}_{NN} \\ \boldsymbol{C}_1 & \cdots & 0 & \boldsymbol{D}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{C}_N & 0 & \cdots & \boldsymbol{D}_N \end{bmatrix} \tag{24}$$

where

$$\boldsymbol{A}_i = \begin{bmatrix} 1 & 1 \\ \theta_i & 1+\gamma_i \end{bmatrix}$$

$$\boldsymbol{C}_i = \begin{bmatrix} 0 & \alpha_i \end{bmatrix}, \quad \boldsymbol{D}_i = \begin{bmatrix} 1 + \dfrac{\alpha_i K}{\bar{w}_{is}^2} \end{bmatrix}$$

$$\boldsymbol{B}_{ij} = \begin{cases} \begin{bmatrix} -\dfrac{\partial}{\partial \bar{w}_j} h_i(\bar{w}_1, \ldots, \bar{w}_N)\Big|_{\vec{w}=\vec{w}_s} \\ (\gamma_i + \theta_i\beta_i)\dfrac{K}{\bar{w}_{is}^2} \end{bmatrix}, & \text{if } i = j \\ \begin{bmatrix} -\dfrac{\partial}{\partial \bar{w}_j} h_i(\bar{w}_1, \ldots, \bar{w}_N)\Big|_{\vec{w}=\vec{w}_s} \\ 0 \end{bmatrix}, & \text{if } i \neq j \end{cases} \tag{25}$$

Note that the detailed derivations of (23)–(25) are shown in the appendix. If we choose $-1 < \alpha_i < 0$, where $|\alpha_i| \ll 1$ in $\boldsymbol{C}_i$, we only have to consider $\boldsymbol{A}_i$ and $\boldsymbol{D}_i$ for stability. Then, to have the stable eigenvalues of $\boldsymbol{A}_i$ and $\boldsymbol{D}_i$, we show the conditions of θ_i, and γ_i by Jury's criterion [21] as follows

$$\theta_i < 0, \quad \gamma_i < 0, \quad \gamma_i > \frac{\theta_i - 4}{2}, \quad -2 < \gamma_i - \theta_i < 0 \tag{26}$$

Then, the steady state is asymptotically stable with the properly chosen parameters such as α_i, γ_i and θ_i that satisfy (26).

5 Simulation results

In this section, we confirm the proposed model to describe the multiuser diversity system and compare the delay performance with the results of Eryilmaz and Srikant [4] and Bui *et al.* [11]. We consider the downlink system with a BS and associated MSs. Note that each MS has only one

doi: 10.1049/iet-com.2012.0101

Table 2 Average throughput and average delay

	User 1	User 2	User 3	User 4
average SNR, dB	−11	−9	−6	−4
throughput of analysis, kbps	1.5567	2.4672	4.9228	7.8022
throughput of simulation, kbps	1.5604	2.4537	4.9525	7.7847
delay requirement, β, ms	500	200	400	300
delay of simulation, ms	508.9076	199.6197	399.0601	300.7367

user with elastic traffic source. We assume that every MSs CSI is available in the BS at the beginning of each time slot. The SNR value is assumed to be fixed within a time slot. However, at each time slot, the SNR value is time-varying with an exponential distribution with the Rayleigh fading channels [18]. We set that the channel bandwidth is 1 MHz and the BER requirement is 10^{-4}. We obtain the result of performance evaluation from our developed discrete-time based MATLAB (Ver. R2010a) simulator. In the simulator, all sources randomly generate packets with Poisson process at each discrete-time, then the packets are injected into each user's transmission queue. The packets are transmitted to each MS without any distortion because we assume that the MSs CSI is available at the BS. Note that since we do not consider the transmission outage or failure in this paper, the packets are always successfully transmitted in the simulation. For the comparison with the result in [4], our developed simulator assumes that the channels between the BS and each of the users fade independently and the BS is only allowed to serve a single queue in a given slot. Finally, we measure the average throughput as the amounts of transmitted bits during a simulation time and the average delay as the mean value of waiting time of each packet until transmission is completed.

First, we confirm that the proposed scheme guarantees a controlled delay with the proportional fairness. We set that there are four users with different characteristics in the network in Table 2. Also, we set the control parameters to satisfy (26) as $\alpha_i = -0.02$, $\gamma_i = -0.2$ and $\theta_i = -0.001$ for all users. As shown in Figs. 1–3, the average values of state variables named by the queue length, the scheduling weight, and the input arrival rate are converging to the desired steady state. With the analysis of the proposed scheme, we obtain the average throughput analytically in Table 2, and show that their values are close to the average throughput of simulation. Also, the average delay of simulation is positioned around the delay requirement.

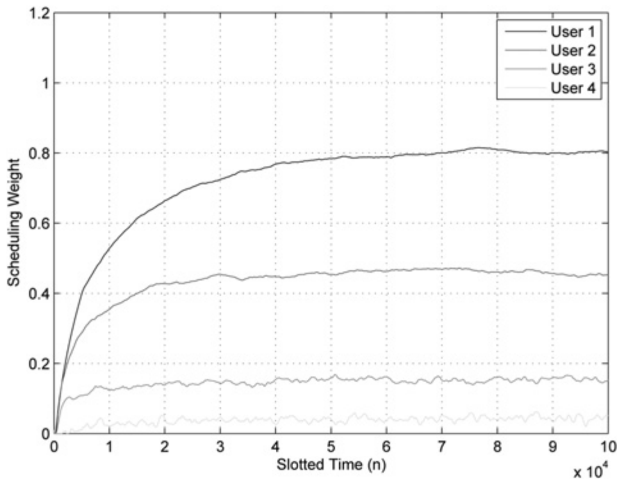

Fig. 2 *Trajectory of scheduling weight*

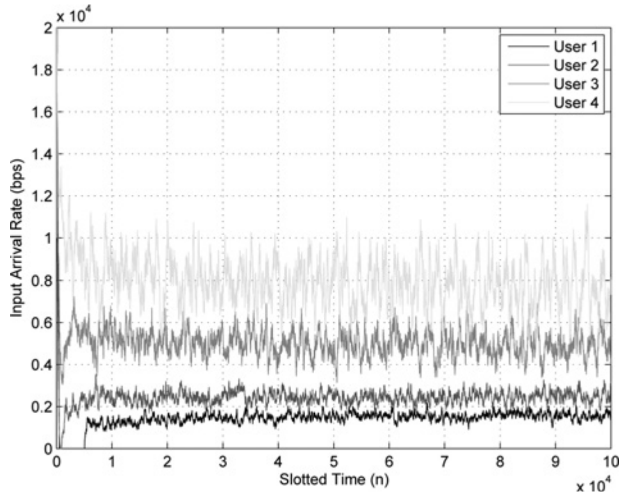

Fig. 3 *Trajectory of input arrival rate*

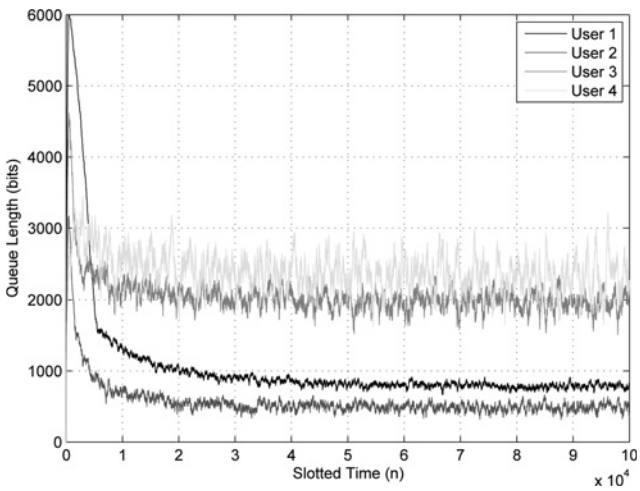

Fig. 1 *Trajectory of queue length*

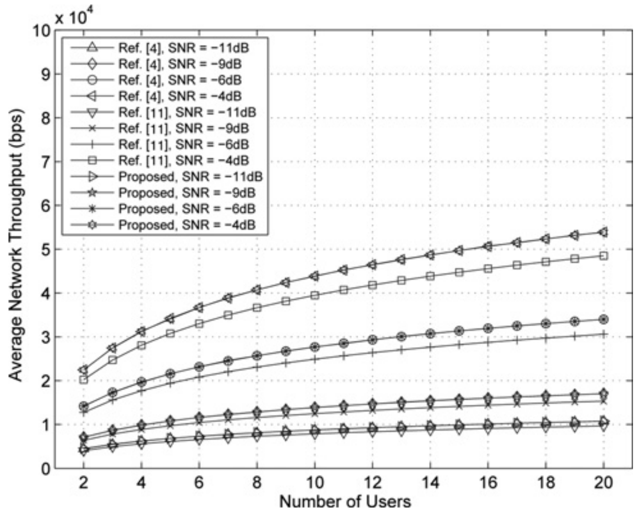

Fig. 4 *Comparison of network throughput with various SNRs*

In addition to the confirmation of analytical results, we compare the throughput and delay performance to the results of Eryilmaz and Srikant [4] and Bui *et al.* [11] by varying the number of users in the network. In order to show the multiuser diversity gain because of the number of users effectively, we set that the average SNR values of all users are same. Also, we set β_i as 200 ms and K as 1500 for all users. As shown in Fig. 4, the average network throughput increases when the number of users increases and the throughput of the proposed scheme is indistinguishable from the result of Eryilmaz and Srikant [4] for the average SNR values of −11, −9, −6 and −4 dB. This result confirms that the multiuser diversity [2] increases the network throughput and the higher average SNR value results in the better network throughput. Note that since the result of Bui *et al* [11] reduces the input arrival rate with a shadow queue for delay reduction, the throughput performance is degraded. Moreover, we show the average queue length of each user in the network in Figs. 5–7. Since there is no delay control scheme for each user, the average queue length with the previous scheme [4] increases. Compared with the previous result, the average queue length with the proposed scheme shows the decreasing behaviour when the number of users increases. The reason why the average queue length of each user decreases is that the congestion controller adjust the queue length when each user's throughput decreases as the number of users increases. However, when the number of users increases, the average network throughput increases as shown in Fig. 4. Moreover, although the queue length is reduced with the result of Bui *et al* [11] as shown in Fig. 6, it is not easy to control the queue length to compensate the throughput degradation. Thus, we show that the proposed scheme controls the delay more easily only adjusting β_i.

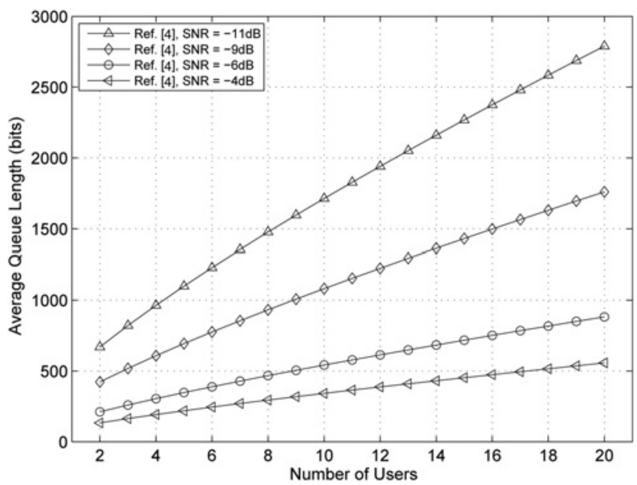

Fig. 5 *Comparison of queue length with various SNRs in [4]*

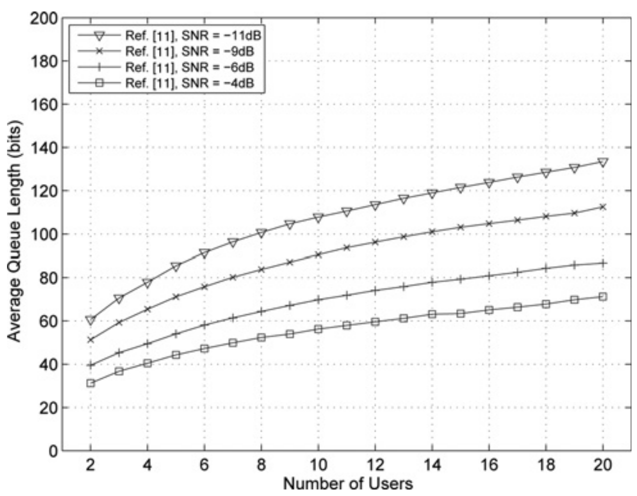

Fig. 6 *Comparison of queue length with various SNRs in [11]*

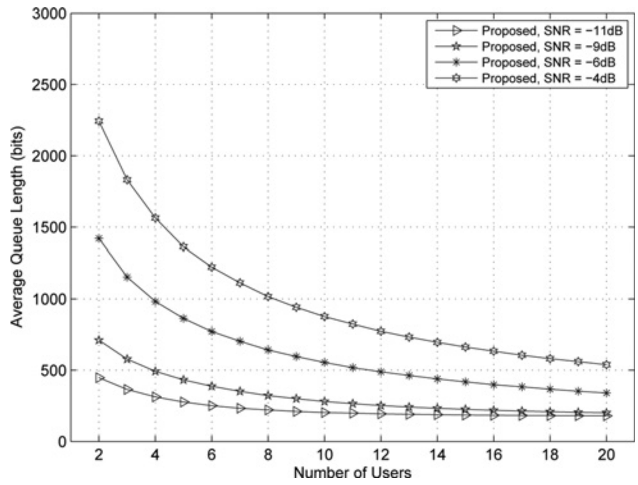

Fig. 7 *Comparison of queue length with various SNRs in the proposed scheme*

6 Conclusion

We propose the congestion control scheme with the scheduling weight, which controls the delay with proportional fairness of the network. From the pdf of channel capacity, we obtain the average performance of throughput and delay and show the multiuser diversity gain analytically in the Rayleigh-fading wireless channels. Finally, we show that the delay of the proposed scheme is controlled with the delay parameter without degrading the throughput compared to the existing results.

7 Acknowledgment

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0004861).

8 References

1 Kelly, F.P., Maulloo, A., Tan, D.: 'Rate control for communication networks: shadow prices, proportional fairness and stability', *J. Oper. Res. Soc.*, 1998, **49**, (3), pp. 237–252
2 Viswanath, P., Tse, N.C., Laroia, R.L.: 'Opportunistic beamforming using dumb antennas', *IEEE Trans. Inf. Theory*, 2002, **48**, (6), pp. 1277–1294
3 Tassiulas, L., Ephremides, A.: 'Dynamic server allocation to parallel queues with randomly varying connectivity', *IEEE Trans. Inf. Theory*, 1993, **39**, (2), pp. 466–478
4 Eryilmaz, A., Srikant, R.: 'Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control', *IEEE/ACM Trans. Netw.*, 2007, **15**, (6), pp. 1333–1344
5 Neely, M.J., Modiano, E., Li, C.-P.: 'Fairness and optimal stochastic control for heterogeneous networks', *IEEE/ACM Trans. Netw.*, 2008, **16**, (2), pp. 396–409

6 Toumpis, S., Goldsmith, A.J.: 'Capacity regions for wireless ad hoc networks', *IEEE Trans. Wirel. Commun.*, 2003, **2**, (4), pp. 736–748
7 Weeraddana, P.C., Codreanu, M., Latva-aho, M., Ephremides, A.: 'Resource allocation for cross-layer utility maximization in wireless networks', *IEEE Trans. Veh. Technol.*, 2011, **60**, (6), pp. 2790–2809
8 Gatsis, N., Ribeiro, A., Giannakis, G.B.: 'A class of convergent algorithms for resource allocation in wireless fading networks', *IEEE Trans. Wirel. Commun.*, 2010, **9**, (5), pp. 1808–1823
9 Neely, M.J.: 'Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks', *IEEE J. Sel. Areas Commun.*, 2006, **24**, (8), pp. 1489–1501
10 Tahir, M., Mazumder, S.K.: 'Delay constrained optimal resource utilization of wireless networks for distributed control systems', *IEEE Commun. Lett.*, 2008, **12**, (4), pp. 289–291
11 Bui, L., Srikant, R., Stolyar, A.: 'Novel architectures and algorithms for delay reduction in back-pressure scheduling and routing'. Proc. IEEE INFOCOM, April 2009, pp. 2936–2940
12 Huang, L., Neely, M.J.: 'Delay reduction via Lagrange multipliers in stochastic network optimization', *IEEE Trans. Autom. Control*, 2011, **56**, (4), pp. 842–857
13 Khodaian, A.M., Khalaj, B.H.: 'Delay-constrained utility maximization in multi-hop random access networks', *IET Commun.*, 2010, **4**, (16), pp. 1908–1918
14 Jaramillo, J.J., Srikant, R., Ying, L.: 'Scheduling for optimal rate allocation in ad hoc networks with heterogeneous delay constraints', *IEEE J. Sel. Areas Commun.*, 2011, **29**, (5), pp. 979–987
15 Lee, H.-J., Lim, J.-T.: 'Cross-layer congestion control for power efficiency over wireless multihop networks', *IEEE Trans. Veh. Technol.*, 2009, **58**, (9), pp. 5274–5278
16 Tang, T., Heath, Jr. R.W.: 'Opportunistic feedback for downlink multiuser diversity', *IEEE Commun. Lett.*, 2005, **9**, (10), pp. 948–950
17 Cruz, R., Santhanam, A.: 'Optimal routing, link scheduling, and power control in multi-hop wireless networks'. Proc. IEEE INFOCOM, San Francisco, USA, April 2003, pp. 702–711
18 Rappaport, T.S.: 'Wireless communications: principles and practice' (Prentice-Hall, 2002)
19 Lin, X., Shroff, N.B.: 'The impact of imperfect scheduling on cross-layer congestion control in wireless networks', *IEEE/ACM Trans. Netw.*, 2006, **14**, (2), pp. 302–315
20 Bertsekas, D.P., Gallager, R.: 'Data networks' (Prentice-Hall, 1987)
21 Jury, E.I.: 'Theory and application of the z-transform method' (Krieger Publishing Co., 1973)

9 Appendix

9.1 Derivation of (9)

$E[C_i(n)I_i(n)|\vec{\boldsymbol{q}}(n), \vec{\boldsymbol{w}}(n), \vec{\boldsymbol{x}}(n)]$ is described from (5) to (8) as follows

$$
\begin{aligned}
&E[C_i(n)I_i(n)|\vec{\boldsymbol{q}}(n), \vec{\boldsymbol{w}}(n), \vec{\boldsymbol{x}}(n)] \\
&= \int_0^\infty \zeta \prod_{j\neq i}^{N}\left[1-\exp\left(-\lambda_j \frac{w_i(n)}{w_j(n)}\zeta\right)\right]\lambda_i \exp(-\lambda_i\zeta)\,\mathrm{d}\zeta
\end{aligned} \tag{27}
$$

Note that the pdf of user is channel capacity, $f_i(\zeta)$, is $\lambda_i \exp(-\lambda_i\zeta)$. Also from the expectation, we know that the following formula is satisfied

$$
\int_0^\infty \zeta\lambda_i \exp(-\lambda_i\zeta)\,\mathrm{d}\zeta = \frac{1}{\lambda_i} \Leftrightarrow \int_0^\infty \zeta \exp(-\lambda_i\zeta)\,\mathrm{d}\zeta = \frac{1}{\lambda_i^2} \tag{28}
$$

In order to using the formula, we represent $\prod_{j\neq i}^{N}[1-\exp(-\lambda_j(w_i(n)/w_j(n))\zeta)]\lambda_i \exp(-\lambda_i\zeta)$ from (27) as follows

$$
\begin{aligned}
&\prod_{j\neq i}^{N}\left[1-\exp\left(-\lambda_j \frac{w_i(n)}{w_j(n)}\zeta\right)\right]\lambda_i \exp(-\lambda_i\zeta) \\
&= \lambda_i \exp(-\lambda_i\zeta) - \sum_{j\neq i}^{N}\lambda_i \exp\left[-\left(\lambda_j\frac{w_i(n)}{w_j(n)}+\lambda_i\right)\zeta\right] \\
&\quad + \sum_{j\neq i}^{N}\sum_{l\neq i,j}^{N}\lambda_i \exp\left[-\left(\lambda_j\frac{w_i(n)}{w_j(n)}+\lambda_l\frac{w_i(n)}{w_l(n)}+\lambda_i\right)\zeta\right] \\
&\quad + \cdots + (-1)^{N+1}\lambda_i \exp\left[-\left(\sum_{j\neq i}^{N}\lambda_j\frac{w_i(n)}{w_j(n)}+\lambda_i\right)\zeta\right]
\end{aligned} \tag{29}
$$

From (28) to (29), we rewrite (27) as follows

$$
\begin{aligned}
&E[C_i(n)I_i(n)|\vec{q}(n), \vec{w}(n), \vec{x}(n)] \\
&= \int_0^\infty \zeta \prod_{j\neq i}^{N}\left[1-\exp\left(-\lambda_j \frac{w_i(n)}{w_j(n)}\zeta\right)\right]\lambda_i \exp(-\lambda_i\zeta)\,\mathrm{d}\zeta \\
&= \lambda_i\left[\frac{1}{\lambda_i^2} - \sum_{j\neq i}^{N}\frac{1}{(\lambda_j(w_i(n)/w_j(n))+\lambda_i)^2}.\right. \\
&\quad + \sum_{j\neq i}^{N}\sum_{l\neq i,;j}^{N}\frac{1}{(\lambda_j(w_i(n)/w_j(n))+\lambda_l(w_i(n)/w_l(n))+\lambda_i)^2} + \cdots \\
&\quad \left. + \frac{(-1)^{N+1}}{\left(\sum_{j\neq i}^{N}\lambda_j(w_i(n)/w_j(n))+\lambda_i\right)^2}\right]
\end{aligned} \tag{30}
$$

With multiplying $(w_i(n))^2$ to both numerator and denominator, we finally obtain following equation

$$
\begin{aligned}
&E[C_i(n)I_i(n)|\vec{q}(n), \vec{w}(n), \vec{x}(n)] \\
&= \frac{\lambda_i}{(w_i(n))^2}\left[\frac{1}{(\lambda_i/w_i(n))^2} - \sum_{j\neq i}^{N}\frac{1}{(\lambda_j/w_j(n)+\lambda_i/w_i(n))^2}.\right. \\
&\quad + \sum_{j\neq i}^{N}\sum_{l\neq i,j}^{N}\frac{1}{(\lambda_j/w_j(n)+\lambda_l/w_l(n)+\lambda_i/w_i(n))^2} + \cdots \\
&\quad \left. + \frac{(-1)^{N+1}}{(\lambda_1/w_1(n)+\cdots+\lambda_N/w_N(n))^2}\right] \\
&:= h_i(w_1(n), \ldots, w_N(n))
\end{aligned} \tag{31}
$$

9.2 Derivation of (23)–(25)

Define $\Delta\bar{q}_i(n) := \bar{q}_i(n) - \bar{q}_{\text{is}}$, $\Delta\bar{x}_i(n) := \bar{x}_i(n) - \bar{x}_{\text{is}}$ and $\Delta\bar{w}_i(n) := \bar{w}_i(n) - \bar{w}_{\text{is}}$. Then, the average dynamics denoted by $\Delta\bar{q}_i(n)$, $\Delta\bar{x}_i(n)$ and $\Delta\bar{w}_i(n)$ from (10) to (12) ignoring saturation non-linearity is described as follows

$$
\begin{aligned}
\Delta\bar{q}_i(n+1) &= \Delta\bar{q}_i(n) + \Delta\bar{x}_i(n) + \bar{x}_{\text{is}} \\
&\quad - h_i(\Delta\bar{w}_1(n)+\bar{w}_{1\text{s}}, \ldots, \Delta\bar{w}_N(n)+\bar{w}_{N\text{s}})
\end{aligned} \tag{32}
$$

$$\Delta\bar{x}_i(n+1) = \Delta\bar{x}_i(n) + \gamma_i\left\{\Delta\bar{x}_i(n) + \bar{x}_{\text{is}} - \frac{K}{\Delta\bar{w}_i(n) + \bar{w}_{\text{is}}}\right\} + \theta_i\left\{\Delta\bar{q}_i(n) + \bar{q}_{\text{is}} - \beta_i\frac{K}{\Delta\bar{w}_i(n) + \bar{w}_{\text{is}}}\right\} \quad (33)$$

$$\Delta\bar{w}_i(n+1) = \Delta\bar{w}_i(n) + \alpha_i\left\{\Delta\bar{x}_i(n) + \bar{x}_{\text{is}} - \frac{K}{\Delta\bar{w}_i(n) + \bar{w}_{\text{is}}}\right\} \quad (34)$$

Linearisation of (32)–(34) around its steady state, that is, $\Delta\bar{q}_i(n) = 0$, $\Delta\bar{w}_i(n) = 0$ and $\Delta\bar{x}_i(n) = 0$, results in the following linear system

$$\Delta\bar{q}_i(n+1) = \Delta\bar{q}_i(n) + \Delta\bar{x}_i(n) - \sum_{j=1}^{N}\frac{\partial}{\partial\bar{w}_j}h_i(\bar{w}_1, \ldots, \bar{w}_N)\bigg|_{\vec{w}=\vec{w}_{\text{s}}}\Delta\bar{w}_j(n) \quad (35)$$

$$\begin{aligned}\Delta\bar{x}_i(n+1) &= \Delta\bar{x}_i(n) + \gamma_i\Delta\bar{x}_i(n) - \frac{\partial}{\partial\Delta\bar{w}_i}\left\{\frac{\gamma_i K}{\Delta\bar{w}_i + \bar{w}_{\text{is}}}\right\}\bigg|_{\Delta\vec{w}=0}\Delta\bar{w}_i(n) \\ &\quad + \theta_i\bar{q}_i(n) - \frac{\partial}{\partial\Delta\bar{w}_i}\left\{\theta_i\beta_i\frac{K}{\Delta\bar{w}_i + \bar{w}_{\text{is}}}\right\}\bigg|_{\Delta\vec{w}=0}\Delta\bar{w}_i(n) \\ &= \Delta\bar{x}_i(n) + \gamma_i\Delta\bar{x}_i(n) + \theta_i\Delta\bar{q}_i(n) \\ &\quad + \left\{(\gamma_i + \theta_i\beta_i)\frac{K}{\bar{w}_{\text{is}}^2}\right\}\Delta\bar{w}_i(n)\end{aligned} \quad (36)$$

$$\begin{aligned}\Delta\bar{w}_i(n+1) &= \Delta\bar{w}_i(n) + \alpha_i\Delta\bar{x}_i(n) - \frac{\partial}{\partial\Delta\bar{w}_i}\left\{\frac{\alpha_i K}{\Delta\bar{w}_i + \bar{w}_{\text{is}}}\right\}\bigg|_{\Delta\vec{w}=0}\Delta\bar{w}_i(n) \\ &= \Delta\bar{w}_i(n) + \alpha_i\Delta\bar{x}_i(n) + \left\{\frac{\alpha_i K}{\bar{w}_{\text{is}}^2}\right\}\Delta\bar{w}_i(n)\end{aligned} \quad (37)$$

Note that $(\partial/(\partial\Delta\bar{w}_j))h_i(\Delta\bar{w}_1 + \bar{w}_{1s}, \ldots, \Delta\bar{w}_N + \bar{w}_{N\text{s}})|_{\Delta\vec{w}=0} = (\partial/(\partial\bar{w}_j))h_i(\bar{w}_1, \ldots, \bar{w}_N)|_{\vec{w}=\vec{w}_{\text{s}}}$. Also, we represent $\boldsymbol{Z}(n)$ from (22) as follows

$$\begin{aligned}\boldsymbol{Z}(n) &:= [\bar{q}_1(n) - \bar{q}_{1\text{s}}\bar{x}_1(n) - \bar{x}_{1\text{s}}, \ldots, \bar{q}_N(n) - \bar{q}_{N\text{s}}\bar{x}_N(n) \\ &\quad - \bar{x}_{N\text{s}}\bar{w}_1(n) - \bar{w}_{1\text{s}}, \ldots, \bar{w}_N(n) - \bar{w}_{N\text{s}}]^{\text{T}} \\ &= [\Delta\bar{q}_1(n)\Delta\bar{x}_1(n), \ldots, \Delta\bar{q}_N(n)\Delta\bar{x}_N(n)\Delta\bar{w}_1(n), \ldots, \Delta\bar{w}_N(n)]^{\text{T}}\end{aligned} \quad (38)$$

Thus, we can describe the linearised system from (35) to (37) as the vector–matrix form as shown in (23)–(25).

IET Commun., 2012, Vol. 6, Iss. 17, pp. 2816–2824
doi: 10.1049/iet-com.2012.0101