This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2008 proceedings.

Wireless medium access via adaptive backoff: Delay and loss minimization

Gardar Hauksson and Murat Alanyali
Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215
E-mail: {gardarh,alanyali}@bu.edu

Abstract— **We consider packet transmission scheduling at the MAC-layer via adaptive backoff algorithms that are favorable in terms of queue occupancies in a wireless network. General network topologies are considered under the operational constraint that transmitters within close proximity of each other cannot be transmitting simultaneously. Transmitters probe the channel at random instants and transmit if the channel is idle. We adopt a measurement based framework in which channel probing rates are adaptively determined based on feedback measured from the network. We consider two separate objectives associated with minimization of packet delay and packet loss rate in the network. We obtain dynamic algorithms that strictly improve channel access rates in a related fluid model. Analytical development of the algorithms is based on a convenient decomposition technique that decouples channel access and queue occupancy statistics, and that leads to a favorable tradeoff between analytical insight and modeling accuracy. Obtained algorithms are oblivious to network load and topology. We also consider versions of these algorithms that are suitable for distributed implementation and study their effectiveness numerically.**

I. Introduction

In this paper we study wireless transmission scheduling at the MAC layer from the viewpoint of packet delay and packet loss. Scheduling is an essential component of radio communications on a common frequency spectrum as simultaneous transmissions in close proximity destructively interfere with each other, thereby resulting in reduced spectrum utilization. Hence, ideally, geographically separated communication sessions should be active at a time; furthermore transmission decisions should be local and require computational complexities that scale well with network size. While simultaneous activation of only non-interfering transmissions can be achieved via distributed collision-avoidance techniques, in typical situations of interest there are multiple such patterns of transmissions that can be scheduled at a time. The goal of this paper is to shed some light on practically suitable algorithms to activate non-interfering transmissions in a way to satisfy a range of objectives related to network performance.

Consideration of transmission scheduling in wireless networks has focused to a large extent on achieving throughput optimality, or a fraction thereof, via distributed and computationally efficient algorithms. The issue has been extensively studied by several authors at the MAC layer (e.g. [2, 3, 7, 9, 10]) and also at the networking layer (e.g. [3, 4, 8, 11]). We do not attempt here a satisfactory survey of important results in this area but rather refer the reader to the detailed accounts given in [9, 10]. Interest on performance measures beyond stability, on the other hand, has been remarkably less pronounced.

Here we bypass the issue of stability by adopting a model that involves finite but otherwise arbitrary queue capacities, and focus on packet transmission scheduling algorithms that are favorable in terms of queue occupancies. The scope of the paper is limited to backoff-based channel access schemes under which transmission attempts are postponed to a randomly chosen future time if channel is sensed busy. Such schemes are practically appealing as they lend themselves to distributed implementation. Our main interest is on distributed and adaptive determination of mean backoff durations that maximize certain measures of performance in the network.

We consider general network topologies under a Markovian model that involves dynamic packet arrivals to individual transmitters. It is assumed that each transmitter maintains a channel access rate that is dynamically adjusted based on feedback measured from the network. The model includes both queueing and possible loss of packets due to finite queue capacities. We consider two objectives associated with packet delay and packet loss rate in the network. By way of studying a fluid model we obtain update rules for the channel access rates that render each objective function monotonically decreasing in time. The analysis is based on a convenient decomposition technique for channel access and equilibrium queue lengths. This technique appears to lead to a remarkably favorable tradeoff between analytical insight and modeling accuracy. The resulting adaptive algorithms are oblivious to network topology and arrival rates. We also consider versions of these algorithms that are suitable for distributed implementation, and numerically verify their effectiveness.

The paper is organized as follows: We continue with a formal specification of the network model in Section II. Section III describes the simplifying decomposition employed in the analysis of the paper. The two optimization problems are then defined in Section IV and corresponding adaptive algorithms are developed in Section V. Section VI discusses implementation issues and Section VII reports numerical evaluation of the algorithms. The paper concludes with final remarks in Section VIII.

This work was supported in part by the US National Science Foundation under grants CCF-0430983, CNS-0721860 and ANI-0238397.

978-1-4244-2026-1/08/$25.00

II. Model

Consider a wireless network consisting of m transmitters that operate on a common frequency band. Operational consequences of interference in this network are represented via a graph G in which each node represents a distinct transmitter and it is understood that two transmitters cannot transmit simultaneously if their associated nodes are connected with an edge. Hence a collection of transmitters can be simultaneously active only if the collection forms an independent set in G.

We consider the network under a dynamic traffic model in which each transmitter receives a stream of packets to be transmitted subject to the alluded constraints. Packets arrive at transmitters according to independent Poisson processes and the rate of packet arrivals to transmitter i is denoted by λ_i. Transmission of a packet takes an exponentially distributed amount of time with unit mean. Hence packets may be queued until completion of their transmission, and we assume that each transmitter i has room for storing C_i packets including any packet in transmission. A packet is lost if it arrives at a transmitter when the transmitter has no room in its queue.

We focus on randomized backoff algorithms to regulate channel access in the network. Namely, we consider a scheme where each transmitter i makes autonomous channel access decisions via a local parameter $r_i \geq 0$ in the following manner: After either completing transmission of a packet or sensing another transmission on the channel, the transmitter sets a random timeout that is exponentially distributed with mean r_i^{-1}. When the timeout expires, the transmitter does a carrier-sense on the channel and claims the channel by starting a new packet transmission if the channel is idle, or sets another independently chosen timeout otherwise. A sketch of randomized backoff is given in Figure 1.

Note that channel access decisions are made without regard to the queue occupancy at the transmitter; in particular a transmitter may gain access of the channel when it has no packets awaiting transmission. In that case we assume that the transmitter holds the channel for the duration of a hypothetical packet anyway, which may be interpreted as transmission of a ghost packet. The attendant underutilization of spectrum clearly leads to a lower bound on the achievable performance in the network. Nevertheless this assumption substantially simplifies the subsequent analysis, and the resulting algorithms can be applied without resorting to ghost packet transmissions.

Let n_i denote the queue length at transmitter i (including the packet being transmitted if any) and let s_i indicate activity of the transmitter so that

$$s_i = \begin{cases} 1 & \text{if transmitter } i \text{ is holding the channel} \\ 0 & \text{otherwise.} \end{cases}$$

Note that for fixed choice $\boldsymbol{r} = (r_1, r_2, \cdots, r_m)$ of channel access rates, $\boldsymbol{s} = (s_1, s_2, \cdots, s_m)$ is an ergodic Markov process on the state space $\mathcal{S}_G$ that is defined as

$$\mathcal{S}_G = \{\boldsymbol{s} \in \{0,1\}^m : s_i s_j = 0 \text{ if } i, j \text{ are neighbors in } G\}.$$

Namely, $\mathcal{S}_G$ is the collection of independent sets of G. The joint process $(\boldsymbol{s}, \boldsymbol{n}) = (s_1, n_1, s_2, n_2, \cdots, s_m, n_m)$ is also

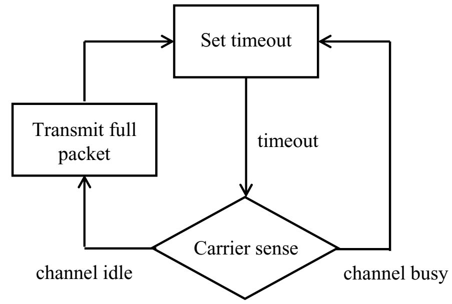


Fig. 1. Local channel access algorithm at each transmitter i. Timeouts and packet transmission times are independently and exponentially distributed, with respective means r_i^{-1} and 1.

Markovian, and it is irreducible on a state space that is finite owing to the finite queue capacities. In particular the process $(\boldsymbol{s}, \boldsymbol{n})$ is ergodic regardless of the choice of $\boldsymbol{r}$. Given a nonnegative vector $\boldsymbol{r}$ we shall denote by $\pi^{\boldsymbol{r}}$ the equilibrium distribution of $(\boldsymbol{s}, \boldsymbol{n})$ when each transmitter i operates with the fixed channel access rate r_i.

We study adaptive algorithms to determine favorable channel access rates $\boldsymbol{r}$ via differential systems of type

$$\dot{\boldsymbol{r}} = f(\pi^{\boldsymbol{r}}). \tag{1}$$

Here $\dot{\boldsymbol{r}}$ denotes time derivative and f is a Lipschitz continuous function to be determined. The differential system above has a natural interpretation in terms of time-scale separation between the evolution of $\boldsymbol{r}$ and the process $(\boldsymbol{s}, \boldsymbol{n})$. Namely, it is understood that the rates $\boldsymbol{r}$ are updated on a slower time scale and only after the distribution of $(\boldsymbol{s}, \boldsymbol{n})$ settles to its equilibrium. An important advantage of this model is that evolution of channel access rates $\boldsymbol{r}$ does not require explicit knowledge of network topology, packet arrivals rates, or prior values of $\boldsymbol{r}$. Rather, the model is tailored for identifying adaptive algorithms that update $\boldsymbol{r}$ based on suitable measurement of network statistics.

While the model (1) may be justified via a fluid limit analysis, here we do not pursue that direction but concentrate on the analytical insight obtained from this model. Our main focus is thus to identify appropriate definitions of the mapping f under which $\boldsymbol{r}$ is driven towards values that achieve certain objectives such as minimization of packet delay and loss in the network.

III. Approximate Occupancy Model

Analyzing solutions of (1) for a given f is typically hindered by difficulties in computing the distribution $\pi^{\boldsymbol{r}}$: Although the channel access process $\boldsymbol{s}$ is reversible and the network occupancy process $\boldsymbol{n} = (n_1, n_2, \cdots, n_m)$ is reversible provided that $\boldsymbol{s}$ is fixed, the combined process $(\boldsymbol{s}, \boldsymbol{n})$ is not reversible and obtaining a closed form for $\pi^{\boldsymbol{r}}$ appears difficult. This section introduces an approximation to $(\boldsymbol{s}, \boldsymbol{n})$ that is based on a suitable decoupling of the component processes $\boldsymbol{s}$ and $\boldsymbol{n}$, and that leads to significant algorithmic insight at the expense of reasonable reduction in modeling accuracy.

Given $\boldsymbol{r}$, detailed balance equations (see for example [5]) can be verified to show that the marginal distribution of $\boldsymbol{s}$ in equilibrium is given by $\pi_1^{\boldsymbol{r}}$ where

$$\pi_1^{\boldsymbol{r}}(\boldsymbol{s}) = \frac{1}{Z(G,\boldsymbol{r})}\prod_{i=1}^{m} r_i^{s_i}, \qquad \boldsymbol{s} \in \mathcal{S}_G \tag{2}$$

and

$$Z(G,\boldsymbol{r}) = \sum_{\boldsymbol{s}\in\mathcal{S}_G}\prod_{i=1}^{m} r_i^{\boldsymbol{s}_i}.$$

In particular a given transmitter i holds the channel for a fraction $\mu_i^{\boldsymbol{r}}$ of time where

$$\mu_i^{\boldsymbol{r}} = \pi_1^{\boldsymbol{r}}(s_i = 1) = r_i\frac{Z(G_{-i},\boldsymbol{r})}{Z(G,\boldsymbol{r})}. \tag{3}$$

Here G_{-i} denotes the subgraph of G obtained by deleting node i and its neighbors, along with any edges that are incident on these nodes.

Let $\pi_2^{\boldsymbol{r}}$ denote the equilibrium occupancy distribution of m independent M/M/1 queues where the ith queue has arrival rate λ_i, service rate $\mu_i^{\boldsymbol{r}}$ and total waiting room for C_i packets. Namely,

$$\pi_2^{\boldsymbol{r}}(\boldsymbol{n}) = \prod_{i=1}^{m}\frac{\rho_i^{n_i}}{\sum_{k=0}^{C_i}\rho_i^k},$$

where

$$\rho_i = \lambda_i/\mu_i^{\boldsymbol{r}}.$$

Note that the dependence of ρ_i on $\boldsymbol{r}$ is suppressed for the sake of notational convenience.

We approximate $\pi^{\boldsymbol{r}}$ with the distribution $\pi_o^{\boldsymbol{r}}$ where

$$\pi_o^{\boldsymbol{r}}(\boldsymbol{s},\boldsymbol{n}) = \pi_1^{\boldsymbol{r}}(\boldsymbol{s})\pi_2^{\boldsymbol{r}}(\boldsymbol{n}).$$

Note that sample paths of processes that admit $\pi_o^{\boldsymbol{r}}$ as equilibrium distribution differ than those of the process $(\boldsymbol{s},\boldsymbol{n})$ specified in Section II as the former may allow simultaneous channel access by neighboring transmitters. The equilibrium distributions, however, typically display remarkable agreement. Example 1 below numerically illustrates accuracy of the approximate distribution $\pi_o^{\boldsymbol{r}}$ in an arbitrarily chosen network topology.

Example 1: A comparison of expected queue lengths and queue length distributions with respect to $\pi^{\boldsymbol{r}}$ (obtained by simulation) and $\pi_o^{\boldsymbol{r}}$ is illustrated in Figure 2. Here 7 transmitters are considered under a network topology G that is represented by adjacency matrix

$$A = [A_{ij}] = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix},$$

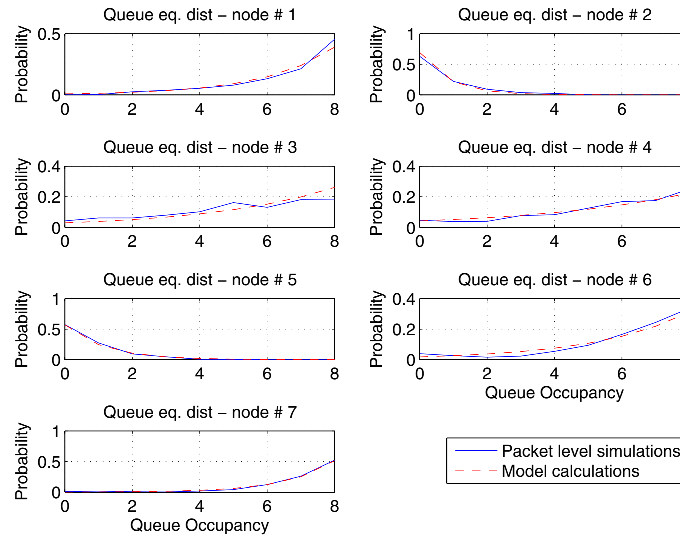


Fig. 2. *Queue occupancy distributions in Example 1 obtained by simulation of the original system and via the approximate distribution $\pi_2^{\boldsymbol{r}}$. The total variation of the distributions of the 7 nodes is* 0.032 ± 0.006.

where $A_{ij} = 1$ if and only if there is an edge between nodes i and j. Channel access rates $\boldsymbol{r}$, traffic loads $\boldsymbol{\lambda}$ and buffer capacities $\boldsymbol{C} = (C_1, \cdots, C_7)$ are taken as

$$\begin{aligned} \boldsymbol{r} &= (1,1,1,1,1,1,1) \\ \boldsymbol{\lambda} &= (0.26, 0.06, 0.42, 0.30, 0.17, 0.46, 0.41) \\ \boldsymbol{C} &= (8,8,8,8,8,8,8). \end{aligned}$$

Decoupling techniques that are similar in flavor to the one employed here have found fruitful application in several engineering contexts [1]. Typical applications therein rely on a time-scale separation between coordinate processes. There is no such separation here between processes $\boldsymbol{n}$ and $\boldsymbol{s}$; the present approach merely matches the fraction of time that the channel is available to each transmitter with the actual model and approximates equilibrium occupancies via those of a convenient model.

IV. Problem Definition

Let $E_{\boldsymbol{r}}$ denote expectation with respect to distribution $\pi_o^{\boldsymbol{r}}$ and let $w_i > 0$ be an arbitrary but fixed weight for each transmitter i. We are interested in identifying channel access rates $\boldsymbol{r}$ that are favorable in minimizing the following quantities:

1) (Packet delay)

$$J_1(\boldsymbol{r}) = E_{\boldsymbol{r}}\left[\sum_i w_i n_i\right],$$

2) (Packet loss rate)

$$J_2(\boldsymbol{r}) = \sum_i w_i\lambda_i\pi_o^{\boldsymbol{r}}(n_i = C_i).$$

Both objective functions are bounded and differentiable at $\boldsymbol{r} > 0$. The functions are separable so that

$$J_k(\boldsymbol{r}) = \sum_i J_{ik}(\mu_i^{\boldsymbol{r}}) \tag{4}$$

where each $J_{ik} : [0,1] \mapsto \mathbb{R}_+$ is strictly decreasing. Let $N(i)$ denote the neighborhood of each transmitter i (including i itself) in the graph network G. Since for each choice of $\boldsymbol{r}$

$$\sum_{j \in N(i)} \mu_j^{\boldsymbol{r}} \leq 1 \quad \text{for} \quad i = 1, 2, \cdots, m, \tag{5}$$

it follows that

$$J_k(\boldsymbol{r}) \geq \min \; \{ \sum_i J_{ik}(\mu_i) : \sum_{j \in N(i)} \mu_j \leq 1 \text{ for all } i \; \}. \tag{6}$$

Note that tightness of this lower bound is not immediate as it is not clear whether given $\mu_1, \mu_2, \cdots, \mu_m$ that satisfy (5) one can find access rates $\boldsymbol{r}$ such that $\mu_i^{\boldsymbol{r}} = \mu_i$ for all i.

We make the following observations concerning possible minima of $J_1(\boldsymbol{r}), J_2(\boldsymbol{r})$.

Lemma 4.1: For any transmitters i, j and $\boldsymbol{r}$ with $r_i > 0$

$$\frac{d}{dr_i}\mu_j^{\boldsymbol{r}} = r_i^{-1}(\pi_o^{\boldsymbol{r}}(s_i = s_j = 1) - \pi_o^{\boldsymbol{r}}(s_i = 1)\pi_o^{\boldsymbol{r}}(s_j = 1)).$$

Proof: By definition (3)

$$\mu_j^{\boldsymbol{r}} = \frac{\sum_{\boldsymbol{s} \in \mathcal{S}_G : s_j = 1} \prod_{k=1}^m r_k^{s_k}}{\sum_{\boldsymbol{s} \in \mathcal{S}_G} \prod_{k=1}^m r_k^{s_k}};$$

in turn

$$\begin{aligned} \frac{d}{dr_i}\mu_j^{\boldsymbol{r}} \quad = \quad & \frac{\sum_{\boldsymbol{s}: s_j = 1} \frac{d}{dr_i} \prod_{k=1}^m r_k^{s_k}}{\sum_{\boldsymbol{s}} \prod_{k=1}^m r_k^{s_k}} \\ & - \frac{\sum_{\boldsymbol{s}: s_j = 1} \prod_{k=1}^m r_k^{s_k}}{\sum_{\boldsymbol{s}} \prod_{k=1}^m r_k^{s_k}} \times \frac{\sum_{\boldsymbol{s}} \frac{d}{dr_i} \prod_{k=1}^m r_k^{s_k}}{\sum_{\boldsymbol{s}} \prod_{k=1}^m r_k^{s_k}}. \end{aligned} \tag{7}$$

If $r_i > 0$ then

$$\frac{d}{dr_i} \prod_{k=1}^m r_k^{s_k} = 1\{s_i = 1\} \frac{1}{r_i} \prod_{k=1}^m r_k^{s_k}.$$

The proof is established by substituting this expression in equality (7). ■

Theorem 4.2: The function $J_k(\boldsymbol{r})$, $k = 1, 2$, has no local minimum in the interior of $\mathbb{R}_+^m$.

Proof: In view of the representation (4) of $J_k(\boldsymbol{r})$ it suffices to establish that given any $\boldsymbol{r} > 0$ (that is, with strictly positive entries) there exist arbitrarily close vectors $\boldsymbol{r}_o$ such that $\mu_i^{\boldsymbol{r}_o} \geq \mu_i^{\boldsymbol{r}}$ for all i and $\mu_i^{\boldsymbol{r}_o} > \mu_i^{\boldsymbol{r}}$ for at least one i. For notational convenience let

$$c_{ij}(\boldsymbol{r}) = \pi_o^{\boldsymbol{r}}(s_i = s_j = 1) - \pi_o^{\boldsymbol{r}}(s_i = 1)\pi_o^{\boldsymbol{r}}(s_j = 1)$$

for $i, j = 1, 2, \cdots, m$. The change in $\mu_i^{\boldsymbol{r}}$ corresponding to an infinitesimal change $(\delta r_1, \delta r_2, \cdots, \delta r_m)$ in $\boldsymbol{r}$ is given by

$$\delta\mu_j^{\boldsymbol{r}} \; = \; \sum_i \left(\frac{d}{dr_i}\mu_j^{\boldsymbol{r}}\right)\delta r_i \; = \; \sum_i \frac{1}{r_i} c_{ij}(\boldsymbol{r}) \delta r_i, \tag{8}$$

where the last equality is due to Lemma 4.1. Let us define the matrices

$$\begin{aligned} \delta\boldsymbol{\mu} &= [\delta\mu_i^{\boldsymbol{r}}]_{1 \times m} \\ \delta\boldsymbol{r} &= [\delta r_i]_{1 \times m} \\ R &= Diag[r_i^{-1}]_{m \times m} \\ \Sigma &= [c_{ij}(\boldsymbol{r})]_{m \times m} \end{aligned}$$

so that equalities (8) can be expressed in vector form

$$\delta\boldsymbol{\mu} = \delta\boldsymbol{r} R \Sigma. \tag{9}$$

Note that Σ is a covariance matrix and it is positive definite for $\boldsymbol{r} > 0$. Hence $R\Sigma$ is invertible and given $\delta\boldsymbol{\mu}$ there exists a vector $\delta\boldsymbol{r}$ such that equality (9) is realized. In particular one can choose $\delta\boldsymbol{\mu}$ such that $\delta\boldsymbol{\mu} \geq 0$ with strict inequality in at least one entry; so $\boldsymbol{r} > 0$ cannot be a local minimum of $J_k(\boldsymbol{r})$. ■

Corollary 4.3: A network model with $r_i = 0$ for some transmitter i is equivalent to a reduced model obtained by excluding all such transmitters so that $\boldsymbol{r} > 0$ and by inheriting the topology from the interference relations in the original model. Since Theorem 4.2 is valid regardless of the network topology, the considered objective functions do not have any local minima on the facets of $\mathbb{R}_+$ either.

V. Adaptive Backoff Policies

The goal of this section is to develop dynamic schemes to adaptively adjust channel access rates $\boldsymbol{r}$ in order to improve the two objective functions introduced in the previous section. We start with reciting a useful fact in the present context.

Lemma 5.1: If, for $k = 1, 2$,

$$\dot{r}_i = -\frac{d}{dr_i} J_k(\boldsymbol{r}), \qquad i = 1, 2, \cdots, m,$$

then $J_k(\boldsymbol{r})$ strictly decreases with time provided that $\boldsymbol{r}$ has positive initial value.

Proof: Since

$$\dot{J}_k(\boldsymbol{r}) = \sum_i \dot{r}_i \frac{d}{dr_i} J_k(\boldsymbol{r}) = -\sum_i \left(\frac{d}{dr_i} J_k(\boldsymbol{r})\right)^2,$$

$J_k(\boldsymbol{r})$ is strictly decreasing along trajectories of $\boldsymbol{r}$ satisfying the hypothesis, as long as not all derivatives $\frac{d}{dr_i} J(\boldsymbol{r})$ are simultaneously 0. By Corollary 4.3 no such $\boldsymbol{r}$ exists. ■

Lemma 5.1 stops short of promising convergence to a local minimum as no such value exists due to Corollary 4.3. Nevertheless the lemma still suggests effective policies for access rate adaptation as described by the following two theorems.

Theorem 5.2: (Packet delay) Let $\sigma_{\boldsymbol{r}}^2(n_i)$ denote the variance of queue length n_i with respect to distribution $\pi_o^{\boldsymbol{r}}$. If for each i

$$\dot{r}_i = \frac{1}{r_i} \sum_j \left(\pi_o^{\boldsymbol{r}}(s_i = 1 | s_j = 1) - \pi_o^{\boldsymbol{r}}(s_i = 1)\right) w_j \sigma_{\boldsymbol{r}}^2(n_j) \tag{10}$$

then $J_1(\boldsymbol{r})$ strictly decreases with time provided that $\boldsymbol{r}$ has positive initial value.

Proof: We prove the theorem by identifying the right hand side of equation (10) with $-\frac{d}{dr_i}J_1(\boldsymbol{r})$ and then by appealing to Lemma 5.1. Note that

$$\begin{aligned} \frac{d}{dr_i}J_1(\boldsymbol{r}) &= \sum_j w_j \frac{d}{dr_i} E_{\boldsymbol{r}}[n_j] \\ &= \sum_j w_j \frac{d\rho_j}{dr_i}\frac{d}{d\rho_j}E_{\boldsymbol{r}}[n_j], \end{aligned} \tag{11}$$

where

$$\begin{aligned} \frac{d}{d\rho_j}E_{\boldsymbol{r}}[n_j] &= \frac{d}{d\rho_j}\frac{\sum_{k=0}^{C_j} k\rho_j^k}{\sum_{k=0}^{C_j}\rho_j^k} \\ &= \frac{\sum_{k=0}^{C_j} k^2\rho_j^{k-1}}{\sum_{k=0}^{C_j}\rho_j^k} - \frac{\sum_{k=0}^{C_j} k\rho_j^k \sum_{k=0}^{C_i} k\rho_j^{k-1}}{\left(\sum_{k=0}^{C_j}\rho_j^k\right)^2} \\ &= \frac{1}{\rho_j}\left(\frac{\sum_{k=0}^{C_j} k^2\rho_j^k}{\sum_{k=0}^{C_j}\rho_j^k} - \left(\frac{\sum_{k=0}^{C_j} k\rho_j^k}{\sum_{k=0}^{C_j}\rho_j^k}\right)^2\right) \\ &= \frac{1}{\rho_j}\sigma^2_{\boldsymbol{r}}(n_j). \end{aligned} \tag{12}$$

We finally observe that

$$\frac{d\rho_j}{dr_i} = -\frac{d\mu_j^{\boldsymbol{r}}}{dr_i}\frac{\lambda_j}{(\mu_j^{\boldsymbol{r}})^2} \tag{13}$$

where by Lemma 4.1 and definition (3) of $\mu_j^{\boldsymbol{r}}$

$$\frac{d}{dr_i}\mu_j^{\boldsymbol{r}} = \frac{\mu_j^{\boldsymbol{r}}}{r_i}\left(\pi_o^{\boldsymbol{r}}(s_i=1|s_j=1) - \pi_o^{\boldsymbol{r}}(s_i=1)\right). \tag{14}$$

Equalities (11)–(14) lead to the conclusion

$$\dot{r}_i = -\frac{d}{dr_i}J_1(\boldsymbol{r}), \qquad i = 1,2,\cdots,m;$$

in turn assertion of the theorem follows from Lemma 5.1. ■

Theorem 5.3: (Packet loss rate) If for each i

$$\begin{aligned} \dot{r}_i &= \frac{1}{r_i}\sum_j (\pi_o^{\boldsymbol{r}}(s_i=1|s_j=1) - \pi_o^{\boldsymbol{r}}(s_i=1)) \\ &\quad \times w_j\lambda_j\pi_o^{\boldsymbol{r}}(n_j=C_j)\,(C_j - E_{\boldsymbol{r}}[n_j]) \end{aligned} \tag{15}$$

then $J_2(\boldsymbol{r})$ strictly decreases with time provided that $\boldsymbol{r}$ has positive initial value.

Proof: Since

$$J_2(\boldsymbol{r}) = \sum_j w_j\lambda_j \frac{\rho_j^{C_j}}{\sum_{k=0}^{C_j}\rho_j^k}$$

it follows that

$$\frac{d}{dr_i}J_2(\boldsymbol{r}) = \sum_j w_j\lambda_j\frac{d\rho_j}{dr_i}\frac{d}{d\rho_j}\frac{\rho_j^{C_j}}{\sum_{k=0}^{C_j}\rho_j^k} \tag{16}$$

The last derivative on the right hand side is evaluated as

$$\begin{aligned} \frac{d}{d\rho_j}\frac{\rho_j^{C_j}}{\sum_{k=0}^{C_j}\rho_j^k} &= \frac{C_j\rho_j^{C_j-1}\sum_{k=0}^{C_j}\rho_j^k - \rho_j^{C_j}\sum_{k=0}^{C_j}k\rho_j^{k-1}}{\left(\sum_{k=0}^{C_j}\rho_j^k\right)^2} \\ &= \frac{1}{\rho_j}\frac{\rho_j^{C_j}}{\sum_{k=0}^{C_j}\rho_j^k}\left(C_j - \frac{\sum_{k=0}^{C_j}k\rho_j^k}{\sum_{k=0}^{C_j}\rho_j^k}\right) \\ &= \frac{1}{\rho_j}\pi_o^{\boldsymbol{r}}(n_j=C_j)\,(C_j - E_{\boldsymbol{r}}[n_j]) \end{aligned} \tag{17}$$

Equalities (16)–(17) together with equalities (13)–(14) lead to the conclusion

$$\dot{r}_i = -\frac{d}{dr_i}J_2(\boldsymbol{r}), \qquad i = 1,2,\cdots,m.$$

The theorem now follows from Lemma 5.1. ■

VI. Implementation Issues

Theorems 5.2–5.3 prescribe adaptation of the channel access rates $\boldsymbol{r}$ according to the common template

$$\dot{r}_i = \frac{1}{r_i}\sum_{j=1}^m \gamma_{ij}(\boldsymbol{r})\phi_j(\boldsymbol{r}), \tag{18}$$

where $1/r_i$ is a local parameter at transmitter i, $\phi_j(\boldsymbol{r})$ is locally measurable at each remote transmitter j and

$$\gamma_{ij}(\boldsymbol{r}) = \pi_o^{\boldsymbol{r}}(s_i=1|s_j=1) - \pi_o^{\boldsymbol{r}}(s_i=1)$$

is a measure of the statistical correlation between channel access by transmitters i and j. This quantity determines both the strength and and the sign of how $\phi_j(\boldsymbol{r})$ influences r_i. In particular if the two transmitters have statistically independent channel access patterns then $\phi_j(\boldsymbol{r})$ is not taken into account in updating r_i.

A remarkable property of the general structure (18) is that the same value $\phi_j(\boldsymbol{r})$ from transmitter j is required to determine evolution of the channel access rate r_i of each transmitter i. Hence structural properties of Theorems 5.2–5.3 suggest algorithmic possibilities that are based on conveying the quantities $\phi_j(\boldsymbol{r})$ to all nodes in the network, which then update their channel access rates to mimick the evolution indicated by the theorems.

A significant impediment in realizing this scenario, however, is determining the correlation values $\gamma_{ij}(\boldsymbol{r})$. On the one hand computation of $\gamma_{ij}(\boldsymbol{r})$ requires knowledge of both $\boldsymbol{r}$ and the network graph G. On the other hand measuring $\gamma_{ij}(\boldsymbol{r})$ requires an oracle that has access to the whole network state with possibly very accurate timing. The situation is relieved to some extent by observing that the correlation between activities of transmitters i and j is relatively small if i, j are sufficiently separated in the network graph G; hence restricting the sum in (18) to a small subset of nodes j for which $\gamma_{ij}(\boldsymbol{r})$ can be determined may possibly lead to practically acceptable performance loss in the network. The following example illustrates the decay rate of $\gamma_{ij}(\boldsymbol{r})$ with the hop distance between transmitters i, j in an arguably pessimistic topology which has high correlation between node pairs.

Example 2: (Infinite line) Consider a large number of transmitters that are indexed by integers $(\cdots, -2, -1, 0, 1, 2, \cdots)$, and are arranged on a line so that G has an edge between i, j if $|i-j| = 1$. Let $\boldsymbol{r}$ be such that $r_i = r > 0$ for all i. Let a be the unique positive root of

$$ra^2 + a - 1 = 0$$

which can be verified to be less than 1, and define

$$\sigma = \left[\frac{1}{1+ra}, \quad \frac{ra}{1+ra}\right] \quad \text{and} \quad Q = \left[\begin{array}{cc} a & 1-a \\ 1 & 0 \end{array}\right].$$

Note that the probability vector $\sigma = [\sigma(0), \sigma(1)]$ is an eigenvector of the stochastic matrix Q corresponding to the eigenvalue 1, and that the remaining eigenvalue of Q is $1 - a$. For this particular topology and rate vector $\boldsymbol{r}$ the distribution (2) of $\boldsymbol{s}$ has a convenient from such that for $i < j$

$$\pi_o^{\boldsymbol{r}}(s_i, s_{i+1}, \cdots, s_j) = \sigma(s_i) \prod_{k=i}^{j-1} Q(s_k, s_{k+1})$$

(see, for example, [6]). In particular

$$\pi_o^{\boldsymbol{r}}(s_j | s_i) = Q^{|i-j|}(s_i, s_j),$$

where Q^k denotes matrix product; and therefore

$$\begin{aligned} |\gamma_{ij}(\boldsymbol{r})| &= |\pi_o^{\boldsymbol{r}}(s_j = 1 | s_i = 1) - \pi_o^{\boldsymbol{r}}(s_j = 1)| \\ &= |Q^{|i-j|}(1,1) - \sigma(1)| \\ &= O((1-a)^{|i-j|}). \end{aligned}$$

Hence the correlation $\gamma_{ij}(\boldsymbol{r})$ decays exponentially fast with the hop-distance between transmitters i and j.

A. Distributed implementation

The prospect of fast decay of $\gamma_{ij}(\boldsymbol{r})$ suggests limiting the sum in (18) to a collection transmitters j that are close to i with respect to the network graph G. One such collection that has significant practical appeal is the immediate neighbors of i, which we discuss in this section. Namely, we consider here updating channel access rates via

$$\dot{r}_i = \frac{1}{r_i} \sum_{j \in N(i)} \gamma_{ij}(\boldsymbol{r}) \phi_j(\boldsymbol{r}), \tag{19}$$

where $N(i)$ denotes the set that comprises i and its neighbors in the network graph G. Note that the correlation coefficients involved in (19) are

$$\gamma_{ij}(\boldsymbol{r}) = \begin{cases} 1 - \mu_i^{\boldsymbol{r}} & \text{if } j = i \\ -\mu_j^{\boldsymbol{r}} & \text{if } j \in N(i) - \{i\}. \end{cases}$$

The product $\gamma_{ii}(\boldsymbol{r})\phi_i(\boldsymbol{r})$ is locally computable at transmitter i. The remaining terms of the sum in (19) can be computed at i if every transmitter j piggybacks the value of $\phi_j(\boldsymbol{r})$ on the packets (or a certain fraction thereof) that it transmits. More explicitly, let $\phi^k(\boldsymbol{r})$ denote the kth value that transmitter i obtains from a packet that is transmitted in its neighborhood (that is, by some transmitter that is a neighbor of i in G). Transmitter i may maintain variables ζ, τ via exponential averaging

$$\begin{aligned} \zeta^k &= \beta \zeta^{k-1} + (1-\beta) \phi^k(\boldsymbol{r}) \\ \tau^k &= \beta \tau^{k-1} + (1-\beta) t^k, \end{aligned}$$

where t^k denotes the time since last packet transmission in the neighborhood of i, and $0 < \beta < 1$. The non-local part of the summation on the right hand side of (19) can then be approximated by

$$\sum_{j \in N(i) - \{i\}} \gamma_{ij}(\boldsymbol{r}) \phi_j(\boldsymbol{r}) \approx -\frac{\zeta^k}{\tau^k}.$$

Accuracy of the approximation improves if $\boldsymbol{r}$ does not change substantially before sufficiently many packets are transmitted in each neighborhood. Hence the update rule (19) may be implemented in a distributed fashion and with reasonable control signalling and algorithmic overhead.

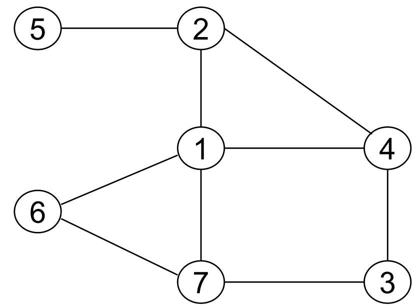


Fig. 3. *The network graph that represents channel access constraints in Example 1.*

VII. Packet-Level Simulations

We numerically verify the analytical insight developed in the earlier sections by comparing the considered differential model with packet level simulations of adaptive backoff policies that are suggested by Theorems 5.2–5.3. In this section we adopt the topology G, capacities $\boldsymbol{C}$, and traffic load $\boldsymbol{\lambda}$ considered in Example 1. The network topology corresponding to the adjacency matrix A is sketched in Figure 3. Both objective functions $J_1(\boldsymbol{r})$ and $J_2(\boldsymbol{r})$ are considered with the following collection of arbitrarily chosen transmitter weights:

$$\boldsymbol{w} = (4.75, 1.16, 3.03, 2.43, 4.46, 3.81, 2.28).$$

Figures 4(a) and 4(b) illustrate the evolution of channel access rates $\boldsymbol{r}$ for the 7 transmitters, respectively under the differential system prescribed by Theorem 5.2 and under simulation of the discrete-time update rule

$$r_i^k = \left(r_i^{k-1} + \varepsilon \frac{1}{r_i^k} \sum_{j=1}^m \gamma_{ij}^k(\boldsymbol{r}) \phi_j^k(\boldsymbol{r}) \right)_+, \tag{20}$$

that mimicks the differential equality (10). Here $(x)_+$ represents $\max(x, 0)$, each $\phi_j^k(\boldsymbol{r})$ is obtained via local measurements, and each $\gamma_{ij}^k(\boldsymbol{r})$ is obtained via centralized averaging of joint activity of transmitters i and j. In view of Corollary 4.3, the figures suggest that $\boldsymbol{r}$ seeks a local minimum of the objective function by increasing indefinitely in all coordinates except r_7, which tends to 0 asymptotically. Note that transmitter 7 has degree 3 in the network graph; hence it uses up considerable network resources. Although transmitter 1 has higher degree, the total weight that is incident to transmitter 7 is 11.60 whereas the same quantity is 9.68 for transmitter 1. In informal terms, the gain due to occupancy reduction in the remaining part of the network outweights the cost incurred by the high (but finite) occupancy of transmitter 7.

Figures 5(a) and 5(b) display simulation traces for the (average) weighted queue length in the network under respectively the update rule (20) and the update rule

$$r_i^k = \left(r_i^{k-1} + \varepsilon \frac{1}{r_i^k} \sum_{j \in N(i)} \gamma_{ij}^k(\boldsymbol{r}) \phi_j^k(\boldsymbol{r}) \right)_+, \tag{21}$$

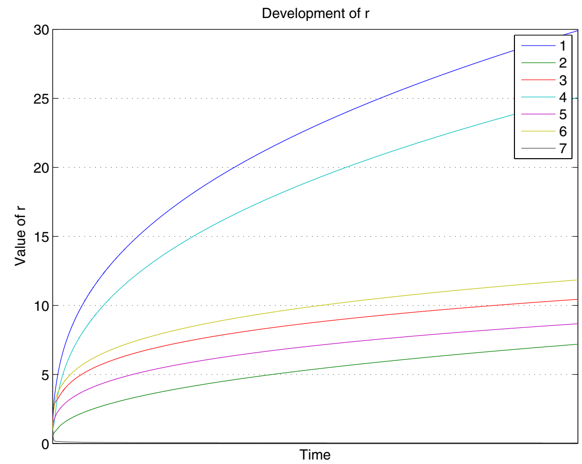
which mimicks the reduced variant (19) of the former rule. The figures also indicate value of the lower bound (6) for the considered objective function. An analogous illustration for the objective function $J_2(\boldsymbol{r})$ is provided by Figures 6(a) and 6(b). The latter two simulation traces can be cross-checked with Figure 7 which plots the $J_2(\boldsymbol{r})$ along solutions of the differential system (15) and its reduced variant (19).

VIII. Final Remarks

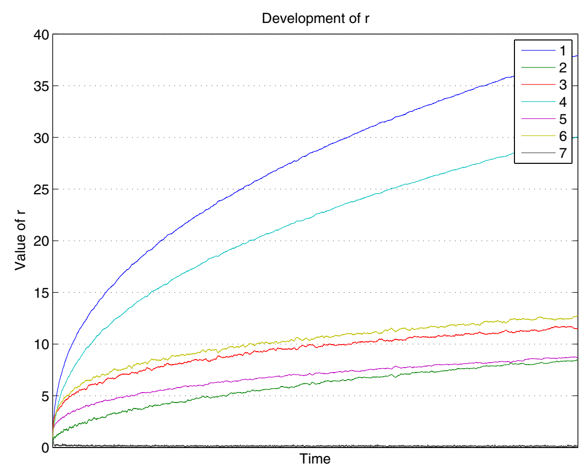
We studied wireless transmission scheduling at the MAC layer, based on a decomposition technique that proved fruitful in leading to a concise approach to seemingly difficult questions considered in the paper. The scope of the paper can be easily extended to cover other objectives that involve a wide range of functions of queue lengths.

No error bounds on accuracy of the approximate occupancy model are readily available. The approximate model is exact in its account for the spatial properties of transmission activity in the network, but involves approximations in the relationship between transmission activity and queue size at individual transmitters; hence its accuracy may be expected to scale gracefully with network size.

A number of questions are left open in the analysis of the approximate model: i) Each objective function has a limit under the associated algorithm, but a characterization of possible such limits is not yet available. ii) Since the considered model involves arbitrarily large buffer sizes the introduced algorithms are possibly throughput optimal within backoff algorithms. At this point it is not clear whether backoff algorithms are throughput optimal. iii) No quantitative bound is available for the performance deterioration due to the reduced-form algorithms identified in Section VI-A. These algorithms possibly minimize related objective functions whose characterization may be instrumental in obtaining the alluded bounds.



(a) Solution of the differential system (10).



(b) Simulation of update rule (20).

Fig. 4. Evolution of channel access rates $\boldsymbol{r}$ for the objective function $J_1(\boldsymbol{r})$.

We finally note that future research might also include finding the convergence rate of the obtained algorithms. This would indicate how adaptive the algorithms would be to modifications of the network topology, which is an important capability in wireless networks.

References

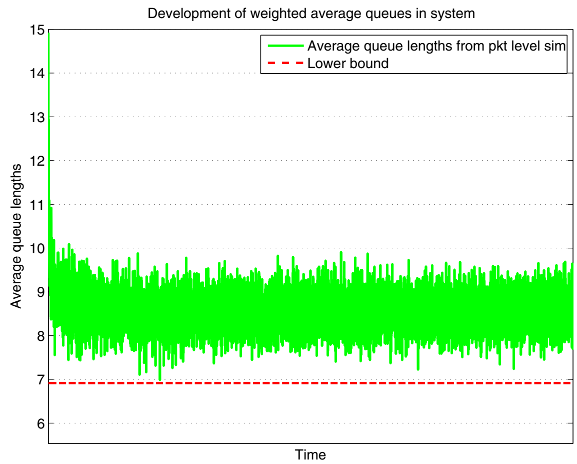
[1] P. J. Courtois. *Decomposability: Queueing and Computer Systems Applications,* Academic Press, New York, 1977.

[2] A. Dimakis and J. Walrand. Sufficient conditions for stability of longest-queue-first scheduling: Second-order properties using fluid limits, *Advances in Applied Probability,* vol. 38, pp. 505–521, 2006.

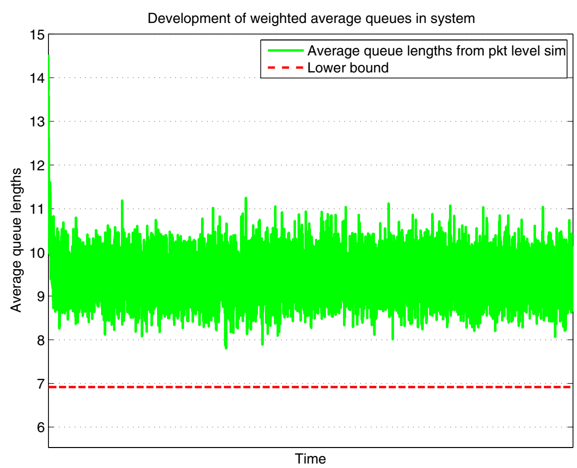
[3] B. Hajek and G. Sasaki. Link scheduling in polynomial time. IEEE Transactions on Information Theory, 34(5):910–917. 1988.

[4] C. Joo and N. B. Shroff. Performance of Random Access Scheduling Schemes in Multi-hop Wireless Networks. IEEE INFOCOM 2007.

[5] F. P. Kelly. *Reversibility and Stochastic Networks,* Wiley, Chichester, 1979.



(a) Simulation of update rule (20).



(b) Simulation of update rule (21).

Fig. 5. Evolution of $m^{-1} \sum_i w_i n_i$ under the two update rules with different information requirements.

[6] F. P. Kelly. Stochastic models of computer communication systems, *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 47, No. 3, pp. 379–395. 1985.

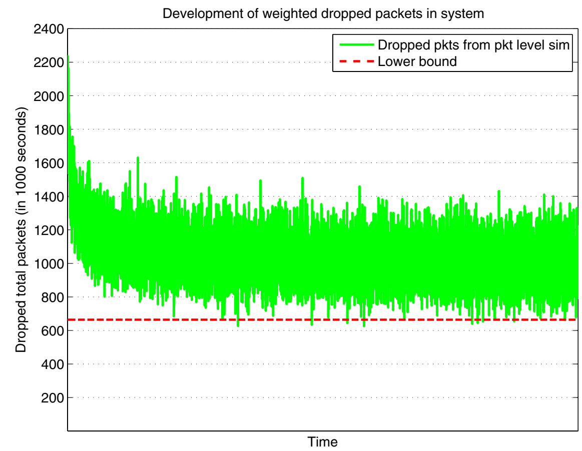
[7] X. Lin and S. Rasool. Constant-Time Distributed Scheduling Policies for Ad Hoc Wireless Networks. IEEE Conference on Decision and Control. 2006.

[8] X. Lin and S. Rasool. A Distributed Joint Channel-Assignment, Scheduling and Routing Algorithm for Multi-Channel Ad Hoc Wireless Networks. IEEE INFOCOM, 2007.

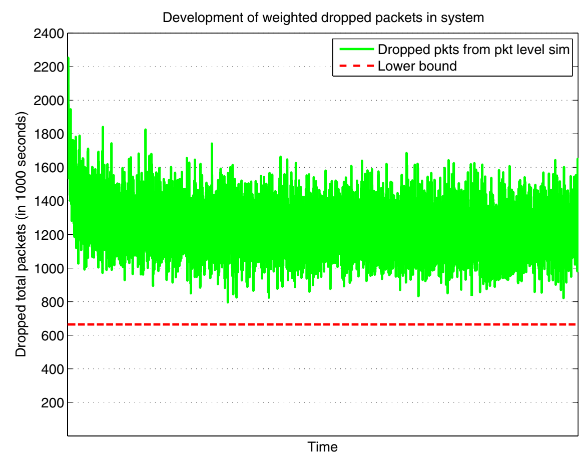
[9] S. Ray and S. Sarkar, Arbitrary Throughput Versus Complexity Tradeoffs in Wireless Networks using Graph Partitioning. Proceedings of Information Theory and Applications Second Workshop, University of California, San Diego, 2007.

[10] S. Sanghavi, L. Bui, and R. Srikant. Distributed link scheduling with constant overhead. SIGMETRICS '07. ACM Press, New York, NY, 313–324. 2007

[11] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. IEEE Transactions on Automatic Control, 37(12):1936–1948, Dec 1992.



(a) Simulation of update rule (20).



(b) Simulation of update rule (21).

Fig. 6. Evolution of packet loss rate (number of lost packet per unit time) under the two update rules with different information requirements.

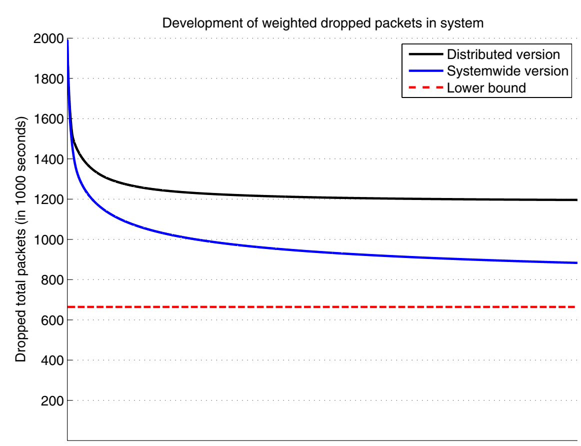


Fig. 7. *Evolution of $J_2(\boldsymbol{r})$ along trajectories of $\boldsymbol{r}$ that solve differential systems (15) and (19).*