Queueing Syst (2014) 77:211–242
DOI 10.1007/s11134-014-9408-8

Delay performance in random-access networks

**Niek Bouman · Sem C. Borst ·
Johan S. H. van Leeuwaarden**

Received: 15 May 2013 / Revised: 26 January 2014 / Published online: 24 April 2014

Abstract We explore the achievable delay performance in wireless random-access networks. While relatively simple and inherently distributed in nature, suitably designed queue-based random-access schemes provide the striking capability to match the optimal throughput performance of centralized scheduling mechanisms in a wide range of scenarios. The specific type of activation rules for which throughput optimality has been established, may however yield excessive queues and delays. Motivated by that issue, we examine whether the poor delay performance is inherent to the basic operation of these schemes, or caused by the specific kind of activation rules. We derive delay lower bounds for queue-based activation rules, which offer fundamental insight in the cause of the excessive delays. For fixed activation rates, we obtain lower bounds indicating that delays can grow dramatically with the load in certain topologies as well.

Keywords Delay performance · Random-access algorithms · Wireless networks

Mathematics Subject Classification 68M20 · 90B18 · 90B22

1 Introduction

Emerging wireless mesh networks typically lack any centralized access control entity, and instead vitally rely on the individual nodes to operate autonomously and fairly share the medium in a distributed fashion. A particularly popular mechanism for

N. Bouman (✉) · S. C. Borst · J. S. H. van Leeuwaarden
Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: n.bouman@tue.nl

S. C. Borst
e-mail: s.c.borst@tue.nl

J. S. H. van Leeuwaarden
e-mail: j.s.h.v.leeuwaarden@tue.nl

distributed medium access control is provided by the so-called carrier-sense multiple-access (CSMA) protocol. In the CSMA protocol, each node attempts to access the medium after a certain random back-off time, but nodes that sense activity of interfering nodes freeze their back-off timer until the medium is sensed idle.

While the CSMA protocol is fairly easy to understand at a local level, the interaction among interfering nodes gives rise to quite intricate behavior and complex throughput characteristics on a macroscopic scale. In recent years, relatively parsimonious models have emerged that provide a useful tool in evaluating the throughput characteristics of CSMA-like networks. These models were originally developed by Boorstyn et al. [1], and further pursued by Wang and Kar [2], Durvy et al. [3,4] and Garetto et al. [5]. Although the representation of the CSMA back-off mechanism in the above-mentioned models is less detailed than in the landmark work of Bianchi [6], they accommodate a general interference graph and thus cover a broad range of topologies. Experimental results of Liew et al. [7] demonstrate that these models, while idealized, provide throughput estimates that match remarkably well with measurements in actual real-life networks.

Despite their asynchronous and distributed nature, CSMA-like algorithms have been shown to offer the capability of achieving the full capacity region and thus match the optimal throughput performance of centralized scheduling mechanisms operating in slotted time, see for instance Jiang and Walrand [8], Liu et al. [9] and Tassiulas and Ephremides [10]. Based on this observation, various clever algorithms have been developed for finding the back-off rates that yield a particular target throughput vector or that optimize a certain concave throughput utility function in scenarios with saturated buffers, see for instance Jiang et al. [11], Jiang and Walrand [8] and Marbach and Eryilmaz [12].

In the same spirit, several powerful approaches have been devised for adapting the transmission periods based on the queue lengths in non-saturated scenarios, see for instance Rajagopalan et al. [13], Shah and Shin [14] and Shah et al. [15]. Roughly speaking, the latter algorithms provide maximum-stability guarantees under the condition that the transmission durations of the various nodes behave as logarithmic functions of the queue lengths.

Unfortunately, however, simulation experiments demonstrate that such activation rules can induce excessive queues and delays, which has sparked a strong interest in developing approaches for improving the delay performance, see for instance Ghaderi and Srikant [16], Lam et al. [17], Lotfinezhad and Marbach [18], Ni et al. [19] and Shah and Shin [20]. In particular, it has been shown that more aggressive schemes, where the transmission durations grow faster as function of the queue lengths, can reduce the delays, see for instance Bouman et al. [21].

In order to gain insight in the root cause for the poor delay performance, we establish in the present paper lower bounds for the average steady-state delay. To the best of our knowledge, the derivation of lower bounds for the average steady-state delay in random-access networks has received hardly any attention so far. An interesting paper by Shah et al. [22] showed that low-complexity schemes cannot be expected to achieve low delay in arbitrary topologies (unless P equals NP), since that would imply that certain NP-hard problems could be solved efficiently. However, the notion of delay in [22] is a transient one, and it is not exactly clear what the implications are for the average steady-state delay in specific networks, if any.

Jiang et al. [23] and Jiang and Walrand [24] derived *upper* bounds for the average steady-state delay based on mixing time results for Glauber dynamics, where the mixing time describes the amount of time required for the process to come close to its equilibrium distribution. The bounds show that for sufficiently low load the delay only grows polynomially with the number of nodes in bounded-degree interference graphs. Subramanian and Alanyali [25] presented similar upper bounds for bounded-degree interference graphs with low load based on analysis of neighbor sets and stochastic coupling arguments. While some of the conceptual notions in the present paper are similar (cliques, mixing times), we focus on *lower* rather than upper bounds, and exploit quite different techniques.

The lower bounds that we derive for queue-based activation schemes provide fundamental insight why the kind of rules that guarantee maximum stability yield excessive delays. We further obtain lower bounds for the delay in the case fixed back-off rates are used. In both cases, the bounds bring to light that the delay can grow dramatically with the load of the system. Specifically, we establish that the expected delay grows as $F(1/(1-\rho))$ as $\rho \uparrow 1$, where ρ is the load and $F(\cdot)$ is a superlinear function, implying that the growth rate may be polynomially or even exponentially faster than is typically the case in queueing systems at high load. The specific form and growth rate of the function $F(\cdot)$ depend on the activation rule as well as the topology of the network, as we will show for several scenarios of interest. The present paper is based on Bouman et al. [26], with various partial versions of the results having appeared earlier in [27,28].

The remainder of the paper is organized as follows. In Sect. 2, we present a detailed model description, followed by some preliminary results in Sect. 3. In Sect. 4, we derive delay lower bounds for queue-based activation schemes. We establish generic lower bounds for the delay in the case of fixed back-off rates in Sect. 5. In Sects. 6 and 7, we apply these generic bounds to a canonical class of partite interference graphs, which includes several specific cases of interest such as grid topologies. Simulation experiments are conducted in Sect. 8 to support the analytical results. In Sect. 9, we make some concluding remarks and identify topics for further research.

2 Model description

2.1 Network, interference graph, and traffic model

We consider a network of several nodes sharing a wireless medium according to a random-access mechanism. The network is represented by an undirected graph $G = (V, E)$ where the set of vertices $V = \{1, \dots, N\}$ corresponds to the various nodes and the set of edges $E \subseteq V \times V$ indicates which pairs of nodes interfere. Nodes that are neighbors in the interference graph are prevented from simultaneous activity, and thus the independent sets of G correspond to the feasible joint activity states of the network. A node is said to be blocked whenever the node itself or any of its neighbors is active, and unblocked otherwise. Define $\Omega \subseteq \{0, 1\}^N$ as the set of all feasible joint activity states of the network.

Packets arrive at node i as a Poisson process of rate λ_i. The packet transmission times of node i are generally distributed with mean β_i and second moment $\beta_i^{(2)}$. We assume that $\beta_i^{(2)} < \infty$ for all $i \in V$ and that transmission times are independent of each other and of the arrival process. Denote by $\rho_i = \lambda_i \beta_i$ the traffic intensity of node i.

Let $U(t) \in \Omega$ represent the joint activity state of the network at time t, with $U_i(t)$ indicating whether node i is active at time t or not. Denote by $L_i(t)$ the number of packets at node i at time t (including any packet that may be in the process of being transmitted).

2.2 Random-access mechanism

The nodes share the medium according to a random-access mechanism. When a node ends an activity period (consisting of possibly several back-to-back packet transmissions), it starts a back-off period. The back-off times of node i are independent and exponentially distributed with mean $1/\nu_i$. The back-off period of a node is suspended whenever it becomes blocked by activity of any of its neighbors, and is only resumed once the node becomes unblocked again. Thus, the back-off period of a node can only end when none of its neighbors are active. Now suppose a back-off period of node i ends at time t. Then the node starts a transmission with probability $\phi_i(L_i(t))$, and begins a next back-off period otherwise. Thus, when inactive and unblocked, node i starts transmitting at the instants of a time-inhomogeneous Poisson process of intensity $f_i(L_i(t))$ at time t, with $f_i(L_i(t)) = \nu_i \phi_i(L_i(t))$.

When a transmission of node i ends at time t, it releases the medium and begins a back-off period with probability $\psi_i(L_i(t^+))$, or starts the next transmission otherwise. We allow for $\phi_i(0) > 0$ and $\psi_i(0) < 1$, so a node may be active even when its buffer is empty, and transmit dummy packets.

For conciseness, the functions $f_i(\cdot)$ and $\psi_i(\cdot)$ will be referred to as activation and de-activation functions, respectively, and we define $h_i(\cdot) = f_i(\cdot)/\psi_i(\cdot)$ as the nominal activation function.

2.3 Stability

In general, it is difficult to establish under what conditions the system is stable. Denoting by conv$(\cdot)$ the convex hull operator and by int(conv$(\cdot)$) its interior, it is easily seen that $(\rho_1, \ldots, \rho_N) \in \text{int}(\text{conv}(\Omega))$ is a necessary condition for stability.

In [13,14,16], it is shown that in the case of exponentially distributed transmission times, this condition is in fact also sufficient for activation and de-activation functions $f_i(l) = r_i(l)/(1 + r_i(l))$ and $\psi_i(l) = 1/(1 + r_i(l))$ with suitably chosen $r_i(\cdot)$, e.g., $r_i(l) = \log(l + 1)$. For more aggressive queue-based activation functions, [29] shows that the necessary condition is not always sufficient though.

In the case of fixed activation and de-activation rates, a simple necessary and sufficient condition for stability is $\rho_i < \theta_i$, for all $i = 1, \ldots, N$, where θ_i denotes the stationary fraction of time that node i is active. Furthermore, there exists a unique vector $(\sigma_1, \ldots, \sigma_N)$ that yields $(\theta_1, \ldots, \theta_N) \in \text{int}(\text{conv}(\Omega))$

[8,30]. Hence, for any traffic intensity vector obeying the necessary stability condition, $(\rho_1, \ldots, \rho_N) \in \text{int}(\text{conv}(\Omega))$, there exists a vector $(\sigma_1, \ldots, \sigma_N)$ such that $(\rho_1, \ldots, \rho_N) < (\theta_1, \ldots, \theta_N) \in \text{int}(\text{conv}(\Omega))$, though determining the right vector $(\sigma_1, \ldots, \sigma_N)$ is non-trivial in general.

3 Preliminary results

In this section, we state some preliminary results in preparation for the derivation of delay lower bounds in the next sections. Throughout we assume that the system under consideration is stable, because otherwise such lower bounds are not particularly meaningful. More specifically, we derive lower bounds for the expected aggregate weighted stationary queue length in subsets of nodes $\mathcal{A} \subseteq V$. By virtue of Little's law, this also provides a lower bound for the expected aggregate weighted stationary delay. That is, $\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} \geq \alpha$ implies that $\sum_{i \in \mathcal{A}} w_i \lambda_i \mathbb{E}\{W_i\} \geq \alpha$, with W_i a random variable representing the delay (waiting time plus service time) of an arbitrary packet at node i.

The notion of a *clique* will play a pivotal role in the derivation of the lower bounds. A clique is a subset $\mathcal{C} \subseteq V$ of vertices in the interference graph G such that the subgraph induced by $\mathcal{C}$ is complete. Note that, in a clique, at most one node can be active at a time. The aggregate load in a clique should therefore be less than one if the system is to be stable. We use the notation $\rho_{\mathcal{C}} = \sum_{j \in \mathcal{C}} \rho_j$ for the aggregate load in a clique and we say that a clique $\mathcal{C}$ is in heavy traffic when $\rho_{\mathcal{C}}$ is close to one.

An immediate lower bound for the expected aggregate weighted queue length in any system follows from the workload decomposition property for M/G/1 systems with service interruptions [31,32]. Denote by $L_{i,\mathcal{C}}$ the number of packets at node i at an arbitrary epoch during a non-serving interval for the clique $\mathcal{C}$, i.e., a time interval during which none of the nodes in $\mathcal{C}$ is transmitting an actual packet. The workload decomposition property (applied to the total workload at the nodes in the clique $\mathcal{C}$) implies

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_i\} = \frac{\sum_{i \in \mathcal{C}} \lambda_i \beta_i^{(2)}}{2(1-\rho_{\mathcal{C}})} + \sum_{i \in \mathcal{C}} \rho_i \left(\beta_i - \frac{\beta_i^{(2)}}{2\beta_i} \right) + \sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_{i,\mathcal{C}}\} \tag{1}$$

$$= \frac{\rho_{\mathcal{C}} \sum_{i \in \mathcal{C}} \lambda_i \beta_i^{(2)}}{2(1-\rho_{\mathcal{C}})} + \sum_{i \in \mathcal{C}} \rho_i \beta_i + \sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_{i,\mathcal{C}}\} \tag{2}$$

$$= \frac{\rho_{\mathcal{C}} \sum_{i \in \mathcal{C}} \lambda_i \beta_i^{(2)}}{2(1-\rho_{\mathcal{C}})} + \sum_{i \in \mathcal{C}} \rho_i \beta_i. \tag{3}$$

Thus, as we would expect from elementary queueing theory, the total expected number of packets grows at least linearly in $1/(1-\rho)$ as ρ increases to one, with $\rho = \max_{\mathcal{C}} \rho_{\mathcal{C}}$ the maximum traffic intensity in any clique.

The lower bound in (3) is only based on sheer load considerations and does not account for the effect of the back-off mechanism. In the next sections, we will derive lower bounds for queue-based strategies as well as fixed-rate strategies that do capture the effect of the back-off mechanism, and turn out to be considerably tighter and exhibit superlinear growth in $1/(1-\rho_{\mathcal{C}})$.

The derivation of the lower bounds starts from the observation that stability of the system requires the non-serving intervals for a clique in heavy traffic to be short or happen infrequently. That is, in each clique, most of the time, one of the nodes should be active, since otherwise the average rate of arriving packets would exceed the average rate of departing packets. For this to be the case, the activity factors should be big at high load. The next lemma quantifies this statement. Denote by $\hat{L}_{i,\mathcal{C}}$ the number of packets at node i at an arbitrary epoch during an inactive interval for the clique $\mathcal{C}$, i.e., a time interval during which none of the nodes in $\mathcal{C}$ is active.

Lemma 1 *Assume that the system is stable. Then, for any clique $\mathcal{C} \subseteq V$ containing node i,*

$$\frac{\mathbb{E}\{f_i(\hat{L}_{i,\mathcal{C}})\}}{\mathbb{E}\{\psi_i(L_i)\}} \geq \frac{\lambda_i}{1-\rho_{\mathcal{C}}}. \tag{4}$$

Proof Observing that the mean number of activations at node i equals the mean number of de-activations at node i per unit of time, we obtain

$$\mathbb{E}\left\{f_i(L_i)\mathbf{1}_{\{U_j=0 \text{ for all } j\in\mathcal{N}_i^+\}}\right\} = \frac{\theta_i}{\beta_i}\mathbb{E}\{\psi_i(L_i^d)\}, \tag{5}$$

where $\mathcal{N}_i^+$ denotes the set of neighbors of node i in the graph G, along with i itself, and L_i^d denotes the number of packets waiting for transmission at node i at a departure epoch. Note that L_i^d is, in distribution, equal to L_i.

Further,

$$\begin{aligned}\mathbb{E}\{f_i(L_i)\mathbf{1}_{\{U_j=0 \text{ for all } j\in\mathcal{N}_i^+\}}\} &\leq \mathbb{E}\{f_i(L_i)\mathbf{1}_{\{U_j=0 \text{ for all } j\in\mathcal{C}\}}\} \\ &= \mathbb{E}\left\{f_i(\hat{L}_{i,\mathcal{C}})\right\}\mathbb{P}\{U_j=0 \text{ for all } j\in\mathcal{C}\}. \end{aligned} \tag{6}$$

Since the events $\{U_j = 1\}$ are mutually exclusive for all $j \in \mathcal{C}$, it follows that

$$\begin{aligned}\mathbb{P}\{U_j=0 \text{ for all } j\in\mathcal{C}\} &= 1-\mathbb{P}\{U_j=1 \text{ for some } j\in\mathcal{C}\} \\ &= 1-\sum_{j\in\mathcal{C}}\mathbb{P}\{U_j=1\} = 1-\sum_{j\in\mathcal{C}}\theta_j. \end{aligned} \tag{7}$$

Thus, combining (7) with (5) and (6), we find

$$\mathbb{E}\{f_i(\hat{L}_{i,\mathcal{C}})\}\left(1-\sum_{j\in\mathcal{C}}\theta_j\right) \geq \frac{\theta_i}{\beta_i}\mathbb{E}\{\psi_i(L_i)\},$$

and (4) follows from the fact that $\rho_i \leq \theta_i$ for all $i \in V$ is a necessary condition for stability of the system. □

Remark 1 For fixed-rate strategies, it follows that stability entails

$$\sigma_i \geq \frac{\rho_i}{1 - \rho_{\mathcal{C}}}. \tag{8}$$

Remark 2 If all neighbors of node i belong to the clique $\mathcal{C}$, then the inequality in (6) is an equality. If furthermore the activity functions are such that node i never transmits dummy packets, i.e., if $f_i(0) = 0$ and $\psi_i(0) = 1$, then the inequality in (4) is in fact an equality.

Lemma 1 shows that the activity factors in each clique should be big at high load. In the next sections, we will demonstrate that this also causes the delay to grow dramatically in heavy traffic.

3.1 Queue-based strategies

For queue-based strategies, we examine in Sect. 4 activation functions that are such that a node becomes increasingly more aggressive when the total number of packets at that node increases. For that natural class of activation functions, we exploit the result of Lemma 1 to find a lower bound of the form $h^{-1}(c/(1 - \rho_{\mathcal{C}}))$ for the total amount of work in the clique $\mathcal{C}$, where $h^{-1}(\cdot)$ is the inverse function of $h(\cdot)$ and c depends on the parameters of the system, but does not vanish as $\rho_{\mathcal{C}} \uparrow 1$.

A prominent example is $f(l) = r(l)/(1 + r(l))$ and $\psi(l) = 1/(1 + r(l))$ with $r(l) = \log(l + 1)$, so that $h^{-1}(l) = \exp(l) - 1$, the class of backlog-based strategies for which maximum stability is guaranteed as mentioned earlier. In this case, we find that the queue length scales at least exponentially in $1/(1 - \rho_{\mathcal{C}})$.

3.2 Fixed-rate strategies

In the case of fixed-rate strategies, the derivation of the delay lower bounds revolves around two simple observations: (i) high activation rates cause long mixing times, in particular slow transitions between dominant activity states; (ii) slow transitions between dominant states imply long starvation periods for some nodes, which cause huge queue lengths and delays. In Sect. 5, we formalize (ii), and establish lower bounds for the expected aggregate weighted queue length and delay in terms of the expected return times of the activity process $\{U(t)\}$.

In order to lower bound these return times, we will build in Sects. 6 and 7 on insight (i) for a canonical class of partite interference graphs. That is, we examine topologies where the nodes belong to one of K different components such that nodes in the same component do not interfere with each other and every node belongs to a clique of size K (of which the other $K - 1$ nodes necessarily belong to $K - 1$ different components). This class of K-partite interference graphs covers a wide range of network topologies with nearest-neighbor interference, e.g., linear topologies, ring networks with an even number of nodes, two-dimensional grid networks, tori (two-dimensional grid networks with a wrap-around boundary), and complete K-partite graphs, where all nodes are connected except those that belong to the same component, with star topologies as a prime example.

In Sect. 6, we prove that if the interference graph is a complete K-partite graph, then the expected queue length grows at least as fast as $1/(1-\rho)^{M-1}$, with M the size of the largest component. Based on observations (i) and (ii), this may be heuristically explained as follows. In order for the system to be stable, each node must at least have an activation rate of the order $1/(1-\rho)$, see Lemma 1. In turn, the transition times between the various activity states as governed by the maximum-size component occur on a time scale of the order ν^{M-1}, when each node has a fixed activation rate ν.

In Sect. 7, we extend the results of Sect. 6 to the broader class of K-partite interference graphs and show that the expected queue length grows at least as fast as $1/(1-\rho)^{M(1-H^*)}$ as ρ approaches 1. The coefficient H^* depends on the topology and is in general hard to calculate. We however know that $\frac{1}{M} \leq H^* \leq 1$ and for some specific topologies we can explicitly determine H^*.

Mathematically, the expected return time to a set equals the reciprocal of the bottleneck ratio, also known as conductance, of that set. We will therefore focus on deriving an upper bound for the bottleneck ratio instead. In the literature, the bottleneck ratio is often used to find bounds on the mixing time, either direct or via the spectral gap of the Markov chain, see e.g., [33]. Building on those results, one can derive lower bounds for the mixing time of the activity process $\{U(t)\}$ using the bounds on the bottleneck ratio we derive in Sects. 6 and 7. The heavy-traffic behavior of these lower bounds is similar to that of the corresponding delay lower bound.

4 Queue-based strategies

In this section, we derive delay lower bounds for queue-based strategies that use a concave activation function or a convex de-activation function.

For compactness, we use the notation $\beta_{\mathcal{C}} = \sum_{i\in\mathcal{C}} \beta_i$, $\beta_{\max,\mathcal{C}} = \max_{i\in\mathcal{C}} \beta_i$, and $\rho_{\min,\mathcal{C}} = \min_{i\in\mathcal{C}} \rho_i$.

Theorem 1 *Assume that the system is stable. Then, for any clique $\mathcal{C} \subseteq V$,*

(i) *If $f_i(\cdot) \equiv f(\cdot)$, with $f(0) = 0$, for $i \in \mathcal{C}$ is an increasing concave function and $\psi_i(\cdot) \geq \xi_i > 0$ with $\psi_i(0) = 1$ for $i \in \mathcal{C}$, then*

$$\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\} \geq \frac{\rho_{\mathcal{C}} \sum_{i\in\mathcal{C}} \lambda_i \beta_i^{(2)}}{2(1-\rho_{\mathcal{C}})} + \sum_{i\in\mathcal{C}} \rho_i \beta_i + \beta_{\mathcal{C}} f^{-1}\left(\frac{1}{\beta_{\mathcal{C}}} \frac{\sum_{i\in\mathcal{C}} \rho_i \xi_i}{1-\rho_{\mathcal{C}}}\right). \tag{9}$$

(ii) *If $f_i(\cdot) \leq \xi_i$ for $i \in \mathcal{C}$ and $\psi_i(\cdot) \equiv \psi(\cdot)$ for $i \in \mathcal{C}$ is a decreasing convex function, then*

$$\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\} \geq \beta_{\mathcal{C}} \psi^{-1}\left(\frac{1-\rho_{\mathcal{C}}}{\beta_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \xi_i \beta_i / \lambda_i\right). \tag{10}$$

(iii) *If $f_i(\cdot) \equiv f(\cdot)$, with $f(0) = 0$, for $i \in \mathcal{C}$ is an increasing concave function and $\psi_i(\cdot) \equiv \psi(\cdot)$ for $i \in \mathcal{C}$, with $\psi(0) = 1$, is a decreasing convex function, then*

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_i\} \geq \beta_{\mathcal{C}} h^{-1}\left(\frac{\rho_{\min,\mathcal{C}}}{\beta_{\max,\mathcal{C}}(1-\rho_{\mathcal{C}})}\right), \tag{11}$$

with $h(\cdot) = f(\cdot)/\psi(\cdot)$.

Proof In case (i), the fact that $f(0) = 0$ and $\psi_i(0) = 1$ for $i \in \mathcal{C}$ precludes the transmission of dummy packets by any of the nodes $i \in \mathcal{C}$. In other words, a node $i \in \mathcal{C}$ must be transmitting an actual packet whenever it is active, so that $L_{i,\mathcal{C}}$ and $\hat{L}_{i,\mathcal{C}}$ are identical in distribution. From (4), we thus get that

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \geq \frac{\sum_{i \in \mathcal{C}} \rho_i \xi_i}{1-\rho_{\mathcal{C}}}. \tag{12}$$

Since $f(\cdot)$ is concave, it follows from the probabilistic form of Jensen's inequality that

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \leq \sum_{i \in \mathcal{C}} \beta_i f(\mathbb{E}\{L_{i,\mathcal{C}}\}).$$

Now applying the finite form of Jensen's inequality gives

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \leq \beta_{\mathcal{C}} f\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_{i,\mathcal{C}}\}\right). \tag{13}$$

Because $f(\cdot)$ is increasing, we thus get

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_{i,\mathcal{C}}\} \geq \beta_{\mathcal{C}} f^{-1}\left(\frac{1}{\beta_{\mathcal{C}}} \frac{\sum_{i \in \mathcal{C}} \rho_i \xi_i}{1-\rho_{\mathcal{C}}}\right),$$

and the assertion for case (i) follows from (2).

The proof for case (ii) proceeds along similar lines. From (4), we obtain

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{\psi(L_i)\} \leq (1-\rho_{\mathcal{C}}) \sum_{i \in \mathcal{C}} \xi_i \beta_i / \lambda_i.$$

Since $\psi(\cdot)$ is convex, it follows from Jensen's inequality that

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{\psi(L_i)\} \geq \beta_{\mathcal{C}} \psi\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_i\}\right). \tag{14}$$

Since $\psi(\cdot)$ is also decreasing, we thus get

$$\sum_{i \in \mathcal{C}} \beta_i \mathbb{E}\{L_i\} \geq \beta_{\mathcal{C}} \psi^{-1}\left(\frac{1-\rho_{\mathcal{C}}}{\beta_{\mathcal{C}}} \sum_{i \in \mathcal{C}} \xi_i \beta_i / \lambda_i\right),$$

yielding (10).

In case (iii), the fact that $f(0) = 0$ and $\psi_i(0) = 1$ for $i \in \mathcal{C}$ again means that $L_{i,\mathcal{C}}$ and $\hat{L}_{i,\mathcal{C}}$ are identical in distribution for all the nodes $i \in \mathcal{C}$. Combining (2) and (13) then gives

$$\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \leq \beta_{\mathcal{C}} f\left(\frac{1}{\beta_{\mathcal{C}}}\left(\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\} - \frac{\rho_{\mathcal{C}} \sum_{i\in\mathcal{C}} \lambda_i \beta_i^{(2)}}{2(1-\rho_{\mathcal{C}})} - \sum_{i\in\mathcal{C}} \rho_i \beta_i\right)\right),$$

and hence, because $f(\cdot)$ is increasing,

$$\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \leq \beta_{\mathcal{C}} f\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\}\right). \tag{15}$$

From (4), we further find

$$\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \geq \frac{1}{1-\rho_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \rho_i \mathbb{E}\{\psi(L_i)\} \geq \frac{\rho_{\min,\mathcal{C}}}{\beta_{\max,\mathcal{C}}(1-\rho_{\mathcal{C}})} \sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{\psi(L_i)\},$$

and, using (14),

$$\sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{f(L_{i,\mathcal{C}})\} \geq \frac{\rho_{\min,\mathcal{C}}}{\beta_{\max,\mathcal{C}}(1-\rho_{\mathcal{C}})} \beta_{\mathcal{C}} \psi\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\}\right).$$

We thus find

$$\beta_{\mathcal{C}} f\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\}\right) \geq \frac{\rho_{\min,\mathcal{C}}}{\beta_{\max,\mathcal{C}}(1-\rho_{\mathcal{C}})} \beta_{\mathcal{C}} \psi\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\}\right),$$

or equivalently,

$$h\left(\frac{1}{\beta_{\mathcal{C}}} \sum_{i\in\mathcal{C}} \beta_i \mathbb{E}\{L_i\}\right) \geq \frac{\rho_{\min,\mathcal{C}}}{\beta_{\max,\mathcal{C}}(1-\rho_{\mathcal{C}})}.$$

Thus, as $h(\cdot) = f(\cdot)/\psi(\cdot)$ is increasing because $f(\cdot)$ is increasing and $\psi(\cdot)$ is decreasing, we get (11). □

Remark 3 For complete conflict graphs, i.e., when $E = V \times V$, the inequality in (12) is an equality when $\mathcal{C} = V$ and $\psi_i(0) = 1$ for all $i \in V$, see Remark 2.

In particular, when $\psi_i(l) = 1$ and $f_i(l) = \nu l$, this gives an exact result for the expected stationary amount of work in the system, as $f_i(\cdot)$ is both convex and concave and the upper and lower bounds are equal. Further, when the transmission times are identically distributed, the system behaves exactly like a single-node system with arrival rate $\lambda = \sum_{i=1}^{N} \lambda_i$, for which exact delay results have been derived in [21].

For identical mean transmission times, i.e., $\beta_i \equiv \beta$ for all $i \in \mathcal{C}$, the results in Theorem 1 provide a lower bound for the expected aggregate number of packets in a clique. In addition, a slightly different lower bound can be obtained for case (iii) if the mean transmission times are identical. Denoting $\lambda_{\mathcal{C}} = \sum_{i \in \mathcal{C}} \lambda_i$ and $\lambda_{\min,\mathcal{C}} = \min_{i \in \mathcal{C}} \lambda_i$, the results for this case are presented in the following corollary.

Corollary 1 *Assume that the system is stable. Then, for any clique $\mathcal{C} \subseteq V$ with $\beta_i \equiv \beta$ for all $i \in \mathcal{C}$,*

(i) If $f_i(\cdot) \equiv f(\cdot)$, with $f(0) = 0$, for $i \in \mathcal{C}$ is an increasing concave function and $\psi_i(\cdot) \geq \xi_i > 0$ with $\psi_i(0) = 1$ for $i \in \mathcal{C}$, then

$$\sum_{i \in \mathcal{C}} \mathbb{E}\{L_i\} \geq \frac{\lambda_{\mathcal{C}} \sum_{i \in \mathcal{C}} \lambda_i \beta_i^{(2)}}{2(1 - \rho_{\mathcal{C}})} + \rho_{\mathcal{C}} + |\mathcal{C}| f^{-1}\left(\frac{1}{|\mathcal{C}|} \frac{\sum_{i \in \mathcal{C}} \lambda_i \xi_i}{1 - \rho_{\mathcal{C}}}\right). \tag{16}$$

(ii) If $f_i(\cdot) \leq \xi_i$ for $i \in \mathcal{C}$ and $\psi_i(\cdot) \equiv \psi(\cdot)$ for $i \in \mathcal{C}$ is a decreasing convex function, then

$$\sum_{i \in \mathcal{C}} \mathbb{E}\{L_i\} \geq |\mathcal{C}| \psi^{-1}\left(\frac{1 - \rho_{\mathcal{C}}}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \xi_i / \lambda_i\right). \tag{17}$$

(iii) If $f_i(\cdot) \equiv f(\cdot)$, with $f(0) = 0$, for $i \in \mathcal{C}$ is an increasing concave function and $\psi_i(\cdot) \equiv \psi(\cdot)$ for $i \in \mathcal{C}$ with $\psi(0) = 1$ is a decreasing convex function, then

$$\sum_{i \in \mathcal{C}} \mathbb{E}\{L_i\} \geq \max\left\{|\mathcal{C}| h^{-1}\left(\frac{\lambda_{\min,\mathcal{C}}}{1 - \rho_{\mathcal{C}}}\right), h^{-1}\left(\frac{1}{|\mathcal{C}|} \frac{\lambda_{\mathcal{C}}}{1 - \rho_{\mathcal{C}}}\right)\right\}, \tag{18}$$

with $h(\cdot) = f(\cdot)/\psi(\cdot)$.

Proof For identical mean transmission times, $\beta_i \equiv \beta$ for all $i \in \mathcal{C}$, (16) and (17) follow immediately from (9) and (10), respectively. Also, for case (iii), it follows immediately from (11) that

$$\sum_{i \in \mathcal{C}} \mathbb{E}\{L_i\} \geq |\mathcal{C}| h^{-1}\left(\frac{\lambda_{\min,\mathcal{C}}}{1 - \rho_{\mathcal{C}}}\right). \tag{19}$$

Further note that (15) gives

$$\sum_{i \in \mathcal{C}} \mathbb{E}\{f(L_{i,\mathcal{C}})\} \leq |\mathcal{C}| f\left(\frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \mathbb{E}\{L_i\}\right),$$

when the mean transmission times are identical, and, hence, because $f(\cdot)$ is increasing and $|\mathcal{C}| \geq 1$,

$$\sum_{i \in \mathcal{C}} \mathbb{E}\{f(L_{i,\mathcal{C}})\} \leq |\mathcal{C}| f\left(\sum_{i \in \mathcal{C}} \mathbb{E}\{L_i\}\right).$$

Further, using (4) and applying Jensen's inequality,

$$\sum_{i\in\mathcal{C}}\mathbb{E}\{f(L_{i,\mathcal{C}})\}\geq\frac{1}{1-\rho_{\mathcal{C}}}\sum_{i\in\mathcal{C}}\lambda_i\mathbb{E}\{\psi(L_i)\}\geq\frac{\lambda_{\mathcal{C}}}{1-\rho_{\mathcal{C}}}\psi\left(\frac{1}{\lambda_{\mathcal{C}}}\sum_{i\in\mathcal{C}}\lambda_i\mathbb{E}\{L_i\}\right).$$

Thus, as $\lambda_i\leq\lambda_{\mathcal{C}}$ for $i\in\mathcal{C}$ and because $\psi(\cdot)$ is decreasing we obtain

$$\sum_{i\in\mathcal{C}}\mathbb{E}\{f(L_{i,\mathcal{C}})\}\geq\frac{\lambda_{\mathcal{C}}}{1-\rho_{\mathcal{C}}}\psi\left(\sum_{i\in\mathcal{C}}\mathbb{E}\{L_i\}\right),$$

yielding

$$h\left(\sum_{i\in\mathcal{C}}\mathbb{E}\{L_i\}\right)\geq\frac{1}{|\mathcal{C}|}\frac{\lambda_{\mathcal{C}}}{1-\rho_{\mathcal{C}}}.$$

Again noting that $h(\cdot)=f(\cdot)/\psi(\cdot)$ is increasing because $f(\cdot)$ is increasing and $\psi(\cdot)$ is decreasing, we get

$$\sum_{i\in\mathcal{C}}\mathbb{E}\{L_i\}\geq h^{-1}\left(\frac{1}{|\mathcal{C}|}\frac{\lambda_{\mathcal{C}}}{1-\rho_{\mathcal{C}}}\right),$$

and (18) follows by combining this with (19). □

The three cases covered in Theorem 1 and Corollary 1 all reveal the same effect, namely that the mean number of packets in a clique is at least of the order of $h^{-1}(c/(1-\rho_{\mathcal{C}}))$, where $h^{-1}(\cdot)$ is the inverse function of $h(\cdot)$ and c depends on the parameters of the system. In case (ii), this effect is observed because the argument of $\psi^{-1}(\cdot)$ is reciprocal. Further, noting that $f(l)=\log(l+1)/(1+\log(l+1))$ is an increasing concave function and $\psi(l)=1/(1+\log(l+1))$ is a decreasing convex function, we have, with $\beta_i=1$ for all $i\in\mathcal{C}$,

$$\sum_{i\in\mathcal{C}}\mathbb{E}\{L_i\}\geq\max\left\{|\mathcal{C}|\left(\operatorname{Exp}\left(\frac{\rho_{\min,\mathcal{C}}}{1-\rho_{\mathcal{C}}}\right)-1\right),\operatorname{Exp}\left(\frac{\rho_{\mathcal{C}}}{|\mathcal{C}|(1-\rho_{\mathcal{C}})}\right)-1\right\}$$

for the class of functions for which maximum stability is guaranteed.

The results of Theorem 1 and Corollary 1 suggest that in order to improve the delay performance one should use more aggressive access schemes. In fact, if $h(\cdot)$ is a superlinear function, i.e., if $h(\cdot)$ grows faster than linear, we find a lower bound that is loose in heavy traffic and (3) provides a better lower bound in that case. Remember however that maximum stability is not guaranteed in the case that a superlinear function $h(\cdot)$ is used, hence the delay performance might actually deteriorate, and even instability could occur as shown in [29].

5 Fixed-rate strategies

In the previous section, we derived delay bounds for queue-based activation rules and we saw that the type of activation rules for which throughput optimality has been established yield excessive delays and queues. We now proceed to construct lower bounds for the expected aggregate weighted queue length and delay in the case of fixed activation and de-activation rates, i.e., we take $\phi_i(\cdot) \equiv \phi_i$ and $\psi_i(\cdot) \equiv \psi_i$. Note that, the activity process $\{U(t)\}$ then does not depend on the process $\{L(t)\}$. We further assume that packet transmission times at node i are exponentially distributed with mean $\beta_i = 1/\mu_i$. Then, $\{U(t)\}$ in fact evolves as a reversible Markov process with state space Ω. Transitions (due to activations) from a state u with $u_i = 0$ and $u_j = 0$ for all neighbors j of node i to $u + e_i$ occur at rate $\nu_i \phi_i$ and transitions (due to transmission completions followed by a back-off period) from a state u with $u_i = 1$ to $u - e_i$ occur at rate $\mu_i \psi_i$.

For any $u \in \Omega$, define $\pi(u) = \lim_{t\to\infty} \mathbb{P}\{U(t) = u\}$ as the steady-state probability that the activity process $\{U(t)\}$ resides in state u. The steady-state probabilities have a product-form distribution [1]

$$\pi(u) = Z^{-1} \prod_{i=1}^{N} \sigma_i^{u_i}, \qquad u \in \Omega, \tag{20}$$

with $\sigma_i = \nu_i \phi_i / (\mu_i \psi_i)$ representing a nominal activity factor, and

$$Z = \sum_{u \in \Omega} \prod_{i=1}^{N} \sigma_i^{u_i}$$

denoting a normalization constant.

Further, denote $S^c = \Omega \backslash S$, and define

$$\partial S = \left\{ u \in S : \sum_{u' \in S^c} q(u', u) > 0 \right\} = \left\{ u \in S : \sum_{u' \in S^c} q(u, u') > 0 \right\}$$

as the boundary of the subset S, with $q(u, u')$ denoting the transition rate from state u to state u' of the activity process $\{U(t)\}$ as specified above, i.e., $q(u, u + e_i) = \nu_i \phi_i$ and $q(u + e_i, u) = \mu_i \psi_i$, $u, u + e_i \in \Omega$, and denote the transition rate out of the subset S by

$$Q(S) = \sum_{u \in S} \sum_{u' \in S^c} \pi(u) q(u, u') = \sum_{u \in S^c} \sum_{u' \in S} \pi(u) q(u, u') = \sum_{u \in \partial S} \sum_{u' \in S^c} \pi(u) q(u, u').$$

With minor abuse of notation, denote by $\pi(S) = \sum_{u \in S} \pi(u)$ the fraction of time that the process $\{U(t)\}$ spends in one of the activity states in the subset S. The bottleneck ratio of the subset S is defined as

$$\Phi(S) = \frac{Q(S)}{\pi(S)}.$$

Further define for arbitrary weights $w \in \mathbb{R}_+^N$ and for any $\mathcal{A} \subseteq V$, $S \subseteq \Omega$,

$$Y(w, \mathcal{A}, S) = \max_{u \in S} \sum_{i \in \mathcal{A}} w_i \mu_i u_i,$$

and denote

$$D(w, \mathcal{A}, S) = \sum_{i \in \mathcal{A}} w_i \lambda_i - Y(w, \mathcal{A}, S).$$

The coefficient $Y(w, \mathcal{A}, S)$ represents the maximum aggregate weighted service rate of the nodes in $\mathcal{A}$ when the system resides in one of the activity states in the subset S. Noting that $\sum_{i \in \mathcal{A}} w_i \lambda_i$ is the aggregate weighted arrival rate of the nodes in $\mathcal{A}$, the coefficient $D(w, \mathcal{A}, S)$ may thus be interpreted as the minimum drift in the aggregate weighted queue length of the nodes in $\mathcal{A}$ when the system resides in one of the activity states in the subset S. For compactness, denote $w_{\mathcal{A},S} = \sum_{i \in \mathcal{A}} w_i \max_{u \in S} u_i$.

Proposition 1 *For any $w \in \mathbb{R}_+^N$, $\mathcal{A} \subseteq V$,*

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} \geq \frac{1}{2} \max_{S \subseteq \Omega} \pi(S) \left(D(w, \mathcal{A}, S) \frac{1}{\Phi(S)} - w_{\mathcal{A},S} \right). \tag{21}$$

Proof Consider the process $X(t)$ with $X(0) = 0$ and $X(t) = 0$ for all t such that $U(t) \notin S$. Further, $X(t)$ behaves identically to $L(t)$ when $U(t) \in S$, i.e., $X_i(t) - X_i(s) = L_i(t) - L_i(s)$ for all i when $U(\tau) \in S$ for all $\tau \in [s, t]$. Thus, by construction, $L(t) \geq X(t)$ for all t, and we have

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \sum_{i \in \mathcal{A}} w_i L_i(t) \mathrm{d}t \geq \lim_{T \to \infty} \frac{1}{T} \int_0^T \sum_{i \in \mathcal{A}} w_i X_i(t) \mathrm{d}t. \tag{22}$$

Now assume $U(0) = u_0$ for some $u_0 \in \partial S$ and note that

$$T_{u_0} = \inf\{t > 0 : U(t) = u_0, U(t^-) \notin S\}$$

is a regeneration epoch for $X(t)$. Denote by $X^u(t)$ the process started at $U(0) = u$. The renewal-reward theorem, see e.g., [34], gives

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \sum_{i \in \mathcal{A}} w_i X_i^{u_0}(t) \mathrm{d}t = \frac{\mathbb{E}\left\{ \int_0^{T_{u_0}} \sum_{i \in \mathcal{A}} w_i X_i^{u_0}(t) \mathrm{d}t \right\}}{\mathbb{E}\{T_{u_0}\}}. \tag{23}$$

Now, let $T_S^{u,k}$ be the duration of the k-th visit to S with starting state $u \in \partial S$, and let K_u be a random variable representing the number of visits to S with starting state u during $[0, T_{u_0}]$. Now, observe that the random variables $T_S^{u,k}$, $k = 1, \dots, K_u$, are independent and identically distributed copies of a random variable T_S^u. Because T_{u_0} is a stopping time, Wald's lemma gives

$$\mathbb{E}\left\{\int_0^{T_{u_0}} \sum_{i\in\mathcal{A}} w_i X_i^{u_0}(t)\mathrm{d}t\right\} = \sum_{u\in\partial S} \mathbb{E}\{K_u\}\mathbb{E}\left\{\int_0^{T_S^u} \sum_{i\in\mathcal{A}} w_i X_i^{u}(t)\mathrm{d}t\right\}. \tag{24}$$

Now, define $K = \sum_{u\in\partial S} K_u$ and $p_u = \mathbb{E}\{K_u\}/\mathbb{E}\{K\}$, the probability that a visit to S starts in state u. Also, let T_S be the length of a visit to S in the stationary regime, and let $X^s(t)$ be the process $X(t)$ started at $U(0) = u$ with probability p_u, $u \in \partial S$, then (24) may be written as

$$\mathbb{E}\left\{\int_0^{T_{u_0}} \sum_{i\in\mathcal{A}} w_i X_i^{u_0}(t)\mathrm{d}t\right\} = \mathbb{E}\{K\}\mathbb{E}\left\{\int_0^{T_S} \sum_{i\in\mathcal{A}} w_i X_i^{s}(t)\mathrm{d}t\right\}. \tag{25}$$

Thus, $X^s(t)$, $t \in [0, T_S]$, describes the evolution of the process $X(t)$ during a visit to the set S in the stationary regime. Let $(x_0, y_0) \in S^c \times S$ be the transition that caused the entrance into S at $t = 0$, $(x_b, y_b) \in S^2$ the b-th transition during the visit occurring at time epoch $t_b \in (0, T_S)$, $b = 1, \dots, B$, and $(x_{B+1}, y_{B+1}) \in S \times S^c$ the transition that triggered the exit from S.

Denote by $X_i^{\uparrow}(t)$ the number of service completions at node i by time $t \in [0, T_S]$, not counting the service completion that may have occurred at time 0 and caused the entrance into S. Define $T_i(t) = \int_{s=0}^{t} U_i(s)\mathrm{d}s$ as the cumulative amount of time that node i has been active by time $t \in [0, T_S]$.

Denote by u_l and v_l the time epochs of the j-th service initiation and j-th service completion at node i during the visit, $j = 1, \dots, J$, $J = X_i^{\uparrow}(T_S) + U_i(T_S)$, with the convention that $u_1 = 0$ and $v_J = T_S$ if $U_i(0) = 1$ and $U_i(T_S) = 1$, respectively. Then

$$\int_{t=0}^{T_S} X_i^{\uparrow}(t)\mathrm{d}t = \sum_{j=1}^{J} (T_S - v_j). \tag{26}$$

Since the activity process is reversible, it would have been equally likely to observe an entrance into S at time 0 through a transition $(y_{B+1}, x_{B+1}) \in S^c \times S$, a transition $(y_b, x_b) \in S^2$ during the visit at time epoch $T_S - t_b$, $b = 1, \dots, B$, and then an exit from S through a transition $(y_0, x_0) \in S \times S^c$ at time T_S. In particular, it would be equally likely for service completions at node i to occur at time epochs $T_S - u_j$, $j = 1, \dots, J$. In that case,

$$\int_{t=0}^{T_S} X_i^{\uparrow}(t)\mathrm{d}t = \sum_{j=1}^{J} u_l. \tag{27}$$

Adding (26) and (27) yields

$$\begin{aligned}\sum_{j=1}^{J}(T_S - v_j) + \sum_{j=1}^{J} u_j = \sum_{j=1}^{J}(T_S - (v_j - u_j)) &= JT_S + \sum_{j=1}^{J}(v_j - u_j)\\ &= (X_i^{\uparrow}(T_S) + U_i(T_S))T_S - T_i(T_S).\end{aligned}$$

Thus, we obtain

$$\mathbb{E}\left\{\int_{t=0}^{T_S} X_i^{\uparrow}(t)\mathrm{d}t\right\} = \frac{1}{2}\left[\mathbb{E}\{X_i^{\uparrow}(T_S)T_S\} + \mathbb{E}\{U_i(T_S)T_S\} - \mathbb{E}\{T_i(T_S)\}\right].$$

Now, observe that $M(t) = t(X_i^{\uparrow}(t) - \mu_i T_i(t))$ is a martingale with respect to the natural filtration. Since T_S has bounded expectation, Doob's stopping time theorem gives

$$\mathbb{E}\{X_i^{\uparrow}(T_S)T_S\} = \mu_i \mathbb{E}\{T_i(T_S)T_S\}.$$

Since $\mathbb{E}\{U_i(T_S)T_S\} \leq \mathbb{E}\{T_S\}\max_{u\in S} u_i$ and $\mathbb{E}\{T_i(T_S)\} \geq 0$, it follows that

$$\mathbb{E}\left\{\int_{t=0}^{T_S} X_i^{\uparrow}(t)\mathrm{d}t\right\} \leq \frac{1}{2}\left[\mu_i \mathbb{E}\{T_i(T_S)T_S\} + \mathbb{E}\{T_S\}\max_{u\in S} u_i\right].$$

We thus obtain from (25) that

$$\begin{aligned}&\mathbb{E}\left\{\int_{0}^{T_{u_0}} \sum_{i\in\mathcal{A}} w_i X_i^{u_0}(t)\mathrm{d}t\right\} = \mathbb{E}\{K\}\sum_{i\in\mathcal{A}}\mathbb{E}\{\int_0^{T_S} w_i X_i^{s}(t)\mathrm{d}t\}\\
&= \mathbb{E}\{K\}\left(\sum_{i\in\mathcal{A}}\mathbb{E}\left\{\int_0^{T_S} w_i\left[\lambda_i t - X_i^{\uparrow}(t)\mathrm{d}t\right]\right\}\right)\\
&\geq \mathbb{E}\{K\}\left(\sum_{i\in\mathcal{A}}\left[\int_0^{T_S} w_i\lambda_i t\mathrm{d}t - \frac{1}{2}\mu_i w_i \mathbb{E}\{T_i(S)T_S\}\right] - \frac{1}{2}w_{\mathcal{A},S}\mathbb{E}\{T_S\}\right)\\
&= \mathbb{E}\{K\}\left(\sum_{i\in\mathcal{A}}\left[w_i\lambda_i\frac{1}{2}\mathbb{E}\{T_S^2\} - \frac{1}{2}\mu_i w_i \mathbb{E}\left\{T_S\int_0^{T_S} U_i(t)\mathrm{d}t\right\}\right] - \frac{1}{2}w_{\mathcal{A},S}\mathbb{E}\{T_S\}\right)\end{aligned}$$

$$\geq \mathbb{E}\{K\} \left(\sum_{i\in\mathcal{A}} w_i \lambda_i \frac{1}{2}\mathbb{E}\{T_S^2\} - \frac{1}{2}\mathbb{E}\left\{ T_S \int_0^{T_S} Y(w, \mathcal{A}, S)\mathrm{d}t \right\} - \frac{1}{2} w_{\mathcal{A},S}\mathbb{E}\{T_S\} \right)$$

$$= \mathbb{E}\{K\} \left(\sum_{i\in\mathcal{A}} w_i \lambda_i \frac{1}{2}\mathbb{E}\{T_S^2\} - \frac{1}{2} Y(w, \mathcal{A}, S)\mathbb{E}\{T_S^2\} - \frac{1}{2} w_{\mathcal{A},S}\mathbb{E}\{T_S\} \right)$$

$$= \mathbb{E}\{K\} \left(\frac{1}{2} D(w, \mathcal{A}, S)\mathbb{E}\{T_S^2\} - \frac{1}{2} w_{\mathcal{A},S}\mathbb{E}\{T_S\} \right).$$

Thus, because

$$\begin{aligned}\mathbb{E}\{T_{u_0}\} = \mathbb{E}\{\sum_{k=1}^{K_u} \sum_{u\in\partial S} T_S^{u,k} + T_{S^c}^{u,k}\} &= \sum_{u\in\partial S} \mathbb{E}\{K_u\} \left(\mathbb{E}\{T_S^u\} + \mathbb{E}\{T_{S^c}^u\}\right) \\ &= \sum_{u\in\partial S} \mathbb{E}\{K\} p_u \left(\mathbb{E}\{T_S^u\} + \mathbb{E}\{T_{S^c}^u\}\right) \\ &= \mathbb{E}\{K\}\big(\mathbb{E}\{T_S\} + \mathbb{E}\{T_{S^c}\}\big),\end{aligned}$$

we find from (22) and (23),

$$\sum_{i\in\mathcal{A}} w_i \mathbb{E}\{L_i\} \geq \frac{\frac{1}{2} D(w, \mathcal{A}, S)\mathbb{E}\{T_S^2\} - \frac{1}{2} w_{\mathcal{A},S}\mathbb{E}\{T_S\}}{\mathbb{E}\{T_S\} + \mathbb{E}\{T_{S^c}\}}.$$

Now, since $\mathbb{E}\{T_{S^c}\} = \frac{1-\pi(S)}{\pi(S)} T_S$, we get

$$\begin{aligned}\sum_{i\in\mathcal{A}} w_i \mathbb{E}\{L_i\} &\geq \frac{1}{2}\pi(S) \left(D(w, \mathcal{A}, S)\frac{\mathbb{E}\{T_S^2\}}{\mathbb{E}\{T_S\}} - w_{\mathcal{A},S} \right) \\ &\geq \frac{1}{2}\pi(S) \left(D(w, \mathcal{A}, S)\mathbb{E}\{T_S\} - w_{\mathcal{A},S} \right).\end{aligned}$$

Finally, because $Q(S)$ is the expected number of times the activity process $\{U(t)\}$ enters S per unit of time and $\mathbb{E}\{T_S\}$ is the expected amount of time the process stays in S after entering, the expected fraction of the time the process resides in S, $\pi(S)$, is given by $\pi(S) = Q(S)\mathbb{E}\{T_S\}$. Thus, $\mathbb{E}\{T_S\} = \frac{1}{\Phi(S)}$, and (21) follows. □

The question arises how to choose S such that the maximum and thus the tightest possible lower bound in (21) is obtained. Evidently, the more S includes states with some of the nodes in $\mathcal{A}$ active, the larger the potential aggregate weighted service rate of the nodes in $\mathcal{A}$, i.e., the larger $Y(w, \mathcal{A}, S)$, and the smaller $D(w, \mathcal{A}, S)$. In other words, we need to ensure that S excludes some of the states with nodes in $\mathcal{A}$ active. Indeed, if S includes all states with maximal subsets of the nodes in $\mathcal{A}$ active, then $Y(w, \mathcal{A}, S) = \max_{u\in\Omega} \sum_{i=1}^N \hat{w}_i \mu_i u_i$, with $\hat{w}_i = w_i$ if $i \in \mathcal{A}$ and $\hat{w}_i = 0$ otherwise. The fact that $(\rho_1, \ldots, \rho_N) \in \mathrm{int}(\mathrm{conv}(\Omega))$ then implies that $Y(w, \mathcal{A}, S) \geq \sum_{i=1}^N \hat{w}_i \mu_i \rho_i = \sum_{i=1}^N \hat{w}_i \lambda_i = \sum_{i\in\mathcal{A}} w_i \lambda_i$, so that $D(w, \mathcal{A}, S) \leq 0$, yielding an irrelevant lower bound. However, observe that the expected equilibrium return time to S^c, denoted $\mathbb{E}\{T_S\}$, may be small when S includes very few states. Hence, to obtain the sharpest possible lower bound, it may not necessarily be optimal to exclude all the states with nodes in $\mathcal{A}$ active from S. For high values of ν, which are necessary for stability at high load as Lemma 1 showed, the above argument suggests that we should choose S so that it contains a state with many active nodes, while the boundary of S only contains states with few active nodes. Denoting $K(S, \mathcal{A}) = \max_{u \in S} \sum_{i \in \mathcal{A}} u_i$, we thus need to find a subset S such that $K(S, V)$ is large, $K(\partial S, V)$ is small, and $K(S, \mathcal{A})$ is small, in order to get a tight lower bound in (21).

We will now first give an example to illustrate the use of Proposition 1.

Example 1 Suppose that S is such that $u + e_i \notin S^c$ for all $u \in \partial S$. In the case that $\phi_i \equiv 1$, $\psi_i \equiv 1$, $\mu_i \equiv 1$, and $\nu_i \equiv \nu \geq 1$, we then have $Q(S) \leq N\pi(\partial S)$, and thus using (20),

$$\frac{1}{\Phi(S)} = \frac{\pi(S)}{Q(S)} \geq \frac{\sum\limits_{u \in S} \pi(u)}{N \sum\limits_{u \in \partial S} \pi(u)} = \frac{\sum\limits_{u \in S} \nu^{\sum_{i=1}^N u_i}}{N \sum\limits_{u \in \partial S} \nu^{\sum_{i=1}^N u_i}} \geq \frac{1}{N} \nu^{K(S,V) - K(\partial S, V)}.$$

We thus see that in this example we indeed need to choose S such that $K(S, V) - K(\partial S, V)$ is maximized.

Now, suppose the interference graph is a symmetric complete bipartite graph. That is, the nodes in $V_1 = \{1, \ldots, N/2\}$ interfere with, and only with, the nodes in $V_2 = \{N/2 + 1, \ldots, N\}$. In this case, we have $K(S, V) \leq N/2$. Further, as S is such that $u + e_i \notin S^c$, we have $S = \Omega$ if and only if $K(\partial S, V) = 0$. Thus, because $S = \Omega$ yields an irrelevant lower bound, we have $K(\partial S, V) \geq 1$.

Assuming that $\mathcal{A} \subseteq V_1$, it is clear that $K(S, \mathcal{A}) = 0$ if S only contains states where nodes in V_2 are active. Hence in this case we should choose $S = \{u \in \Omega : \sum_{i \in V_2} u_i \geq 1\}$, the set of activity states where at least one of the nodes in V_2 is active, as this gives $K(S, V) = N/2$, $K(\partial S, V) = 1$ and $K(S, \mathcal{A}) = 0$. We thus see that the delay grows at least as fast as $\nu^{N/2-1}$.

6 Complete partite graphs

In the previous section, we derived generic lower bounds for the expected aggregate weighted queue length and delays in terms of the bottleneck ratio of any subset $S \subseteq \Omega$. In this section and the next, we describe how to find a subset $S \subseteq \Omega$ with the desired properties discussed in the previous section, for a broad class of K-partite interference graphs. We additionally assume that each of the nodes belongs to at least one clique of size K (of which the other $K - 1$ nodes necessarily belong to $K - 1$ different components).

We first introduce some further notation and state a few preparatory lemmas. Denote by $V_k \subseteq V$ the subset of nodes that belong to the k-th component and $M_k = |V_k|$, $k = 1, \ldots, K$. For compactness, define

$$\Upsilon_k = \prod_{i \in V_k} (1 + \sigma_i) - 1 = \sum_{I \subseteq V_k} \prod_{i \in I} \sigma_i - 1 = \sum_{\emptyset \neq I \subseteq V_k} \prod_{i \in I} \sigma_i.$$

In particular when $\sigma_i \equiv \hat{\sigma}_k$ for all $i \in V_k$, we have $\Upsilon_k = (1 + \hat{\sigma}_k)^{M_k} - 1$.

Throughout we assume that $\rho_i = \hat{\rho}_k > 0$ for all $i \in V_k$, and denote $\rho = \sum_{k=1}^K \hat{\rho}_k$, and $\rho_{\min} = \min_{k=1,\ldots,K} \hat{\rho}_k$. For convenience, we also assume $\phi_i \equiv 1$, $\psi_i \equiv 1$, $\mu_i \equiv 1$, so that $\sigma_i = \nu_i$ for all $i = 1, \ldots, N$. Define $M = \max_{k=1,\ldots,K} M_k$ as the maximum component size.

In order to gain some useful intuition, we focus first on complete K-partite graphs, where all nodes are connected except those that belong to the same component. In other words, the complement of the graph consists of K fully connected components. Thus, transmission activity is mutually exclusive across the various components.

In this case, the normalization constant in (20) satisfies

$$Z = 1 + \sum_{k=1}^{K} \sum_{\emptyset \neq I \subseteq V_k} \prod_{i \in I} \sigma_i = 1 + \sum_{k=1}^{K} \Upsilon_k.$$

For any $k = 1, \ldots, K$, define $S_k = \{u \in \Omega : \sum_{i \in V_k} u_i \geq 1\}$ as the set of activity states where at least one of the nodes in V_k is active. We will use these sets to find a lower bound for the delay. As discussed in Example 1, these sets are likely to provide a tight lower bound.

Lemma 2 *For any activation rate vector* $(\nu_1, \ldots, \nu_N)$ *such that the system is stable, for any* $k = 1, \ldots, K$*,*

$$Q(S_k) = Q(\Omega \backslash S_k) < M_k \left(1 - \sum_{l \neq k} \hat{\rho}_l\right) \left(\frac{1-\rho}{\hat{\rho}_k}\right)^{M_k - 1}, \tag{28}$$

$$\hat{\rho}_k < \pi(S_k) < 1 - \sum_{l \neq k} \hat{\rho}_l, \tag{29}$$

$$\sum_{l \neq k} \hat{\rho}_l < \pi(\Omega \backslash S_k) < 1 - \hat{\rho}_k. \tag{30}$$

Proof Using (20) and $\theta_i = \sum_{u \in \Omega} \pi(u) u_i$, we obtain

$$\theta_i = Z^{-1} \sum_{u \in \Omega, u_i = 1} \prod_{j=1}^{N} \sigma_j^{u_j} \geq Z^{-1} \sigma_i \prod_{l \in V_k \backslash \{i\}} (1 + \sigma_l) = Z^{-1} \frac{\sigma_i}{1 + \sigma_i} (\Upsilon_k + 1).$$

Also, from (20), we know

$$\theta_i = \sigma_i \mathbb{P}\{U_j = 0 \text{ for all } j \in \mathcal{N}_i^+\}.$$

Furthermore,

$$\mathbb{P}\{U_j = 0 \text{ for all } j \in \mathcal{N}_i^+\} \leq \mathbb{P}\{U_j = 0 \text{ for all } j \in \mathcal{C}\},$$

and hence we get

$$\theta_i \leq \sigma_i \left[1 - \sum_{j \in \mathcal{C}} \theta_j\right],$$

from (7), which may be rewritten as

$$\theta_i \leq \frac{\sigma_i}{1+\sigma_i}\left[1 - \sum_{j \in \mathcal{C}\setminus\{i\}} \theta_j\right],$$

so that

$$\Upsilon_k \leq Z\left[1 - \sum_{j \in \mathcal{C}\setminus\{i\}} \theta_j\right] - 1,$$

and thus, using the fact that $\rho_i < \theta_i$ for all $i \in V$ is a necessary condition for stability,

$$\Upsilon_k < Z\left[1 - \sum_{j \in \mathcal{C}\setminus\{i\}} \rho_j\right] - 1. \tag{31}$$

Next, note that $Q(\Omega \setminus S_k) = \pi(0) \sum_{i \in V_k} \sigma_i$, and similarly,

$$Q(S_k) = \pi(0) \sum_{i \in V_k} \sigma_i = \frac{1}{Z} \sum_{i \in V_k} \sigma_i.$$

Using this we get,

$$Q(S_k) \leq \frac{M_k}{Z} \max_{i \in V_k} \sigma_i = M_k \max_{i \in V_k} \frac{\frac{\sigma_i}{1+\sigma_i}(\Upsilon_k + 1)}{\frac{1}{1+\sigma_i}(\Upsilon_k + 1)} = \frac{M_k}{Z} \max_{i \in V_k} \frac{\frac{\sigma_i}{1+\sigma_i}(\Upsilon_k + 1)}{\prod_{l \in V_k\setminus\{i\}} (1+\sigma_l)}$$
$$< \frac{M_k(\Upsilon_k + 1)}{Z(1 + \min_{i \in V_k} \sigma_i)^{M_k - 1}}.$$

Invoking Lemma 1 and (31) gives (28).

Also, because $\hat{\rho}_k < \pi(S_k) = \Upsilon_k / Z$ is needed for stability, (31) gives,

$$\pi(S_k) < 1 - \sum_{l \neq k} \hat{\rho}_l - \frac{1}{Z},$$

which proves (29). Noting that $\pi(S_k) + \pi(\Omega \setminus S_k) = 1$ gives (30). □

Using Lemma 2, we can find a lower bound for the expected aggregate weighted queue length at some subset of nodes in $\mathcal{A} \subseteq V_k$.

Theorem 2 *For any activation rate vector* $(\nu_1, \ldots, \nu_N)$ *such that the system is stable and for any* $w \in \mathbb{R}_+^N$, $\mathcal{A} \subseteq V_k$,

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} > \frac{1}{2M} (\rho_{\min})^{M+1} \sum_{i \in \mathcal{A}} w_i \lambda_i \left(\frac{1}{1-\rho} \right)^{M-1}.$$

For the symmetric scenario $M_k \equiv M$ *and* $\hat{\rho}_k \equiv \rho/K$ *for all* $k = 1, \ldots, K$,

$$\mathbb{E}\{L_i\} > \frac{(K-1)^2 \rho^{M+2}}{2MK^{M+1}(K-(K-1)\rho)} \left(\frac{1}{1-\rho} \right)^{M-1}.$$

Proof The proof relies on applying Proposition 1, taking S to be (i) $\Omega \backslash S_k$ and (ii) S_l, $l \neq k$. In either case, $u_i = 0$ for all $i \in \mathcal{A}$, $u \in S$, so that $w_{\mathcal{A},S} = 0$ and $Y(w, \mathcal{A}, S) = 0$, i.e.,

$$D(w, \mathcal{A}, S) = \sum_{i \in \mathcal{A}} w_i \lambda_i.$$

First, consider case (i). In this case, we obtain the lower bound

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} > \frac{\left(\sum_{l \neq k} \hat{\rho}_l \right)^2}{2M_k} \sum_{i \in \mathcal{A}} w_i \lambda_i \left(\frac{\hat{\rho}_k}{1-\rho} \right)^{M_k - 1}$$

from Proposition 1 and Lemma 2. Taking $\mathcal{A} = V_k$ yields the second statement of the lemma for a symmetric scenario.

In order to complete the proof of the first part of the lemma, we now turn to case (ii). Using Proposition 1 and Lemma 2, we arrive at the lower bound

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} > \frac{\hat{\rho}_l^2}{2M_l} \sum_{i \in \mathcal{A}} w_i \lambda_i \left(\frac{\hat{\rho}_l}{1-\rho} \right)^{M_l - 1}.$$

Combining the above two lower bounds yields the first part of the lemma. □

Theorem 2 states that in a complete K-partite graph the expected queue length grows at least as fast as $1/(1-\rho)^{M-1}$, with M the size of the largest component. Based on the observations in Sect. 3, this may be heuristically explained as follows. In order for the system to be stable, each node must at least have an activation rate of the order $1/(1-\rho)$, see Lemma 1. In turn, the transition times between the various activity states as governed by the maximum-size component occur on a time scale of the order ν^{M-1}, when each node has a fixed activation rate ν.

For $M = 1$ (full interference graph), the lower bound established in Theorem 2 is loose, reflecting that it is not the slow transitions between the various components that cause the delays to be long in that case, but the sheer load. For $M = 2$, the lower bound could also have been obtained by treating cliques as single-resource systems. In fact, it follows already from (3) that the queue length is at least of the order $1/(1-\rho)$. For $M \geq 3$, the lower bound is particularly relevant, and reflects that the slow transitions between the various components cause the delays to be exponentially larger than can be explained from sheer load considerations alone.

7 Extensions

In this section, we turn attention to the broader class of (not necessarily complete) K-partite graphs. Thus, transmission activity is no longer mutually exclusive across the various components. However, we make the next assumption implying that joint activity across various components is relatively inefficient. Denote by $v^{(k)} = 1_{V_k}$ the incidence vector of V_k, i.e., $v_i^{(k)} = 1$ if $i \in V_k$ and $v_i^{(k)} = 0$ otherwise, and define $\Omega^* = \{v^{(1)}, \ldots, v^{(K)}\}$. Further, we adopt the notation and assumptions used in Sect. 6.

Assumption 1 For any $u \in \Omega$,

$$H(u) = \sum_{k=1}^{K} \sum_{i \in V_k} \frac{u_i}{M_k} \leq 1,$$

with strict inequality for any $u \notin \Omega^*$.

Based on the above assumption, we define

$$\zeta = 1 - \max_{u \notin \Omega^*} H(u) > 0.$$

An illustrative example is provided by a $2B \times 2B$ grid with nodes labeled as $\{(i, j)\}$, $i, j = 1, \ldots, 2B$, and nearest-neighbor interference. The two components are $V_1 = \{(i, j) : (i + j) \bmod 2 = 1\}$ and $V_2 = \{(i, j) : (i + j) \bmod 2 = 0\}$, with $M_1 = M_2 = 2B^2$. There are no other activity states $u \notin \Omega^*$ with $2B^2$ active nodes. Thus, $\sum_{i=1}^{2B} \sum_{j=1}^{2B} u_{(i,j)} \leq 2B^2 - 1$ for all $u \in \Omega \backslash \Omega^*$, and $\zeta = \frac{1}{2B^2}$.

The next lemma shows that in order for the system to be stable, joint activity across the various components can only occur a negligible fraction of the time at high load.

Lemma 3 *In order for the system to be stable, it should hold that*

$$\sum_{u \in \Omega \backslash \Omega^*} \pi(u) < \frac{1-\rho}{\zeta},$$

and

$$\pi(v^{(k)}) > \hat{\rho}_k - \frac{1-\rho}{\zeta}.$$

Proof In order for the system to be stable, we must have $\rho_i < \theta_i$ for all $i = 1, \ldots, N$. Thus

$$\begin{aligned}
\rho &= \sum_{k=1}^{K} \hat{\rho}_k = \sum_{k=1}^{K} \frac{1}{M_k} \sum_{i \in V_k} \rho_i \\
&< \sum_{k=1}^{K} \frac{1}{M_k} \sum_{i \in V_k} \sum_{u \in \Omega} \pi(u) u_i = \sum_{u \in \Omega} \pi(u) \sum_{k=1}^{K} \frac{1}{M_k} \sum_{i \in V_k} u_i \\
&= \sum_{u \in \Omega^*} \pi(u) \sum_{k=1}^{K} \frac{1}{M_k} \sum_{i \in V_k} u_i + \sum_{u \in \Omega \setminus \Omega^*} \pi(u) \sum_{k=1}^{K} \frac{1}{M_k} \sum_{i \in V_k} u_i \\
&\leq \sum_{u \in \Omega^*} \pi(u) + (1 - \zeta) \sum_{u \in \Omega \setminus \Omega^*} \pi(u) \\
&= 1 - \zeta \sum_{u \in \Omega \setminus \Omega^*} \pi(u),
\end{aligned}$$

where the last inequality follows from Assumption 1. The first part of the lemma follows.

Also, for any $i \in V_k$,

$$\hat{\rho}_k = \rho_i < \sum_{u \in \Omega} \pi(u) u_i \leq \pi(v^{(k)}) + \sum_{u \in \Omega \setminus \Omega^*} \pi(u),$$

which combined with the first statement yields the second part of the lemma. □

In the next lemma, we show that the fraction of time the activity process $\{U(t)\}$ spends in any component V_k relative to the traffic intensity of the nodes in that component, is almost equal for all components if ρ is large enough.

Lemma 4 *Assume the system is stable and* $\rho \geq \rho_\gamma = 1 - \gamma \zeta \rho_{\min}^2$, $\gamma > 0$. *Then*

$$\min_{k=1,\ldots,K} \frac{1}{\hat{\rho}_k} \prod_{i \in V_k} \sigma_i \geq (1 - 3\gamma) \max_{k=1,\ldots,K} \frac{1}{\hat{\rho}_k} \prod_{i \in V_k} \sigma_i.$$

Proof For compactness, we denote $\Pi_k = \prod_{i \in V_k} \sigma_i$ and $R_k = \Pi_k / \hat{\rho}_k$, and define $k_{\min} = \arg\min_{k=1,\ldots,K} R_k$ and $k_{\max} = \arg\max_{k=1,\ldots,K} R_k$.

Lemma 3 implies

$$\begin{aligned}
\hat{\rho}_{k_{\min}} - \gamma \rho_{\min}^2 \leq \hat{\rho}_{k_{\min}} - (1 - \rho)/\zeta \leq \pi(v^{(k_{\min})}) &= Z^{-1} \Pi_{k_{\min}} \\
&\leq \frac{\Pi_{k_{\min}}}{\sum_{k=1}^{K} \Pi_k} = \frac{\Pi_{k_{\min}}}{\Pi_{k_{\min}} + \Pi_{k_{\max}} + \sum_{k \neq k_{\min}, k_{\max}} \Pi_k} \\
&\leq \frac{\Pi_{k_{\min}}}{\Pi_{k_{\min}} + \Pi_{k_{\max}} + \sum_{k \neq k_{\min}, k_{\max}} \frac{\hat{\rho}_k}{\hat{\rho}_{k_{\min}}} \Pi_{k_{\min}}},
\end{aligned}$$

yielding

$$\left(1-\left(\hat{\rho}_{k_{\min}}+\sum_{k\neq k_{\min},k_{\max}}\hat{\rho}_k\right)\left(1-\frac{\gamma\rho_{\min}^2}{\hat{\rho}_{k_{\min}}}\right)\right)\frac{\Pi_{k_{\min}}}{\Pi_{k_{\max}}}\geq\hat{\rho}_{k_{\min}}-\gamma\rho_{\min}^2,$$

or equivalently,

$$(1-(\rho-\hat{\rho}_{k_{\max}})(1-\gamma\rho_{\min}^2/\hat{\rho}_{k_{\min}}))\Pi_{k_{\min}}\geq(\hat{\rho}_{k_{\min}}-\gamma\rho_{\min}^2)\Pi_{k_{\max}}.$$

Using $\rho\geq 1-\gamma\zeta\rho_{\min}^2$ and $\rho_{\min}\leq\hat{\rho}_{k_{\min}}$, it follows that

$$(1-(1-\gamma\zeta\rho_{\min}^2-\hat{\rho}_{k_{\max}})(1-\gamma\rho_{\min}))\Pi_{k_{\min}}\geq(1-\gamma)\hat{\rho}_{k_{\min}}\Pi_{k_{\max}}.$$

This yields

$$(1+2\gamma)\hat{\rho}_{k_{\max}}\Pi_{k_{\min}}\geq(1-\gamma)\hat{\rho}_{k_{\min}}\Pi_{k_{\max}},$$

and thus,

$$R_{k_{\min}}\geq\frac{(1-\gamma)R_{k_{\max}}}{1+2\gamma}\geq(1-3\gamma)R_{k_{\max}}.$$

□

In order to state a lower bound for the expected aggregate weighted queue length at some subset of nodes $\mathcal{A}\subseteq V_k$, we now first introduce some further notation and concepts.

A sequence of states $(u^{(0)},u^{(1)},\ldots,u^{(l)})$, with $u^{(k)}\in\Omega$, $k=0,\ldots,l$, is called a *path* from $u^{(0)}$ to $u^{(l)}$ if $(u^{(k)},u^{(k+1)})$ are feasible transitions, i.e., $q(u^{(k)},u^{(k+1)})>0$ for all $k=0,\ldots,l-1$. For a given path $p=(u^{(0)},u^{(1)},\ldots,u^{(l)})$, denote by $m(p)=\min_{k=0,1,\ldots,l}H(u^{(k)})$ the minimum value of the function $H(\cdot)$, as defined in Assumption 1, along the path. For given states $u,v\in\Omega$, denote by $P(u,v)$ the collection of all paths from u to v. Define $M(u,v)=\max_{p\in P(u,v)}m(p)$ as the maximum of the minimum value of the function $H(\cdot)$ along any path from state u to state v, with the convention that $M(u,u)=\infty$.

For all $\mathcal{A}\subseteq V$ such that $\mathcal{A}\subseteq V_k$ for some $k\in\{1,\ldots,K\}$, denote by $\Delta(\mathcal{A})$ the set of states in which the expected drift of the aggregate weighted queue length in $\mathcal{A}$ is non-positive, i.e.,

$$\Delta(\mathcal{A})=\left\{u\in\Omega:\sum_{i\in\mathcal{A}}w_i\lambda_i\leq\sum_{i\in\mathcal{A}}w_i\mu_iu_i\right\}.$$

Further, define $\delta(\mathcal{A})$ as the minimal expected drift of the aggregate weighted queue length in $\mathcal{A}$ if the system does not reside in of one of the states in $\Delta(\mathcal{A})$, i.e.,

$$\delta(\mathcal{A}) = \sum_{i \in \mathcal{A}} w_i \lambda_i - \max_{u \in \Omega \setminus \Delta(\mathcal{A})} \sum_{i \in \mathcal{A}} w_i \mu_i u_i.$$

Note that $\delta(\mathcal{A}) > 0$ by construction. For all $l \neq k$, define $m_l(\mathcal{A}) = \max_{u \in \Delta(\mathcal{A})} M(v^{(l)}, u)$, and

$$S_l(\mathcal{A}) = \{u \in \Omega : M(v^{(l)}, u) > m_l(\mathcal{A})\}$$

as the set of states that can be reached from $v^{(l)}$ via a path p with $m(p) > m_l(\mathcal{A})$. Also, define $m_k(\mathcal{A}) = \max_{l \neq k} m_l(\mathcal{A})$, and

$$S_k(\mathcal{A}) = \{v \in \Omega : \max_{u \in \Delta(\mathcal{A})} M(u, v) > m_k(\mathcal{A})\}$$

as the set of states that can be reached from $\Delta(\mathcal{A})$ via a path p with $m(p) > m_k(\mathcal{A})$. Finally, define $H_l(\mathcal{A}) = \max_{u \in \partial S_l(\mathcal{A})} H(u)$, $H^*(\mathcal{A}) = \min_{l=1,\ldots,K} H_l(\mathcal{A})$ and $H^*_{\min} = \min_{\mathcal{A} \subseteq V : \exists k : \mathcal{A} \subseteq V_k} H^*(\mathcal{A})$.

In deriving lower bounds for the delay in this section, we will use the sets $\Omega \setminus S_k(\mathcal{A})$ and $S_l(\mathcal{A})$, $l \neq k$, as the bottleneck sets, S. Note that, for a complete partite graph, these sets coincide with the bottleneck sets used in the previous section. Furthermore, for more general topologies, these sets are optimal in the sense that adding or removing one state will make $K(S, V)$ smaller, or $K(\partial S, V)$ larger, or $K(S, \mathcal{A})$ larger.

In the remainder of this section, we will assume that the activation rates of nodes in the same component are equal, i.e., $\sigma_i = \hat{\sigma}_k$ if $i \in V_k$ for all $k = 1, \ldots, K$. Denote $\sigma^* = \min_{k=1,\ldots,K} \hat{\sigma}_k^{M_k}$ and $k^* = \operatorname{argmin}_{k=1,\ldots,K} \hat{\sigma}_k^{M_k}$.

Remark 4 It is not clear when there exists an activation rate vector $(\nu_1, \ldots, \nu_N)$ with $\nu_i = \nu_j$ if $i, j \in V_k$ that stabilizes the system. For symmetric topologies, e.g., ring networks with an even number of nodes or tori with an even number of nodes in both directions, it seems plausible that such an activation rate vector can stabilize the system for any $\rho < 1$. For asymmetric topologies, e.g., linear topologies and two-dimensional grid networks, this is not clear.

In the next lemma, we derive an upper bound for the fraction of the time the system spends in the boundary of $S_l(\mathcal{A})$ for any $l = 1, \ldots, K$.

Lemma 5 *Assume the system is stable and* $\rho \geq \rho_\gamma = 1 - \gamma\zeta\rho^2_{\min}$, $\gamma > 0$. *Then*

$$\max_{u \in \partial S_l(\mathcal{A})} \prod_{j=1}^{N} \sigma_j^{u_j} \leq \left(\frac{\sigma^*}{(1 - 3\gamma)\rho_{\min}} \right)^{H_l(\mathcal{A})}.$$

Proof Since $\sigma_i = \hat{\sigma}_k$ for all $i \in V_k$, we obtain

$$\max_{u \in \partial S_l(\mathcal{A})} \prod_{j=1}^{N} \sigma_j^{u_j} = \max_{u \in \partial S_l(\mathcal{A})} \prod_{k=1}^{K} \prod_{i \in V_k} \sigma_i^{u_i} = \max_{u \in \partial S_l(\mathcal{A})} \prod_{k=1}^{K} \hat{\sigma}_k^{\sum_{i \in V_k} u_i}$$
$$= \max_{u \in \partial S_l(\mathcal{A})} \prod_{k=1}^{K} (\hat{\sigma}_k^{M_k})^{\frac{1}{M_k} \sum_{i \in V_k} u_i}.$$

Lemma 4 gives

$$\hat{\sigma}_k^{M_k} \leq \frac{\hat{\rho}_k}{1 - 3\gamma} \min_{l=1,\ldots,K} \frac{1}{\hat{\rho}_l} \hat{\sigma}_l^{M_l} \leq \frac{\sigma^*}{(1 - 3\gamma)\rho_{\min}},$$

and thus,

$$\max_{u \in \partial S_l(\mathcal{A})} \prod_{j=1}^{N} \sigma_j^{u_j} \leq \max_{u \in \partial S_l(\mathcal{A})} \prod_{k=1}^{K} \left(\frac{\sigma^*}{(1 - 3\gamma)\rho_{\min}} \right)^{\frac{1}{M_k} \sum_{i \in V_k} u_i}$$
$$\leq \max_{u \in \partial S_l(\mathcal{A})} \left(\frac{\sigma^*}{(1 - 3\gamma)\rho_{\min}} \right)^{\sum_{k=1}^{K} \frac{1}{M_k} \sum_{i \in V_k} u_i}$$
$$= \left(\frac{\sigma^*}{(1 - 3\gamma)\rho_{\min}} \right)^{\max_{u \in \partial S_l(\mathcal{A})} H(u)} = \left(\frac{\sigma^*}{(1 - 3\gamma)\rho_{\min}} \right)^{H_l(\mathcal{A})}.$$

□

We are now in the position to derive bounds for $Q(S_l(\mathcal{A}))$, $\pi(S_l(\mathcal{A}))$, and $\pi(\Omega \backslash S_l(\mathcal{A}))$ that are qualitatively similar to the bounds in Lemma 2.

Lemma 6 *Assume $\rho \geq \rho_\gamma = 1 - \gamma \zeta \rho_{\min}^2$, $\gamma > 0$. For any activation rate vector $(\nu_1, \ldots, \nu_N)$ such that the system is stable and with $\nu_i = \nu_j$ if $i, j \in V_k$ for some k, for any $l = 1, \ldots, K$,*

$$Q(S_l(\mathcal{A})) = Q(\Omega \backslash S_l(\mathcal{A})) < \frac{2^N}{(1 - 3\gamma)\rho_{\min}} \left(\frac{\rho_{k^*}}{1 - \rho} \right)^{M(H_l(\mathcal{A}) - 1)}, \tag{32}$$

$$(1 - \gamma)\rho_{\min} < \pi(S_l(\mathcal{A})) < 1 - (1 - \gamma)\rho_{\min}, \tag{33}$$

$$(1 - \gamma)\rho_{\min} < \pi(\Omega \backslash S_l(\mathcal{A})) < 1 - (1 - \gamma)\rho_{\min}. \tag{34}$$

Proof First note that

$$Q(S_l(\mathcal{A})) = \sum_{u \in \partial S_l(\mathcal{A})} \pi(u) = Z^{-1} \sum_{u \in \partial S_l(\mathcal{A})} \prod_{j=1}^{N} \sigma_j^{u_j}$$
$$\leq Z^{-1} |\partial S_l(\mathcal{A})| \max_{u \in \partial S_l(\mathcal{A})} \prod_{j=1}^{N} \sigma_j^{u_j}.$$

Noting that $Z \geq \sigma^*$ and $|\partial S_l(\mathcal{A})| \leq 2^N$ yields, using Lemma 5,

$$Q(S_l(\mathcal{A})) \leq \frac{2^N \left(\frac{\sigma^*}{(1-3\gamma)\rho_{\min}}\right)^{H_l(\mathcal{A})}}{\sigma^*} \leq \frac{2^N (\sigma^*)^{H_l(\mathcal{A})-1}}{(1-3\gamma)\rho_{\min}},$$

and (32) follows from Lemma 1.

Further, using Lemma 3,

$$\pi(S_l(\mathcal{A})) \geq \pi(v^{(l)}) > \hat{\rho}_l - \frac{1-\rho}{\zeta} \geq (1-\gamma)\rho_{\min}.$$

Now note that by definition $S_l(\mathcal{A}) \cap \Delta(\mathcal{A}) = \emptyset$ for $l \neq k$ and $\Delta(\mathcal{A}) \subseteq S_k(\mathcal{A})$ if $\mathcal{A} \subseteq V_k$. Hence, for $l \neq k$,

$$\pi(S_l(\mathcal{A})) \leq 1 - \pi(v^{(k)}) < 1 - (1-\gamma)\rho_{\min},$$

and

$$\pi(S_k(\mathcal{A})) \leq 1 - \sum_{l \neq k} \pi(v^{(l)}) < 1 - (K-1)(1-\gamma)\rho_{\min},$$

which gives (33). Noting that $\pi(S_l(\mathcal{A})) + \pi(\Omega \backslash S_l(\mathcal{A})) = 1$ gives (34). □

Using a similar approach as in Sect. 6, the bounds in Lemma 6 can be utilized to establish a lower bound for the expected aggregate weighted queue length in some subset of nodes.

Theorem 3 *Assume $\rho \geq \rho_\gamma = 1 - \gamma\zeta\rho_{\min}^2$, $\gamma > 0$. For any activation rate vector $(\nu_1, \ldots, \nu_N)$, with $\nu_i = \nu_j$ if $i, j \in V_k$ for some k, such that the system is stable and for any $w \in \mathbb{R}_+^N$, $\mathcal{A} \subseteq V_k$,*

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} > \frac{\delta(\mathcal{A})(1-4\gamma)\rho_{\min}^{M+3}}{2^{N+1}} \left(\frac{1}{1-\rho}\right)^{M(1-H^*(\mathcal{A}))} - \frac{1}{2} w_{\mathcal{A},S}(1-(1-\gamma)\rho_{\min}),$$

Proof The proof of this theorem proceeds along similar lines as the proof of Theorem 2 and relies on applying Proposition 1, taking S to be (i) $\Omega \backslash S_k(\mathcal{A})$ and (ii) $S = S_l(\mathcal{A})$,

$l \neq k$. First, note that by definition $S_l(\mathcal{A}) \cap \Delta(\mathcal{A}) = \emptyset$, and thus $D(w, \mathcal{A}, S_l(\mathcal{A})) \geq \delta(\mathcal{A}), l \neq k$. Also note that $\Delta(\mathcal{A}) \subseteq S_k(\mathcal{A})$, so that $D(w, \mathcal{A}, \Omega \backslash S_k(\mathcal{A})) \geq \delta(\mathcal{A})$.

Further, using Lemma 6 we obtain the lower bound

$$\sum_{i \in \mathcal{A}} w_i \mathbb{E}\{L_i\} > \frac{\delta(\mathcal{A})(1-4\gamma)\rho_{\min}^{M+3}}{2^{N+1}} \left(\frac{1}{1-\rho}\right)^{M(1-H_l(\mathcal{A}))} - \frac{1}{2} w_{\mathcal{A},S}(1-(1-\gamma)\rho_{\min}),$$

for $l = 1, \ldots, K$, and the result follows. □

Theorem 3 states that in a general K-partite interference graph, the expected queue length grows at least as fast as $1/(1-\rho)^{M(1-H^*)}$, where the coefficient H^* depends on the specific topology and is in general hard to calculate. We however know that $\frac{1}{M} \leq H^* \leq 1$ and for some specific topologies we can explicitly determine H^*.

The value of the coefficient $H^*(\mathcal{A})$ depends strongly on the specific properties of the interference graph G. As mentioned, for a complete partite graph the sets $S_l(\mathcal{A})$ coincide with those in the previous section, and we have $\partial S_l(\mathcal{A}) = \bigcup_{i \in V_l} \{e_i\}$, so that $H_l(\mathcal{A}) = 1/M_l$, and $H^*(\mathcal{A}) = 1/M$, recovering the result of Theorem 2. On the other hand, when the graph consists of N/K fully connected components, we have $H_l(\mathcal{A}) \equiv 1$, and the result trivializes. An interesting intermediate situation is the $2B \times 2B$ grid mentioned earlier with $M = M_1 = M_2 = 2B^2$, for which we conjecture that for $\rho_{\min}$ sufficiently large, $H^*(\mathcal{A}) = H_1 = H_2 = 1 - 1/B$ or $1 - 1/(2B)$ if $B \geq 2$, depending on whether or not we assume a wrap-around boundary, suggesting that the mean queue lengths would grow as $1/(1-\rho)^B$ or $1/(1-\rho)^{2B}$.

8 Simulation experiments

In this section, we will illustrate the theoretical results for the growth behavior of the aggregate queue length through simulation experiments. For cross comparison, we consider a system that can be represented by a symmetric complete bipartite ($K = 2$) interference graph with components of size $M = 5$. Because of space considerations, we do not report simulation results for other cases, but we observed qualitatively similar behavior in a broad range of scenarios.

To estimate the expected aggregate queue length for a given value of ρ, we set $t = 10^6$ and calculate the average total number of packets in the time intervals $[0, t]$ and $[t + 1, 2t]$, starting from an initially empty system. We take the average of the two values to be our estimate if the values are less than 5% apart. Otherwise, we set $t = 2t$ and repeat the procedure.

Figure 1 shows the average total number of packets in the system for various fixed activation rates. Note that we used a log-lin scale. We see that the simulated curves lie well above the lower bound of Theorem 2 for all chosen values of ν. Note that, the system is not stable for all values of ρ, e.g., for $\nu = 1$ the system is unstable if $\rho \geq 2\theta_i = \frac{32}{63}$, explaining the jumps in the simulation result. Further, note that the expected time between activation of nodes in the two components is smaller for small

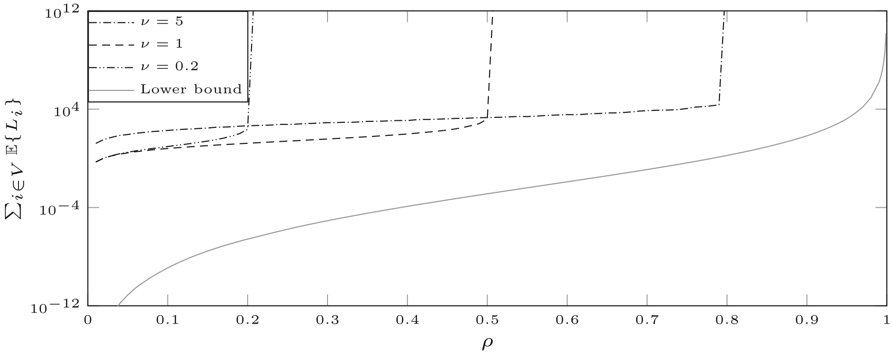

Fig. 1 Average total number of packets for several fixed activation rates

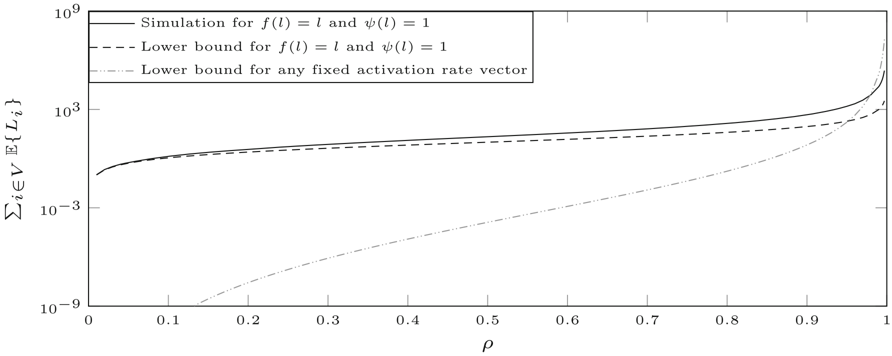

Fig. 2 Average total number of packets for $f(l) = l$ and $\psi(l) = 1$

values of ν. This explains why small values of ν tend to perform better in the case when ρ is small, i.e., for large values of ν the nodes in one component will often be transmitting dummy packets while the nodes in the other component do have packets waiting to be transmitted.

Figure 2 shows the average total number of packets in the system for $f(l) = l$ and $\psi(l) = 1$. We see that the lower bound of Theorem 1 is remarkably close to the simulation result for small values of ρ. For larger values of ρ the bound and simulation result are farther apart. One explanation for this lies in the approximation made in (6). For small values of ρ, this approximation is relatively good while for large values of ρ this approximation is off by a factor of about 2 in this case. While this does not explain the total discrepancy in this case, it does explain all discrepancies in the case when the rate of increase of the activation function is slow, e.g., $f(l) = \log(l + 1)$.

Finally, note that the simulation result lies, for large values of ρ, below the lower bound for fixed activation rates established in Theorem 2. This suggests that the activation function $f_i(l) \equiv f(l) = l$ performs better in heavy traffic than $f_i(l) = \nu_i$ for any choice of the activation rate vector $(\nu_1, \ldots, \nu_N)$ in this topology, but in general topologies this function may not always ensure throughput-optimality [29].

9 Conclusions

We have established lower bounds for the expected queue lengths and delays in wireless random-access networks. Both for queue-based strategies and fixed activation rates, the derivation of the bounds starts from the observation that stability of the system requires the activity factors to be big at high load. The specific subsequent arguments considerably differ however in both cases. Queue-based strategies for which maximum stability has been established, involve slow, logarithmic, activation functions, which require huge queue lengths at every node for the activity factors to be big enough, and cause the exponential delay scaling. In contrast, the delays for fixed activation rates are shown to result from excessive mixing times due to a bottleneck in the network topology together with the big activity factors required for stability. We also observe that the network topology plays a major role in the case of fixed activation rates, while it only appears to matter somewhat implicitly in the case of queue-based strategies as will be further discussed below.

For complete partite interference graphs, a comparison of both cases reveals that the expected delay for queue-based strategies grows faster than the lower bound $1/(1-\rho)^{M-1}$ for fixed activation and de-activation rates when $h(l)$ increases slower than $l^{1/(M-1)}$, with M denoting the maximum component size. This is for example the case if $f(l) = r(l)/(1+r(l))$ and $\psi(l) = 1/(1+r(l))$, with $r(l) = \log(l+1)$. Conversely, when $h(l)$ increases faster than $l^{1/(M-1)}$, the lower bound for fixed activation and de-activation rates could potentially be beaten by sufficiently aggressive queue-based strategies. Simulation experiments demonstrate that the actual expected delays indeed exhibit the cross-over suggested by the lower bounds.

A challenging issue for further research is to examine whether more aggressive queue-based strategies can improve the delay performance in more general topologies as well. As noted earlier, maximum-stability guarantees in arbitrary topologies have only been established so far for nominal activation functions that grow logarithmically with the queue lengths [13–15]. Inspection of the proof arguments indicates that maximum stability will remain guaranteed as long as the fluid limits of the queue length process exhibit fast mixing behavior. This in turn means that the activity process for such queue-based strategies in fact behaves as if the activation rates are essentially fixed. Thus, in arbitrary topologies it is questionable whether queue-based strategies have the capability to outperform fixed-rate strategies.

In some specific topologies, however, maximum stability is maintained for highly aggressive queue-based strategies for which the fluid limits of the queue length process may exhibit slow mixing behavior [29,35]. The complete partite interference graphs considered in the present paper are crucial examples of such topologies. In these scenarios, there seems to be scope for more aggressive queue-based strategies to reduce the delays, as confirmed by the lower bounds and simulation results that we presented.

In conclusion, the question in what kind of scenarios more aggressive queue-based strategies can improve the delay performance appears to be inextricably linked to the question under what conditions such strategies provide maximum-stability guarantees. In both these questions, the mixing properties of the activity process seem to play a central role, and it would be interesting to explore this three-way connection further.

Acknowledgments This work was supported by Microsoft Research through its PhD Scholarship Programme, by the European Research Council (ERC) and by the Netherlands Organisation for Scientific Research (NWO). We thank Florian Simatos and Phil Whiting for many helpful discussions and comments.

References

1. Boorstyn, R., Kershenbaum, A., Maglaris, B., Sahin, V.: Throughput analysis in multihop CSMA packet radio networks. IEEE Trans. Commun. **35**, 267–274 (1987)
2. Wang, X., Kar, K.: Throughput modeling and fairness issues in CSMA/CA based ad-hoc networks. In: Proceedings Infocom 2005 Conference (2005)
3. Durvy, M., Dousse, O., Thiran, P.: Modeling the 802.11 protocol under different capture and sensing capabilities. In: Proceedings Infocom 2007 Conference (2007)
4. Durvy, M., Thiran, P.: A packing approach to compare slotted and non-slotted medium access control. In: Proceedings Infocom 2006 Conference (2006)
5. Garetto, M., Salonidis, T., Knightly, E.: Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks. IEEE/ACM Trans. Netw. **16**(4), 864–877 (2008)
6. Bianchi, G.: Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J. Sel. Areas Commun. **18**(3), 535–547 (2000)
7. Liew, S., Kai, C., Leung, J., Wong, B.: Back-of-the-envelope computation of throughput distributions in CSMA wireless networks. IEEE Trans. Mob. Comput. **9**(9), 1319–1331 (2010)
8. Jiang, L., Walrand, J.: A distributed CSMA algorithm for throughput and utility maximization in wireless networks. IEEE/ACM Trans. Netw. **18**(3), 960–972 (2010)
9. Liu, J., Yi, Y., Proutière, A., Chiang, M., Poor, H.: Towards utility-optimal random access without message passing. Wireless Commun. Mob. Comput. **10**(1), 115–128 (2010)
10. Tassiulas, L., Ephremides, A.: Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. IEEE Trans. Autom. Control **37**, 1936–1948 (1992)
11. Jiang, L., Shah, D., Shin, J., Walrand, J.: Distributed random access algorithm: scheduling and congestion control. IEEE Trans. Inf. Theory **56**(12), 6182–6207 (2010)
12. Marbach, P., Eryilmaz, A.: A backlog-based CSMA mechanism to achieve fairness and throughput-optimality in multihop wireless networks. In: Proceedings Allerton 2008 Conference (2008)
13. Rajagopalan, S., Shah, D., Shin, J.: Network adiabatic theorem: an efficient randomized protocol for contention resolution. In: Proceedings ACM SIGMETRICS/Performance 2009 Conference (2009)
14. Shah, D., Shin, J.: Randomized scheduling algorithm for queueing networks. Ann. Appl. Probab. **22**, 128–171 (2012)
15. Shah, D., Shin, J., Tetali, P.: Medium access using queues. In: Proceedings FOCS 2011 Conference (2011)
16. Ghaderi, J., Srikant, R.: On the design of efficient CSMA algorithms for wireless networks. In: Proceedings CDC 2010 Conference (2010)
17. Lam, K., Chau, C., Chen, M., Liew, S.: Mixing time and temporal starvation of general CSMA networks with multiple frequency agility. In: Proceedings ISIT 2012 Conference (2012)
18. Lotfinezhad, M., Marbach, P.: Throughput-optimal random access with order-optimal delay. In: Proceedings Infocom 2011 Conference (2011)
19. Ni, J., Tan, B., Srikant, R.: Q-CSMA: queue length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks. In: Proceedings Infocom 2010 Mini-Conference (2010)
20. Shah, D., Shin, J.: Delay-optimal queue-based CSMA. In: Proceedings ACM SIGMETRICS 2010 Conference (2010)
21. Bouman, N., Borst, S., Boxma, O., Van Leeuwaarden, J.: Queues with random back-offs. Queueing Syst. **77**(1), 33–74 (2013)
22. Shah, D., Tse, D., Tsitsiklis, J.: Hardness of low delay network scheduling. IEEE Trans. Inf. Theory **57**(12), 7810–7817 (2011)
23. Jiang, L., Leconte, M., Ni, J., Srikant, R., Walrand, J.: Fast mixing of parallel glauber dynamics and low-delay CSMA scheduling. In: Proceedings Infocom 2011 Mini-Conference (2011)
24. Jiang, L., Walrand, J.: Stability and delay of distributed scheduling algorithms for networks of conflicting queues. Queueing Syst. **72**(1), 161–187 (2012)

25. Subramanian, V., Alanyali, M.: Delay performance of CSMA in networks with bounded degree conflict graphs. In: Proceedings ISIT 2011 Conference (2011)
26. Bouman, N., Borst, S., Van Leeuwaarden, J.: Delays and mixing times in random-access networks. In: Proceedings ACM SIGMETRICS, pp. 117–128 (2013)
27. Bouman, N., Borst, S., Van Leeuwaarden, J.: Achievable delay performance in CSMA networks. In: Proceedings 49th Allerton Conference (2011)
28. Bouman, N., Borst, S., Van Leeuwaarden, J., Proutière, A.: Backlog-based random access in wireless networks: fluid limits and delay issues. In: Proceedings ITC 23 (2011)
29. Ghaderi, J., Borst, S., Whiting, P.: Backlog-based random-access algorithms: fluid limits and stability issues. In: Proceedings WiOpt 2012 Conference (2012)
30. Van de Ven, P., Janssen, A., Van Leeuwaarden, J., Borst, S.: Achieving target throughputs in random-access networks. Perform. Eval. **68**(11), 1103–1117 (2011)
31. Boxma, O.: Workloads and waiting times in single-server systems with multiple customer classes. Queueing Syst. **5**, 185–214 (1989)
32. Boxma, O., Groenendijk, W.: Pseudo-conservation laws in cyclic-service systems. J. Appl. Probab. **24**, 949–964 (1987)
33. Levin, D., Peres, Y., Wilmer, E.: Markov Chains and Mixing Times. AMS, Providence (2009)
34. Ross, S.: Stochastic Processes. Wiley, New York (1996)
35. Feuillet, M., Proutière, A., Robert, P.: Random capture algorithms: Fluid limits and stability. In: Proceedings ITA Workshop (2010)