Stability of Spatial Wireless Systems with Random Admissible-Set Scheduling

N. Bouman
Eindhoven University of Technology
EURANDOM
n.bouman@TUE.nl

S.C. Borst
Eindhoven University of Technology
Alcatel-Lucent Bell Labs
s.c.borst@TUE.nl

J.S.H. van Leeuwaarden
Eindhoven University of Technology
EURANDOM
j.s.h.v.leeuwaarden@TUE.nl

ABSTRACT

We examine the stability of wireless networks whose users are distributed over a torus. Users arrive at spatially uniform locations with intensity λ and each user has a random number of packets to transmit with mean β. In each time slot, an *admissible* subset of users is selected uniformly at random to transmit one packet. A subset of users is called admissible when their simultaneous activity obeys the prevailing interference constraints. We consider the SINR model and the protocol model as two canonical models for interference, and denote by μ the maximum number of users in an admissible subset for the model under consideration. We show that the necessary condition $\lambda\beta < \mu$ is also sufficient for random admissible-set scheduling to achieve stability. Thus random admissible-set scheduling achieves stability, if feasible to do so at all, for a wide range of interference scenarios. The proof relies on a description of the system as a measure-valued process and the identification of a Lyapunov function.

Keywords

Wireless networks, stability, Foster-Lyapunov, Harris recurrent, measure-valued process, interference constraints, SINR requirements, protocol model

1. INTRODUCTION

The present paper examines the stability of a broad class of wireless networks whose users arrive to a torus H (of arbitrary dimension). The torus may be interpreted as a square coverage area with a wrap-around boundary, which is a common notion in the modeling of wireless networks to avoid boundary effects. Time is slotted, and users arrive at H according to some spatial stochastic process with mean λ per time slot. Users independently take their locations in H at random according to the uniform distribution. Each user has a random number of packets to transmit, generally distributed with mean β, and can transmit at most one packet per time slot.

In each time slot, we select a set of users for transmission from all admissible sets uniformly at random. A subset of users is called admissible when their simultaneous activity obeys the prevailing interference constraints. In practice, the relevant interference constraints depend on various system-specific properties, such as the propagation environment and the operation of the physical and medium access layers of the network. In the present paper we therefore adopt generic feasibility criteria, which in particular cover the SINR (Signal-to-Interference-and-Noise Ratio) model and the protocol model as two canonical models for interference, see for instance [5].

Let μ be the maximum number of users in an admissible set. It is clear that $\lambda\beta \leq \mu$ is a *necessary* condition for stability: The mean number of packets that arrive per time slot should be no larger than the maximum number of packets that can be transmitted simultaneously. The main result we show is that this necessary condition is also nearly *sufficient* for stability. Specifically, the Markov chain describing the evolution of the system is positive Harris recurrent when $\lambda\beta < \mu$, implying that the network will be empty infinitely often. Thus random admissible-set scheduling is a highly robust strategy in that it achieves stability, if feasible to do so at all, for a wide range of interference scenarios.

As described above, we investigate stability in the context of a model that combines a scheduling discipline operating under interference constraints and a continuous spatial setting. While these two elements have each been considered in isolation before, the present paper is, to the best of our knowledge, the first to capture both features in conjunction. Indeed, stability of wireless networks has been widely studied in the literature, see for instance Bonald and Feuillet [2], Bordenave *et al.* [3] and Wu *et al.* [11]. These papers restrict attention to discrete topologies and interference constraints such that the system can be represented as a conflict graph. Our model does not allow such a representation due to the continuum of locations, and hence these results are not directly applicable to our problem. Stability of queueing networks in continuous space is investigated in Altman and Levy [1], Leskelä and Unger [7] and Robert [10]. These papers prove stability of networks in which only one user is allowed to transmit at a time. In contrast, the present paper focuses on the more complex situation of simultaneous transmissions governed by a scheduling discipline.

While a discrete network structure is a reasonable assumption in case of a relatively small number of long-lived flows, it is less suitable in case of a relatively large number of short-lived flows. The latter scenario is increasingly relevant

as emerging wireless networks support traffic generated by massive numbers of nodes, each of which engage in sporadic transmission activity. The continuous spatial setting also provides useful insights in the scaling behavior of discrete topologies as the number of nodes grows large.

From a methodological perspective, the continuous spatial setting involves major additional challenges compared to a discrete network structure. Since users reside in a continuum of locations, the evolution of the system cannot be represented in terms of a Markov chain with a finite state space, and we therefore introduce a measure-valued process as a description of the system. In order to prove stability, we identify a Lyapunov function which has a negative drift for all but a 'small' set of states, so that the Markov chain is positive Harris recurrent.

The remainder of the paper is organized as follows. In Section 2 we show how the evolution of the system may be described in terms of a Markov chain with a measure-valued state space. The main stability result is presented in Section 3, along with an interpretation and discussion of its ramifications. In Section 4 we provide a proof outline, and in particular identify a Lyapunov function which has negative drift for all but a 'small' set of states. In Appendix A we recall various useful definitions and collect some preliminaries that are needed in order to apply the Foster-Lyapunov approach for our specific Markov chain. Appendix B provides the full proof details for the stability result.

2. MODEL DESCRIPTION

Consider the hypercube $H = [0, 1)^n$. To avoid boundary conditions we connect the boundary points of the hypercube to the opposite boundary points, so that the space can be thought of as the surface of an n-dimensional torus. The distance between $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, $x, y \in H$, is then defined by

$$D(x, y) = \left(\sum_{i=1}^{n} D_1(x_i, y_i)^2 \right)^{1/2},$$

where

$$D_1(a, b) = \min(|a - b|, 1 - |a - b|).$$

Users arrive at locations uniformly distributed in H, independent of the locations of other users. The number of users that arrive during a time slot has a general non-negative discrete distribution with mean λ and is independent of the number of users that arrive in other time slots. Each user has a number of packets to transmit. The number of packets a user has to transmit has a general non-negative discrete distribution with mean β. Denote by $A(t, s)$ the number of packets arriving during the t-th time slot in the orthotope $[0, s_1) \times \cdots \times [0, s_n)$, with $s_1, \ldots, s_n \leq 1$. That is, the numbers $A(t, \mathbf{1})$, $t = 1, 2, \ldots$, are i.i.d. copies of a non-negative discrete random variable A with $\mathbb{E}\{A\} = \lambda\beta$. Here $\mathbf{1}$ is the all-one vector, so that $A(t, \mathbf{1})$ is the total number of packets that arrive during the t-th time slot. We further assume that $\mathbb{E}\{A \log(A) | A > 0\} < \infty$ and that $0 < \mathbb{P}\{A > 0\} < 1$. Note that, because users arrive to a location according to a uniform distribution, the expected number of packets to arrive to a subspace $B \subseteq H$ in one time slot is given by $\mathbb{E}\{A\} \nu(B)$, where $\nu(B)$ denotes the volume or surface area of B.

We denote the number of packets in H at the start of the t-th time slot by $Y(t)$, with $Y(t) = (Y(t, s), 0 < s \leq 1)$

and $Y(t, s)$ denoting the number of packets residing in the orthotope $[0, s_1) \times \cdots \times [0, s_n)$ at the start of the t-th time slot. The state space of this process is denoted by Ψ and consists of all nondecreasing integer-valued step functions on H. So, when $y \in \Psi$, $y(s)$ denotes the number of packets residing in the orthotope $[0, s_1) \times \cdots \times [0, s_n)$. Note that, using the information in Ψ, we can in fact find the number of packets in any Borel set $B \subseteq [0, 1)^n$. We denote this number by $y(B)$. In particular, $y(\{x\})$ denotes the number of packets at location $x \in H$.

At the start of every time slot an admissible subset of users is selected to transmit one packet in that time slot. Here, $z \in \Psi$ is called a subset of $y \in \Psi$ if $z(\{x\}) \leq y(\{x\})$ for all $x \in H$. To decide whether a set is admissible we define a function $F : \Psi \to \{0, 1\}$ and we call $z \in \Psi$ admissible if and only if $F(z) = 1$. The function F follows from the interference constraints of the wireless network, so that an admissible set consists of users whose transmissions are all successful if only the users in that set are transmitting at the same time. In Subsections 2.1 and 2.2 we will discuss some prototypical examples of this function in more detail.

Let $\chi(y)$ be the set of all subsets of y, i.e.

$$\chi(y) = \{z \in \Psi : z(\{x\}) \leq y(\{x\}), \forall x \in H\},$$

and let $R(t, Y(t), s)$ be the number of packets transmitted from the orthotope $[0, s_1) \times \cdots \times [0, s_n)$ in the t-th time slot, given the configuration, $Y(t)$. An admissible subset of packets is selected uniformly at random. Hence, given $Y(t) = y$, $R(t, y) = z$ with probability

$$\frac{F(z) \displaystyle\prod_{x \in H, z(\{x\}) > 0} y(\{x\})}{\displaystyle\sum_{u \in \chi(y)} F(u) \prod_{x \in H, u(\{x\}) > 0} y(\{x\})},$$

where $R(t, Y(t)) = (R(t, Y(t), s), 0 < s \leq 1)$. Note that for $z \notin \chi(y)$ we have $F(z) = 0$ or $\exists x \in H$ such that $z(\{x\}) = 1$ and $y(\{x\}) = 0$, so this probability is always zero in this case. Here and in the remainder of this paper, the value of the empty product is defined as 1.

The evolution of $Y(t, s)$ is described by the recursion

$$Y(t + 1, s) = Y(t, s) + A(t, s) - R(t, Y(t), s).$$

From this description it follows that $(Y(t))_{t \in \mathbb{N}}$ is a Markov chain. We will equip the state space of the Markov chain Ψ with the smallest σ-field $\mathcal{B}(\Psi)$ with respect to which the map $y \to y(B)$ is measurable for any Borel set $B \subseteq H$. That is, we equip Ψ with the Borel σ-field as we will prove in Lemma 8.

Below we discuss two common models for interference, see for instance Gupta and Kumar [5].

2.1 Protocol model

Under the *protocol model* for interference two users can successfully transmit at the same time whenever the distance between these two users is at least r. This minimal distance r is often called the reuse distance. Let $\{U_i \in H : i \in I\}$ be the set of active users in some time slot. The transmission of a user $U_i, i \in I$, is successful if

$$D(U_i, U_j) \geq r \text{ for all } j \in I. \tag{1}$$

Further, the transmission of all users is successful if (1) holds for all $i \in I$.

To decide whether $y \in \Psi$ is admissible we thus take the function $F : \Psi \to \{0,1\}$ such that $F(y) = 1$ if and only if $y(\{x\}) \in \{0,1\}, \forall x \in H$ and $y(\{x\})y(\{w\}) > 0$ only if $D(x,w) \geq r, \forall x \neq w \in H$.

2.2 SINR model

Again, let $\{U_i \in H : i \in I\}$ be the set of active users in some time slot. The transmission of a user $U_i, i \in I$, is successful under the *SINR model* for interference if

$$\frac{P}{\eta + P \sum\limits_{j \in I \setminus \{i\}} D(U_i, U_j)^{-\alpha}} \geq \gamma. \qquad (2)$$

Here η, P and α are constants representing the background noise, the power at which a user transmits and the path loss exponent, respectively. If (2) holds for all $i \in I$ the transmission of all users is successful and thus that set is admissible.

The function $F : \Psi \to \{0,1\}$ associated with this model for interference is now such that $F(y) = 1$ if and only if for all $x \in H$, $y(\{x\}) \in \{0,1\}$ and

$$\frac{Py(\{x\})}{\eta + P \sum\limits_{\substack{w \in H \\ y(\{w\})>0}} D(x,w)^{-\alpha}} \geq \gamma y(\{x\}).$$

3. MAIN RESULT

In this section we present the main result, along with an intuitive explanation.

Let μ be the maximum number of users in an admissible set, i.e. $\mu = \max\{y(\mathbf{1}) : F(y) = 1\}$. To avoid having almost surely no admissible subset of packets of maximum size μ at any point in time, we will assume that there exists an admissible set of maximal size such that the set remains admissible if all packets are moved over a distance of at most $\delta > 0$. More formally, we make the following assumption:

Assumption 1. There exists an $y \in \Psi$ with $y(\mathbf{1}) = \mu$ and $F(y) = 1$ such that for $z \in \Psi$ with $z(\{x_i\}) > 0$ for $i = 1, \ldots, \mu$ and $X = \{x_1, \ldots, x_\mu\} \in \mathcal{X}_\delta$ for some $\delta > 0$ we have $F(z) = 1$. Here $\mathcal{X}_\delta = \{X = \{x_1, \ldots, x_\mu\} : D(x_i, w_i) < \delta, i = 1, \ldots, \mu\}$ and $\{w_1, \ldots, w_\mu\} = \{w \in H : y(\{w\}) > 0\}$.

As an example consider dimension $n = 1$, so that we have a unit circle. Further consider the protocol model for interference with reuse distance r introduced in Subsection 2.1. Then the above assumption boils down to assuming that the value of $1/r$ is non-integer. Alternatively, the constraint in (1) could be assumed to hold with strict inequality.

The following theorem states the main result of this paper and is a shorter version of Theorem 6, which is proven at the end of Section 4.

THEOREM 1. *Under Assumption 1, the Markov chain* $(Y(t))_{t \in \mathbb{N}}$ *is positive Harris recurrent if* $\lambda\beta < \mu$ *and if interference is modeled by the protocol model or the SINR model.*

This theorem states that $Y(t, \mathbf{1}) = 0$ for infinitely many values of t, i.e. the Markov chain will be in the empty configuration infinitely often.

The result of Theorem 1 may be interpreted as follows. Suppose that the total number of users in the system is large. Then there will be a large number of admissible sets of size μ, assuming that the users are sufficiently dispersed

across the network and not concentrated in a few dense areas. In fact, the number of admissible sets of size μ will be overwhelmingly large compared to the number of admissible sets of smaller size. By virtue of random-admissible set scheduling, one of the admissible sets of size μ will then be selected with high probability. Thus, the expected number of removed packets will exceed the expected number of arriving packets, provided $\lambda\beta < \mu$, implying a reduction in the expected number of packets in the system, and preventing the number of packets from growing without bound.

As the above heuristic explanation indicates, it is crucial for the users to be sufficiently spread out and not be clustered in a few hot spots. In order to obtain a rigorous proof, it will hence not suffice to just consider the total number of users, but in fact be necessary to keep track of their individual locations.

Also, while the spatial dispersion of users under random admissible-set scheduling is intuitively plausible, it is certainly not obvious. This is perhaps best illustrated by the fact that the result of Theorem 1 may not necessarily hold for seemingly similar but subtly different scheduling disciplines.

As an example consider a unit circle to which users arrive according to a spatial Poisson process with parameter λ. Every user has exactly one packet to transmit, i.e. $\beta = 1$. Further, take the protocol model with reuse distance r as the model for interference, with r such that $1/r$ is non-integer. For random-admissible set scheduling we see by Theorem 1 that this system is stable if $\lambda < \lfloor 1/r \rfloor$.

Instead of selecting an admissible set at random, now consider the scheduling discipline that gives priority to users that have minimal distance, in anticlockwise direction, to a certain point ζ on the circle. Obeying this priority rule, we select as many users as possible to transmit in a certain time slot. That is, the user closest to the given point gets to transmit, the user closest to the point and at least a distance r away from the first user gets to transmit, and so on until no user can be selected anymore. We call this service discipline maximal scheduling with priorities.

Figure 1 shows a simulation result for both scheduling disciplines with $\zeta = 0.5$, $r = 0.49$ and $\lambda = 1.95$ starting from an empty configuration and running for 10^6 time slots. That is, the figure gives a realization of $Y(t, \mathbf{1})$ for $t = 1, \ldots, 10^6$ given that $Y(0, \mathbf{1}) = 0$. We see that at the start of the simulation the number of packets in the system with random-admissible set scheduling grows faster than the number of packets in the system with maximal scheduling with priorities. This is because maximal scheduling with priorities always selects a subset of maximum size obeying the priority rules, whereas random-admissible set scheduling may select admissible sets of a small size with a certain probability, which gets lower as the number of packets in the system grows. More importantly, after some time the number of packets in the system with random-admissible set scheduling settles around an equilibrium value whereas the number of packets in the system with maximal scheduling with priorities keeps on growing linearly. This suggests that maximal scheduling with priorities is not stable while random-admissible set scheduling is stable for the chosen parameters. The latter will be proven in the next section.

Figure 2 shows the terminal configuration of the simulation of Figure 1, i.e. it gives a realization of $Y(10^6, s)$ for $0 < s \leq 1$ given that $Y(0, \mathbf{1}) = 0$ for both scheduling dis-
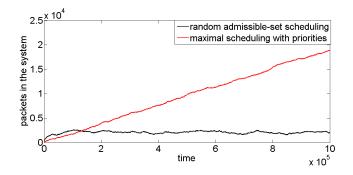
Figure 1: Random-admissible set scheduling and maximal scheduling with priorities for $\zeta = 0.5$, $r = 0.49$ and $\lambda = 1.95$.
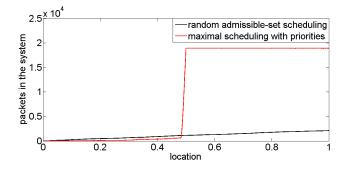

Figure 2: Terminal configuration of the simulation of Figure 1.

ciplines. For random admissible-set scheduling we see that the number of packets in the interval $[0, s)$ is roughly linear in s, indicating that the packets are evenly spread out over the circle. For maximal scheduling with priorities we observe that the number of packets in $[0, s)$ slowly increases with s up to approximately $s = 0.48$, after which the number of packets in the system steeply rises up to $s = 0.5$. For $s \geq 0.5 = \zeta$ the number of packets in the interval $[0, s)$ is (almost) constant, implying that virtually no packets are located in the interval $[0.5, 1)$. Note that users in the interval $[0.48, 0.5)$ have the lowest priority and hence will, whenever they are allowed to, almost always transmit simultaneously with other users, as there are quite some users in the system and outside this interval. However, the users that are allowed to transmit simultaneously with users in $[0.48, 0.5)$ are also allowed to transmit simultaneously with some users in $[0.5, 0.52)$, who have the highest priority. So we infer that the packets in this system are clustered in $[0.48, 0.5)$, and that too high a fraction of the time (larger then 0.02λ) no user in this interval is allowed to transmit a packet, making the system unstable.

4. PROOF OUTLINE

In this section we provide an outline of the proof of Theorem 1. As mentioned earlier, the proof relies on the Foster-Lyapunov criteria and involves the identification of a function which has negative drift for all but a small set of states. Appendix A contains several useful definitions and preliminaries that are needed to apply the Foster-Lyapunov approach for our specific Markov chain. The proofs of the

various lemmas in this section are deferred to Appendix B.

In order to define the Lyapunov function, we partition the torus using K^n hypercubes of size $1/K \times \cdots \times 1/K$. We choose K such that K is a multiple of μ and such that $D(x, w) < \delta$ for x, w inside the same hypercube of size $1/K \times \cdots \times 1/K$. Here $\delta > 0$ is chosen such that Assumption 1 holds. Using hypercubes is not strictly necessary here, and any other partitioning having the same properties could be used as well. However, using hypercubes is convenient as we modeled the torus by the unit hypercube.

We will assume that interference is modeled by the protocol or SINR model and that the parameters of the models are chosen such that there exists a $\delta > 0$ for which Assumption 1 holds. Note that for both the protocol model and the SINR model it holds that no two users can transmit simultaneously in the same hypercube in the same time slot.

Let $x_k(y)$ be the number of packets residing in the k-th region given configuration $y \in \Psi$, where the k-th region is the hypercube

$$H_k = \left[\frac{k-1 \bmod K}{K}, \frac{(k-1 \bmod K) + 1}{K} \right) \times$$
$$\cdots \times \left[\frac{\lfloor \frac{k-1}{K^{n-1}} \rfloor \bmod K}{K}, \frac{(\lfloor \frac{k-1}{K^{n-1}} \rfloor \bmod K) + 1}{K} \right),$$

for $k \in \mathcal{K}$, where $\mathcal{K} = \{1, \ldots, K^n\}$.

Let $\Omega = \mathcal{P}(\mathcal{K})$ be the collection of all subsets of $\{1, \ldots, K^n\}$ and let $\Omega(y)$ be the subsets containing packets, i.e. $\Omega(y) = \{S \in \Omega : x_k(y) \geq 1, \forall k \in S\}$. A subset $S \in \Omega$ is called 'guaranteed' if any subset of packets, with exactly one packet residing in each of the regions contained in S, is admissible, regardless of the exact locations within each of the regions. Let $\Theta \subseteq \Omega$ be the collection of all 'guaranteed' subsets and, again, let $\Theta(y)$ be the subsets containing packets, $\Theta(y) = \{S \in \Theta : x_k(y) \geq 1, \forall k \in S\}$. Further, denote by $q_S(y)$ the probability that the packets that get removed belong to the subset of regions $S \in \Omega$ given that the system is in state y, with $p_k(y) = \sum_{S \in \Omega : S \ni k} q_S(y)$ the probability that a packet in the k-th region gets removed given configuration $y \in \Psi$.

For any $S \in \Omega$, denote

$$w_S(y) = \prod_{k \in S : x_k(y) \geq 1} x_k(y).$$

Further define

$$B(\epsilon) = \left\{ y \in \Psi : w(y) \geq \left(\frac{2|\Omega|}{\epsilon} \right)^{2/\epsilon} \right\},$$

where $\epsilon > 0$ and

$$w(y) = \max_{S \in \Theta(y)} w_S(y).$$

The next lemma shows that the value of $\sum_{k \in S} \log(x_k)$, for the set of users S selected with random admissible-set scheduling, is close to the maximum possible value over all admissible sets with high probability for all states $y \in B(\epsilon)$ if interference is modeled by the protocol or SINR model.

LEMMA 2. *For all states $y \in B(\epsilon)$ we have*

$$\sum_{S \in \Omega(y)} q_S(y) \log(w_S(y)) \geq (1 - \epsilon) \log(w(y)).$$

For all $i \in \mathcal{K}$ define the set $S_i \in \Theta$ such that $i \in S_i$, $j \in S_i$ if and only if $i \in S_j$ and $|S_i| = \mu$. Note that such a set exists for some $i \in \mathcal{K}$ by Assumption 1 and the definition of K. Further note that for both the protocol model and the SINR model for interference such a set then exists for all $i \in \mathcal{K}$ as for these models moving all packets in a set in the same direction over the same distance does not change the classification of that set. Further define $S_i(y) = \{j \in S_i(y) : x_j(y) \geq 1\}$ and note that $S_i(y) \in \Theta(y)$.

Also define the nonnegative function $V : \Psi \to \mathbb{R}$ by

$$V(y) = \sum_{k:x_k(y) \geq 1} x_k(y) \log(x_k(y))$$

and the function $G : \Psi \to \mathbb{R}$ by

$$G(y) = \sum_{k:x_k(y) \geq 1} \log(x_k(y)) \left[\frac{\lambda \beta}{K^n} - p_k(y) \right].$$

Note that

$$V(y) = \sum_{k=1}^{K^n} x_k(y) \log(\max(x_k(y), 1))$$

and

$$V(y) = \sum_{k:x_k(y) \geq 2} x_k(y) \log(x_k(y)).$$

Further observe that the function $V(\cdot)$ only depends on y through the values of $x_k(y)$. However, the $x_k(y)$'s do not constitute a Markov chain, and hence we need to treat $V(\cdot)$ as a function of the full state description y in order for the Foster-Lyapunov approach to apply.

We now first find the relation between the drift of $V(y)$ and $G(y)$, where the drift of $V(y)$ is defined by

$$\Delta V(y) = \mathbb{E}\{V(Y(t+1))|Y(t) = y\} - V(y).$$

After that we will find an upper bound for $G(y)$.

LEMMA 3. $\Delta V(y) = G(y) + G_2(y)$, with $G_2(y)$ a bounded function.

LEMMA 4. Assume $\lambda \beta < \mu$. Then, for all states $y \in B(\epsilon)$ with $\epsilon = \frac{1}{2}(1 - \frac{\lambda \beta}{\mu})$,

$$G(y) \leq -\frac{\epsilon \mu}{K^n} \sum_{k:x_k(y) \geq 1} \log(x_k(y)).$$

By Lemma 3 we know that $\Delta V(y) \leq G(y) + G_2^{\max}$, where $G_2^{\max} = \sup_{y \in \Psi} G_2(y) < \infty$. Now consider the set C where the drift of $V(y)$ might be bigger than -1, i.e. consider

$$C = \{y \in \Psi : G(y) \geq -G_2^{\max} - 1\}.$$

The next lemma shows that this set is *small* (see Definition 3).

LEMMA 5. Assume $\lambda \beta < \mu$. Then, the set C is small.

Using the previous lemmas we can now show stability.

THEOREM 6. Assume $\lambda \beta < \mu$. Then the Markov chain $(Y(t))_{t \in \mathbb{N}}$ is positive Harris recurrent with invariant probability measure π and

$$\pi(f) = \int \pi(dx) f(x) < \infty,$$

where

$$f(y) = \begin{cases} -G(y) - G_2^{\max} & \text{for } y \notin C, \\ 1 & \text{for } y \in C, \end{cases}$$

Moreover,

$$\lim_{t \to \infty} \mathbb{E}\{g(Y(t))|Y(0) = y\} = \int \pi(dx) g(x), \quad \forall y \in \Psi,$$

for any function g satisfying $|g(x)| \leq c(f(x) + 1)$ for all x and some $c < \infty$.

PROOF. In Lemma 9 we have proven that the Markov chain satisfies the irreducibility and aperiodicity properties of Theorem 7. Further, we have proven in Lemma 5 that the set C is small. Thus, as $f \geq 1$ by construction and V is nonnegative and finite everywhere, we need to show that

$$\Delta V(y) \leq -f(y) + b\mathrm{I}_C(y), \quad \forall y \in \Psi, \qquad (3)$$

for some constant $b \in \mathbb{R}$, in order to prove our claim.

For $y \notin C$ we get

$$\Delta V(y) \leq G(y) + G_2^{\max},$$

which holds as we have shown in Lemma 3.

For $y \in C$ we get

$$\Delta V(y) \leq -1 + b.$$

We therefore take $b \geq 1 + \sup_{y \in C} \Delta V(y)$, so that the inequality holds by construction.

Hence we have shown that (3) holds for all $y \in \Psi$, proving our claim. $\square$

Remark 1. Remember that at the beginning of this section we assumed Assumption 1 to hold and that interference is modeled by the protocol or SINR model, introduced in Subsections 2.1 and 2.2, respectively. Further note that Lemma 9 shows that the Markov chain is φ-irreducible, where φ is the Dirac measure on Ψ assigning unit mass to the empty configuration y_0. Hence, by Definition 2, $\psi(\{y_0\}) > 0$. Theorem 1 then follows by the definition of a Harris recurrent chain, see Definition 5.

Remark 2. The Foster-Lyapunov approach may also be leveraged to derive an upper bound for the expected value of functions of the total number of packets in the system.

Specifically, define $J(y) = \sum_{k:x_k(y) \geq 1} \log(x_k(y))$, and denote $G_1^{\max} = \sup_{y \in B(\epsilon)^c} G(y) + J(y) < \infty$. By virtue of Lemma 4 we then have $G(y) \leq G_1^{\max} - \frac{\epsilon \mu}{K^n} J(y)$.

Using Lemma 3 and taking expectations yields

$$\mathbb{E}\{V(Y(t+1))\} - \mathbb{E}\{V(Y(t))\}$$
$$\leq -\frac{\epsilon \mu}{K^n} \mathbb{E}\{J(Y(t))\} + G_1^{\max} + G_2^{\max},$$

for all $t = 1, 2, \ldots$.

Summing over $t = 1, \ldots, T$, we obtain

$$\frac{\epsilon \mu}{K^n} \sum_{t=1}^{T} \mathbb{E}\{J(Y(t))\} \leq \mathbb{E}\{V(Y(1))\} + T(G_1^{\max} + G_2^{\max}),$$

and thus

$$\lim_{T \to \infty} \frac{\epsilon \mu}{TK^n} \sum_{t=1}^{T} \mathbb{E}\{J(Y(t))\} \leq G_1^{\max} + G_2^{\max}.$$

5. CONCLUDING REMARKS

We examined the stability of wireless networks whose users are distributed over a torus. Users arrive at spatially uniform locations with intensity λ and each have a random number of packets to transmit with mean β. In each time slot, an *admissible* subset of users is selected uniformly at random to transmit one packet, as governed by the prevailing interference constraints. We considered the SINR model and the protocol model as two canonical models for interference, and showed that the *necessary* condition $\lambda\beta < \mu$ is also *sufficient* for random admissible-set scheduling to achieve stability, with μ denoting the maximum number of users in an admissible subset.

We observed that the critical property of these interference models was that moving all packets in a set in the same direction over the same distance does not change the classification of that set, so that random admissible-set scheduling achieves stability, if feasible to do so at all, for other interference models satisfying this property as well.

In the present paper we focused on a torus and assumed the arrival process to be spatially uniform. Extensions to more general spaces and non-uniform arrival densities are subjects of current research.

For the sake of convenience, we assumed that the scheduling procedure selects uniformly at random among all admissible subsets of packets. A similar approach can however be applied to prove maximum stability of various modifications of the selection mechanism. We did also demonstrate though that the necessary condition is not sufficient to achieve stability for seemingly similar but subtly different scheduling disciplines such as maximal scheduling with priorities. An interesting topic for further research is to further demarcate the class of scheduling disciplines for which the necessary condition is also sufficient for stability.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] E. Altman and H. Levy. Queueing in space. *Adv. Appl. Prob.*, 26:1095–1116, 1994.

[2] T. Bonald and M. Feuillet. On the stability of flow-aware CSMA. *Perf. Eval.*, 67(11):1219–1229, 2010. (Proc. Performance 2010 Conf., Namur, Belgium, November 15–19).

[3] C. Bordenave, S. Foss, and V. Shneer. A random multiple-access protocol with spatial interactions. *J. Appl. Prob.*, 46:844–865, 2010.

[4] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition*. Springer-Verlag, New York, 2003.

[5] P. Gupta and P. Kumar. The capacity of wireless networks. *IEEE Trans. Inf. Theory*, 46(2):388–404, 2000.

[6] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, 1997.

[7] L. Leskelä and F. Unger. Stability of a spatial polling system with greedy myopic service. *Ann. Oper. Res.*, to appear, 2010.

[8] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

[9] J. Ni, B. Tan, and R. Srikant. Q-CSMA: queue length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks. In *Proc. IEEE Infocom 2010*, 2010. San Diego CA, March 15–19.

[10] P. Robert. The evolution of a spatial stochastic network. *Stochastic Processes and Applications*, 120:1342–1363, 2010.

[11] X. Wu, R. Srikant, and J. Perkins. Queue-length stability of maximal greedy schedules in wireless networks. In *Proc. Inf. Theory Appl. Inaugural Workshop*, San Diego CA, 2006.

APPENDIX

A. DEFINITIONS AND PRELIMINARIES

As mentioned, the stability proof relies on a Foster-Lyapunov approach. In this appendix, we first recall various relevant definitions and a result from Meyn and Tweedie [8]. After that, we prove that our Markov chain $(Y(t))_{t\in\mathbb{N}}$ introduced in Section 2 satisfies all technical conditions for the Foster-Lyapunov approach to apply.

Let $(\hat{Y}(t))_{t\in\mathbb{N}}$ be a Markov chain with state space $\hat{\Psi}$. Further, let $\mathcal{B}(\hat{\Psi})$ be the σ-field of subsets of $\hat{\Psi}$. This σ-field is assumed to be countably generated, i.e. it is generated by some countable class of subsets of $\hat{\Psi}$.

Definition 1. $(\hat{Y}(t))_{t\in\mathbb{N}}$ is said to be φ-irreducible if there exists a measure φ on $\mathcal{B}(\hat{\Psi})$ such that, whenever $\varphi(C) > 0$, we have

$$\mathbb{P}\left\{\min(t : \hat{Y}(t) \in C) < \infty | \hat{Y}(0) = \hat{y}\right\} > 0, \quad \forall \hat{y} \in \hat{\Psi}.$$

Let $\mathbb{P}^m\{\hat{y}, C\}$ denote the m-step transition probability to go from state $\hat{y}$ to the set $C \in \mathcal{B}(\hat{\Psi})$ Further define the transition kernel

$$K_{\frac{1}{2}}(\hat{y}, C) = \sum_{m=0}^{\infty} \mathbb{P}^m\{\hat{y}, C\} 2^{-(m+1)}, \quad \hat{y} \in \hat{\Psi}, C \in \mathcal{B}(\hat{\Psi}).$$

Definition 2. $(\hat{Y}(t))_{t\in\mathbb{N}}$ is said to be ψ-irreducible if it is φ-irreducible for some φ and the measure ψ is a maximal irreducibility measure, i.e. it satisfies the following conditions:

1. For any other measure ϕ' the chain is ϕ'-irreducible if and only if $\psi(C) = 0$ implies $\phi'(C) = 0$.

2. If $\psi(C) = 0$, then

$$\psi(\{\hat{y} : \mathbb{P}\left\{\min(t : \hat{Y}(t) \in C) < \infty | \hat{Y}(0) = \hat{y}\right\} > 0\}) = 0.$$

3. The probability measure ψ is equivalent to

$$\psi'(C) = \int_{\hat{\Psi}} \varphi'(d\hat{y}) K_{\frac{1}{2}}(\hat{y}, C),$$

for any finite measure φ' such that the chain is ϕ'-irreducible.

Note that by [8, Thm. 4.0.1] we know that if there exists a measure φ such that the chain is φ-irreducible, then there exists an (essentially unique) maximal irreducibility measure ψ.

Definition 3. A set $C \in \mathcal{B}(\hat{\Psi})$ is called ξ_m-small if there exists an $m > 0$ and a non-trivial measure ξ_m on $\mathcal{B}(\hat{\Psi})$, such that

$$\mathbb{P}^m\{\hat{y}, D\} \geq \xi_m(D), \quad \forall \hat{y} \in C, D \in \mathcal{B}(\hat{\Psi}).$$

A set is called small if it is ξ_m-small for some $m > 0$ and some non-trivial measure ξ_m.

Definition 4. Suppose $(\hat{Y}(t))_{t \in \mathbb{N}}$ is φ-irreducible. The chain is called strongly aperiodic when there exists a ξ_1-small set C with $\xi_1(C) > 0$.

Definition 5. A ψ-irreducible chain $(\hat{Y}(t))_{t \in \mathbb{N}}$ is said to be Harris recurrent if for all $C \in \mathcal{B}(\hat{\Psi})$ such that $\psi(C) > 0$ we have

$$\mathbb{P}\left\{\sum_{t=1}^{\infty} \mathrm{I}_C(Y(t)) = \infty \Big| \hat{Y}(0) = \hat{y}\right\} = 1, \quad \forall \hat{y} \in C,$$

where $\mathrm{I}_C(y)$ denotes the indicator function of the set C.

If a Harris recurrent chain admits an invariant probability measure it is called positive Harris recurrent.

The following theorem follows from Chapter 14 in [8].

THEOREM 7. *Suppose that the chain $(\hat{Y}(t))_{t \in \mathbb{N}}$ is ψ-irreducible and strongly aperiodic. If there exists some small set $\hat{C}$, a function $\hat{f} \geq 1$ and some nonnegative function $\hat{V}$ that is finite everywhere such that*

$$\Delta \hat{V}(\hat{y}) \leq -\hat{f}(\hat{y}) + \hat{b}\mathrm{I}_{\hat{C}}(\hat{y}), \quad \forall \hat{y} \in \hat{\Psi}, \tag{4}$$

then $(\hat{Y}(t))_{t \in \mathbb{N}}$ is positive Harris recurrent with invariant probability measure π and

$$\pi(\hat{f}) = \int \pi(dx)\hat{f}(x) < \infty.$$

Moreover,

$$\lim_{t \to \infty} \mathbb{E}\left\{\hat{g}(\hat{Y}(t)) | \hat{Y}(0) = \hat{y}\right\} = \int \pi(dx)\hat{g}(x), \quad \forall \hat{y} \in \hat{\Psi},$$

for any function $\hat{g}$ satisfying $|\hat{g}(x)| \leq \hat{c}(\hat{f}(x) + 1)$ for all x and some $\hat{c} < \infty$.

To prove that our Markov chain fulfills the conditions in Theorem 7, we first show that our σ-field, $\mathcal{B}(\Psi)$, is countably generated.

LEMMA 8. *$\mathcal{B}(\Psi)$ is the Borel σ-field. Furthermore, $\mathcal{B}(\Psi)$ is countably generated.*

PROOF. In this proof we will use some definitions and results in measure theory, see [4] and [6] for more details.

First note that H endowed with the topology generated by the open orthotopes defines a complete separable metric space. Then it follows by [4, Thm. A2.6.III] that the Borel σ-field of Ψ is the smallest σ-field with respect to which the map $y \to y(B)$ is measurable for any Borel set $B \subseteq H$. Further it follows that Ψ endowed with the *vague topology* is a complete separable metric space. The vague topology is the topology generated by the mappings $\phi \to \phi g = \int g d\phi$, with ϕ a measure on Ψ, for all continuous functions $g : H \to \mathbb{R}^+$ with compact support.

Since the space is separable, there exists a countable dense set $\mathcal{D}$ in this space. Let $\mathcal{S}_0$ be the class of all finite intersections of all open sets $\{x \in H : D(x, d) < r\}$, with $d \in \mathcal{D}$ and $r \in \mathbb{Q}^+$. Then, by [4, Lemma A2.1.III], $\mathcal{S}_0$ is countable and generates the Borel σ-field, $\mathcal{B}(\Psi)$. $\square$

We will now prove that our Markov chain satisfies the irreducibility and aperiodicity properties of Theorem 7.

LEMMA 9. *$(Y(t))_{t \in \mathbb{N}}$ is φ-irreducible and strongly aperiodic, where φ is the Dirac measure on Ψ assigning unit mass to the empty configuration y_0. Moreover, the level sets of the form $L_m = \{y \in \Psi : y(H) \leq m\}, m > 0$, are small.*

PROOF. The proof proceeds along similar lines as in [7].

Consider an initial configuration y with $y(\mathbf{1}) = n \leq m$ packets, so $y \in L_m$. Then the probability that the system is empty after m time slots is greater than the probability that no packets arrive during the first m time slots times the probability that the n packets are served in the first m time slots. Thus, as in a non-empty configuration the probability that at least one packet is served in a time slot is at least $\frac{1}{2}$,

$$\mathbb{P}^m\{y, \{y_0\}\} \geq \mathbb{P}\{A = 0\}^m \left(\frac{1}{2}\right)^m,$$

which is greater than zero as $\mathbb{P}\{A = 0\} > 0$. This proves that $(Y(t))_{t \in \mathbb{N}}$ is φ-irreducible in this case, because $\varphi(D) > 0$ only when $y_0 \in D$.

Now, define the measure $\xi_m = \mathbb{P}\{A = 0\}^m \left(\frac{1}{2}\right)^m \varphi$. For this measure we have $\mathbb{P}^m\{y, D\} \geq \xi_m(D)$ for all $D \in \mathcal{B}(\Psi)$. So we see that L_m is ξ_m-small. Further, $\{y_0\}$ is ξ_1-small and $\xi_1(\{y_0\}) > 0$, thus $(Y(t))_{t \in \mathbb{N}}$ is strongly aperiodic. $\square$

B. PROOF DETAILS

In this appendix we provide the full proof details of the various lemmas used in establishing the main stability result as stated in Theorem 1.

Proof of Lemma 2

The proof proceeds along similar lines as in [2], [9].

For all states $y \in B(\epsilon)$ define

$$\Upsilon(y) = \{S \in \Omega(y) : \log(w_S(y)) \geq (1 - \frac{\epsilon}{2})\log(w(y))\}.$$

Then,

$$\sum_{S \in \Omega(y)} q_S(y) \log(w_S(y)) \geq (1 - \frac{\epsilon}{2})\log(w(y)) \sum_{S \in \Upsilon(y)} q_S(y). \tag{5}$$

For any $S \in \Omega$, let $v_S(y)$ be the number of admissible subsets of packets of size $|S|$ with exactly one residing in each of the regions contained in S, with the convention that $v_\emptyset(y) = 1$ for all $y \in \Psi$ as the empty set is admissible for interference modeled by the protocol or SINR model. Because every admissible subset can have at most one element residing in each region, there is exactly one $S \in \Omega$ for which this subset is counted in $v_S(y)$. Thus the total number of admissible subsets of packets is given by $\sum_{T \in \Omega} v_T(y)$ and, as an admissible subset of packets is selected uniformly at random,

$$q_S(y) = \frac{v_S(y)}{\sum_{T \in \Omega} v_T(y)} = \frac{v_S(y)}{\sum_{T \in \Omega(y)} v_T(y)}.$$

Further observe that $v_S(y) \leq w_S(y)$ for all $S \in \Omega(y)$, with equality for all $S \in \Theta(y)$. Thus,

$$\sum_{S \notin \Upsilon(y)} q_S(y) = \frac{\sum_{S \notin \Upsilon(y)} v_S(y)}{\sum_{T \in \Omega(y)} v_T(y)} \leq \frac{\sum_{S \notin \Upsilon(y)} w_S(y)}{\sum_{T \in \Theta(y)} w_T(y)}$$

$$\leq \frac{|\Omega| w(y)^{1 - \epsilon/2}}{w(y)} = |\Omega| w(y)^{-\epsilon/2}.$$

The latter quantity is less than $\epsilon/2$ for all states $y \in B(\epsilon)$, and thus

$$\sum_{S \in \Upsilon(y)} q_S(y) \geq 1 - \frac{\epsilon}{2}. \qquad (6)$$

Combining the lower bounds (5) and (6), we obtain

$$\sum_{S \in \Omega(y)} q_S(y) \log(w_S(y)) \geq (1 - \frac{\epsilon}{2}) \log(w(y))(1 - \frac{\epsilon}{2})$$

$$\geq (1 - \epsilon) \log(w(y))$$

for all states $y \in B(\epsilon)$. $\quad \square$

Proof of Lemma 3

Remember that at most one packet can be removed from each region in every time slot. We thus get, with $u_k = x_k(y) + A_k(t)$,

$$\mathbb{E}\{V(Y(t+1))|Y(t) = y\}$$

$$= \mathbb{E}\Big\{\sum_{k=1}^{K^n} x_k(Y(t+1)) \log(\max(x_k(Y(t+1)), 1))\Big|Y(t) = y\Big\}$$

$$= \sum_{k=1}^{K^n} \mathbb{E}\Big\{x_k(Y(t+1)) \log(\max(x_k(Y(t+1)), 1))\Big|Y(t) = y\Big\}$$

$$= \sum_{k:x_k(y)\geq 2} p_k(y)\mathbb{E}\{(u_k - 1)\log(u_k - 1)\}$$

$$+ \sum_{k:x_k(y)\geq 2} (1 - p_k(y))\mathbb{E}\{u_k \log(u_k)\}$$

$$+ \sum_{k:x_k(y)=1} p_k(y)\mathbb{P}\{A_k(t) \geq 1\} \mathbb{E}\{A_k(t) \log(A_k(t))|A_k(t) \geq 1\}$$

$$+ \sum_{k:x_k(y)=1} (1 - p_k(y))\mathbb{E}\{(1 + A_k(t))\log(1 + A_k(t))\}$$

$$+ \sum_{k:x_k(y)=0} \mathbb{P}\{A_k(t) \geq 1\} \mathbb{E}\{A_k(t) \log(A_k(t))|A_k(t) \geq 1\},$$

where $A_k(t)$ denotes the number of packets arriving in the k-th region during the t-th time slot. Remember that $\mathbb{E}\{A_k(t)\} = \lambda\beta/K^n$ and $\mathbb{E}\{A_k(t)\log(A_k(t))|A_k(t) \geq 1\} < \infty$. We further have

$$\sum_{k:x_k(y)\geq 2} p_k(y)\mathbb{E}\{(u_k - 1)\log(u_k - 1)\}$$

$$+ \sum_{k:x_k(y)\geq 2} (1 - p_k(y))\mathbb{E}\{u_k \log(u_k)\}$$

$$= \sum_{k:x_k(y)\geq 2} p_k(y)\mathbb{E}\Big\{(u_k - 1)\Big(\log(x_k(y)) + \log\Big(\frac{u_k - 1}{x_k(y)}\Big)\Big)\Big\}$$

$$+ \sum_{k:x_k(y)\geq 2} (1 - p_k(y))\mathbb{E}\Big\{u_k\Big(\log(x_k(y)) + \log\Big(\frac{u_k}{x_k(y)}\Big)\Big)\Big\}$$

$$= \sum_{k:x_k(y)\geq 2} (x_k(y) + \frac{\lambda\beta}{K^n} - p_k(y)) \log(x_k(y))$$

$$+ \sum_{k:x_k(y)\geq 2} p_k(y)\mathbb{E}\Big\{(u_k - 1)\log\Big(\frac{u_k - 1}{x_k(y)}\Big)\Big\}$$

$$+ \sum_{k:x_k(y)\geq 2} (1 - p_k(y))\mathbb{E}\Big\{u_k \log\Big(\frac{u_k}{x_k(y)}\Big)\Big\}.$$

Now notice that for constants $a \geq 0, b \geq 0, c > 0$

$$\mathbb{E}\Big\{(a + A_k(t))\log\Big(b + \frac{A_k(t)}{c}\Big)\Big\}$$

$$= \mathbb{E}\{(a + A_k(t))(\log(bc + A_k(t)) - \log(c))\}$$

$$\leq \mathbb{E}\{(a + A_k(t))(\log(bc + A_k(t)) - \log(c))|A_k(t) \geq 1\}$$

$$\leq \mathbb{E}\{(a + A_k(t))(bc + \log(A_k(t)) - \log(c))|A_k(t) \geq 1\}$$

$$= a(bc - \log(c)) + \mathbb{E}\{bcA_k(t) + a\log(A_k(t))|A_k(t) \geq 1\}$$

$$\quad + \mathbb{E}\{A_k(t)\log(A_k(t))|A_k(t) \geq 1\}$$

$$\leq a(bc - \log(c)) + \frac{abc\mathbb{E}\{A_k(t)\}}{1 - \mathbb{P}\{A_k(t) = 0\}}$$

$$\quad + \mathbb{E}\{A_k(t)\log(A_k(t))|A_k(t) \geq 1\}.$$

Thus, as $\mathbb{E}\{A_k(t)\}$ and $\mathbb{E}\{A_k(t)\log(A_k(t))|A_k(t) \geq 1\}$ are bounded, we find

$$\mathbb{E}\Big\{(a + A_k(t))\log\Big(b + \frac{A_k(t)}{c}\Big)\Big\} < \infty. \qquad (7)$$

Hence $\Delta V(y) = G(y) + G_2(y)$, with

$$G_2(y) =$$

$$\sum_{k:x_k(y)=1} p_k(y)\mathbb{P}\{A_k(t) \geq 1\} \mathbb{E}\{A_k(t)\log(A_k(t))|A_k(t) \geq 1\}$$

$$+ \sum_{k:x_k(y)=1} (1 - p_k(y))\mathbb{E}\{(1 + A_k(t))\log(1 + A_k(t))\}$$

$$+ \sum_{k:x_k(y)=0} \mathbb{P}\{A_k(t) \geq 1\} \mathbb{E}\{A_k(t)\log(A_k(t))|A_k(t) \geq 1\}$$

$$+ \sum_{k:x_k(y)\geq 2} p_k(y)\mathbb{E}\Big\{(u_k - 1)\log\Big(\frac{u_k - 1}{x_k(y)}\Big)\Big\}$$

$$+ \sum_{k:x_k(y)\geq 2} (1 - p_k(y))\mathbb{E}\Big\{u_k \log\Big(\frac{u_k}{x_k(y)}\Big)\Big\},$$

which is a bounded function by (7) and as $K < \infty$. $\quad \square$

Proof of Lemma 4

Since $\epsilon = \frac{1}{2}(1 - \frac{\lambda\beta}{\mu}) > 0$ we have

$$G(y) = \sum_{k:x_k(y)\geq 1} \log(x_k(y))\Big[\frac{(1 - 2\epsilon)\mu}{K^n} - p_k(y)\Big].$$

Now note that, for all states $y \in B(\epsilon)$, we may write

$$\sum_{k:x_k(y)\geq 1} \log(x_k(y))p_k(y)$$

$$= \sum_{k:x_k(y)\geq 1} \log(x_k(y)) \sum_{S \in \Omega : S \ni k} q_S(y)$$

$$= \sum_{S \in \Omega} q_S(y) \sum_{k \in S : x_k(y) \geq 1} \log(x_k(y))$$

$$= \sum_{S \in \Omega(y)} q_S(y) \log(w_S(y)).$$

Likewise, we may write

$$\sum_{k:x_k(y)\geq 1} \log(x_k(y))$$

$$= \frac{1}{\mu} \sum_{k:x_k(y)\geq 1} \log(x_k(y)) \sum_{i:S_i(y)\ni k} 1$$

$$= \frac{1}{\mu} \sum_{i=1}^{K^n} \sum_{k\in S_i(y)} \log(x_k(y))$$

$$= \frac{1}{\mu} \sum_{i:S_i(y)} \log(w_{S_i(y)}(y)).$$

Substitution of these two equalities in $G(y)$ gives

$$G(y) = -\frac{\epsilon\mu}{K^n} \sum_{k:x_k(y)\geq 1} \log(x_k(y))$$

$$+ (1-\epsilon)\Big[\sum_{i:S_i(y)} \frac{1}{K^n} \log(w_{S_i(y)}(y)) \Big]$$

$$- \sum_{S\in\Omega(y)} q_S(y) \log(w_S(y)).$$

Then, using Lemma 2 and recalling the fact that $S_i(y) \in \Theta(y)$, so that $w_{S_i(y)}(y) \leq w(y)$ for all $i = 1,\dots,K^n$, yields

$$G(y) \leq -\frac{\epsilon\mu}{K^n} \sum_{k:x_k(y)\geq 1} \log(x_k(y))$$

$$+ (1-\epsilon) \sum_{i:S_i(y)} \frac{1}{K^n} \log(w_{S_i(y)}(y)) - \log(w(y))$$

$$\leq -\frac{\epsilon\mu}{K^n} \sum_{k:x_k(y)\geq 1} \log(x_k(y)),$$

for all states $y \in B(\epsilon)$. $\square$

Proof of Lemma 5

Consider the sets

$$\hat{B}(\epsilon) = \left\{ y \in \Psi : x_k(y) \leq \left(\frac{2|\Omega|}{\epsilon}\right)^{2/\epsilon}, \forall k \in \mathcal{K} \right\},$$

with $\epsilon = \frac{1}{2}(1 - \frac{\lambda\beta}{\mu}) > 0$ and

$$\hat{C} = \left\{ y \in \Psi : \log x_k(y) \leq \frac{(G_2^{\max}+1)K^n}{\epsilon\mu}, \forall k \in \mathcal{K} \right\}.$$

We see that $B(\epsilon)^c \subseteq \hat{B}(\epsilon)$ as subsets of one region are guaranteed and thus $w(y) \geq x_k(y)$, for all $k \in \mathcal{K}$. Further we see that $C \setminus B(\epsilon)^c \subseteq \hat{C}$, as follows from the upper bound for $G(y)$ found in Lemma 4. Hence,

$$C \subseteq B(\epsilon)^c \cup (C \setminus B(\epsilon)^c) \subseteq \hat{B}(\epsilon) \cup \hat{C}.$$

Thus, as

$$\hat{B}(\epsilon) \cup \hat{C} = \{ y \in \Psi : x_k(y) \leq M, \forall k \in \mathcal{K} \}$$

with

$$M = \max\left(e^{\frac{(G_2^{\max}+1)K^n}{\epsilon\mu}}, \left(\frac{2|\Omega|}{\epsilon}\right)^{2/\epsilon} \right),$$

we find $C \subseteq L_{K^n M}$, where L_m is the level set, $L_m = \{ y \in \Psi : y(\mathbf{1}) \leq m \}$. We know by Lemma 9 that $L_{K^n M}$ is small and hence, by definition, C is small. $\square$