Submitted to the Annals of Applied Probability

ALOHA THAT WORKS

By S. Rajagopalan D. Shah J. Shin*

MIT

The popularity of *Aloha*(-like) algorithms for resolution of contention between multiple entities accessing common resources is due to their extreme simplicity and distributed nature. Example applications of such an algorithm include Ethernet and recently emerging wireless multi-access networks. For more than four decades, various researchers have established the inefficiency of (the known versions of) such algorithms to varying degrees in various setups. However, the question that has remained unresolved is that of designing an algorithm that is *essentially* as simple and distributed as Aloha while being efficient.

In this paper, we resolve this question successfully for a network of queues when contention is modeled through independent set constraints over the network graph. The work by Tassiulas and Ephremides (1992) suggests that an algorithm that schedules queues so that the summation of "weight" of scheduled queues is maximized subject to constraints, is efficient. However, implementing such an algorithm using Aloha like mechanism has remained a mystery. We design such an algorithm building upon a Metropolis-Hastings sampling mechanism along with selection of "weight" as an appropriate function of the queue size. The key ingredient in establishing the efficiency of the algorithm is a novel *adiabatic*-like theorem for the underlying queueing network, which may be of general interest in the context of dynamical systems.

1. Introduction. A multiple-access channel is a broadcast channel that allows multiple users to communicate with each other by sending messages onto the channel. If two or more users simultaneously send messages, then the messages interfere with each other (collide), and the messages are not transmitted successfully. The channel is not centrally controlled. Instead, the users need to use a contention-resolution protocol to resolve collisions. The popular Aloha protocol or algorithm was developed more than four decades ago to address this (e.g. see [1]). The key behind such protocols is using collision or busyness of the channel as a signal of congestion and then reacting to it appropriately.

Although the most familiar multiple-access channels are wireless multiple-access medium (a la IEEE 802.11 standards) and local-area networks (such as the Ethernet network) which are wired,

*Author names appear in the alphabetical order of their last names. All authors are with Laboratory for Information and Decision Systems, MIT. This work was supported in parts by NSF projects HSD 0729361, CNS 0546590, TF 0728554 and DARPA ITMANET project. Authors' email addresses: {vatsa, devavrat, jinwoos}@mit.edu

AMS 2000 subject classifications: Primary 60K20, 68M12; secondary 68M20

Keywords and phrases: Wireless multi-access, Markov chain, Mixing time, Aloha

multiple-access channels are being implemented using a variety of technologies including packet-radio, fibre-optics, free-space optics and satellite transmission (e.g. see [12]). These multiple-access channels are used for communication in many distributed networked systems, including emerging communication networks such as the wireless *mesh* networks [27].

Despite the long history and great importance of multi-access contention resolution protocols, the question of designing an efficient Aloha-like simple algorithm or protocol[1] has remained unresolved in complete generality so far even for one multiple-access channel. In this paper, we are interested in designing a distributed contention resolution protocol for a *network* of multiple-access channels in which various subsets of these network users (nodes) interfere with each other. For example, in a wireless network placed in a geographic area two users interfere with each other if they are near by and do not interfere if they are far apart. Such networks can be naturally modeled as queueing networks with contentions modeled through independent set constraints over the interference network graph. For this setup, we will design a simple randomized, Aloha-like, algorithm that is efficient. Indeed, as a special case, it resolves the classical multiple-access single broadcast channel problem as well.

1.1. *Related work.* Design and analysis of multiple-access contention resolution algorithms have been of great interest for four decades across research communities. Due to its long and rich history, it will be impossible for us to provide a complete history. We will describe a few of the related works that are closer to our results. An interested reader is referred to an excellent online survey of literature (until October 2002) on contention resolution that is maintained by Leslie Ann Goldberg [11].

The research on contention resolution in earlier work concentrated on a single broadcast channel while more recently it has been addressing the network version. First we start with literature on the protocol design for a single channel and then we will discuss the recent work on the protocol design for the network version.

1.1.1. *Single multiple-access channel protocols.* For a single broadcast channel, there are two broad classes of models that are considered: (1) The *Queue-free* model, where users arrive to the system over time with each user willing to transmit exactly one message. A user leaves the system once it has transmitted its message successfully. Here the goal is to keep the number of waiting users, i.e. number of messages to be transmitted, as small as possible. (2) The *Queuing* model, where the number of users is fixed but each user has messages arriving to it over time. The messages not transmitted are queued at the user where they arrive. The goal is to keep the total queue size over users as small as possible.

There is no single definition which is used to determine whether or not a contention-resolution protocol is "good" or "efficient" or "stable" but the basic idea is that the protocol should allow good enough utilization of the multiple-access channel. More formally, since typically such a system can be modeled as a Markov chain (or process) the capacity region of a protocol can be defined in terms of the problem parameters (say arrival rates of messages) for which the Markov

[1]In this paper, we will use words *protocol* and *algorithm* interchangeably.

chain is positive recurrent or positive Harris recurrent. See a discussion in the paper by Hajek and Ephremides [6].

Most research on the contention-resolution protocols has focussed on the design of simple protocols because they are easiest to implement and hence relevant to practice, as well as (in a sense) easiest to understand. The first such protocol is popularly known as *Aloha* (see [1]). In what follows, we will describe certain Aloha-like protocols that are classified based on certain key features.

The first class of protocols are the *age-based* protocols. In an age-based protocol [18] for a discrete-time system, a message that is head-of-the-line of a queue decides to transmit data in a given time slot with pre-determined probability $p_t, t \geq 0$ if it has been waiting to be transmitted for t time slots. In a queue-free model, each message is at the head-of-line of its own queue. In such an age-based protocol, the transmission probabilities $p_t, t \geq 0$, determine the protocol completely. One of the general results on the inefficiency of age-based protocols was established in a sequence of papers by Kelly and McPhee (see [18], [19] and [20]) which showed that in the queue-free model, for any given age-based protocol the "critical" arrival rate is 0 if and only if $\sum_{i \leq t} p_i = \omega(\log t)$. Here "critical" arrival rate is the maximal rate at which messages arrive in the queue-free model so that an infinite number of transmissions is possible. It does not imply that the number of messages in the system remains finite.

The second class of protocols, closely related to the age-based protocols, are *backoff* protocols. In a backoff protocol, a message that is head-of-the-line and has been unsuccessful in transmitting for t times so far, decides to transmit with a pre-determined probability p_t. Again, in the queue-free version of the model, the work by Kelly and MacPhee [19] implies that if $1/p_t$ scales sub-exponentially in t then the "critical" arrival rate is 0. Therefore, even for infinitely many transmissions, one should have p_t scaling exponentially. A specific selection of such $p_t = 2^{-t}$ is popularly known as binary exponential backoff protocol proposed in [23]. Using a very elegant coupling argument, Aldous [2] established that for every positive arrival rate, the binary exponential backoff protocol has $o(t)$ expected transmissions in t time slots. Thus, the system may have infinitely many transmission over infinite time horizon, but the rate at which messages are transmitted goes to 0. This led MacPhee [20] to pose the question of whether there exists a backoff protocol which is recurrent for some positive arrival rate in the queue-free model. In [10], Goldberg, Jerrum, Kannan and Paterson established that no backoff protocol is recurrent for rates larger than 0.42, i.e. the capacity of every backoff protocol is at most 0.42. For the queuing model, the backoff protocols were studied by Hastad, Leighton and Rogoff [15]. They showed that if there are N users with each having rate λ/N, then binary exponential backoff is unstable (i.e. the network Markov chain is not positive recurrent) if $\lambda > 0.568$. But, if the backoff probabilities are "polynomial" (i.e. $p_t = (1+t)^{-\alpha},\ \alpha > 1$) then it is stable (i.e. Markov chain is positive recurrrent) for any $\lambda < 1$.

Finally, there is an extensive literature on the full-sensing protocol in the context of the queue-free model. An important result in that context is due to Mosely and Humblet [26] which found existence of a "tree protocol" with capacity 0.48776. This was followed up by a sequence of results including that of Tsybakov and Likhanov [34] that established that no protocol can

achieve capacity higher than 0.568 in the queue-free model.

In summary, for a single multiple-access channel, in the queue-free model there is no algorithm that is known to have full capacity. For the queuing model, when all the nodes have exactly the same rate requirement then polynomial backoff protocol by Hastad et al. achieves the capacity. However, if rate requirement of all nodes were different (but their summation less than 1) then there is no known algorithm that is provably stable.

1.1.2. *Multiple-access network protocols.* The recent emergence of wireless multi-hop networks as a canonical architecture for an access network in a residential area and a metro-area network in a dense city has led to a lot of exciting activity in the context of distributed protocol design for multiple-access networks. Such a network can be modeled as a queuing network with each queue having an exogenous arrival process. The queues interfere or contend for transmission depending upon their proximity and communication protocol. Therefore, such a network can be modeled as a constrained queueing network where simultaneously transmitting queues (or nodes) at any time must be a valid *independent set* of the network interference graph (see section 2 for a detailed formal description). In such a network, we need an algorithm to determine such a schedule of simultaneously transmitting queues every time while satisfying the scheduling constraints.

Now ignoring the implementation concerns, i.e. not worrying about algorithm being simple and distributed, the work by Tassiulas and Ephremides [33] implies that an algorithm that selects queues that can be scheduled with maximum summation of their weights, where the weight of a queue is its queue-size, is stable as long as the arrival rate is in the capacity region. That is, such a maximum weight (MW) scheduling algorithm is optimal in terms of its capacity. However, implementing MW algorithm, i.e. finding maximum weighted independent set in the network interference graph in a distributed and simple manner is a daunting task. Ideally, one wishes to design a MW algorithm that is as simple as the random access protocols. This has led researchers to exploit two approaches: (1) design of random access algorithms with access probabilities that are arrival rate aware, and (2) design of distributed implementations of MW algorithms.

We begin with the first line of approach, where various researchers have addressed the question of designing efficient random access scheduling algorithms. Here the question boils down to finding appropriate channel access probabilities for head-of-line packets as a function of their local history (i.e. age or backoff). In most of the works along these lines, authors try to determine the *saturated* capacity region of the algorithm of interest. That is, assuming that all queues are backlogged infinitely (equivalently, queues never become empty), find the set of arrival rates for which the allocated service rate is at least as high as the arrival rate for each queue.

An exception to this is a recent exciting work by Bordenave, McDonald and Proutiére [3], where they study the capacity region of network in large (or mean field) limit of random access protocols with given access probabilities without assumption of the saturated system. On the flip side, this work provides an approximate characterization of the capacity region for a small network (hard to quantify exact approximation error for small network). Also, a fixed set of access probabilities is unlikely to work for all arrival rates in the capacity region. Therefore, to be able to support a larger capacity region, one needs to select access probabilities that should

be adjusted depending on system arrival process.

In an earlier work motivated by this concern, Marbach [21] as well as Eryilmaz, Marbach and Ozdaglar [22] considered the selection of access probabilities based on the arrival rates. In a certain asymptotic sense, they established that their rate-aware selection of the access probabilities allocate the rates to queues so that the allocated rates are no less than the arrival rates. A caveat of their approach was "saturated system" analysis and the goodness of the algorithm in an asymptotic sense.

Another work by Gupta and Stolyar [14] and Stolyar [32] considered random access algorithms where the access probabilities are determined as a function of the queue-sizes by means of solving an optimization problem in a distributed manner. This algorithm has certain throughput (pareto) optimality property. It should be noted that in principle this algorithm is distributed as the access probabilities are determined by means of an iterative algorithm for solving an optimization problem based on queue-sizes. However, they are not "simple enough" as determination of access probabilities for each time step requires a lot of "control overhead" due to the need for solving an optimization problem every time. More recently, work by Liu and Stolyar [17] adapted this approach for congestion-controlled multi-hop random-access network where the algorithm requires very minimal control overhead each time and leads to a throughput-optimal algorithm in the "saturated system". Finally, we take note of very recent work by Jiang and Walrand [16] that provides a rate-aware distributed algorithm to determine the access probabilities for a similar multi-hop congestion-controlled setup. Their approach differs from that of Stolyar and co-authors as it is a rate-aware algorithm and tries to solves a different optimization problem. Also, in [16] authors provide intuitive relation between their rate-aware algorithm and a queue-based algorithm – the class of algorithms of interest in this paper.

The second approach has been design of distributed implementation of the MW mechanism in a simple manner. The work by Modiano, Shah and Zussman [25] and its natural extension described in a survey by Shah [28] provides a totally distributed, simple *gossip* mechanism to find an approximate MW schedule each time. This algorithm is throughput optimal and like the standard MW algorithm does not require information about arrival rates and does not have the caveat of "saturated system" analysis. Thus, this result indeed proves that there exists a totally distributed efficient algorithm. However, in such an algorithm the selection of schedule involves computing summation of weights of network nodes in a distributed manner. In principle, this summation can be amortized over time leading to $O(1)$ control overhead per time-step. However, in authors' opinion (and should be clear to an informed reader that) this is merely (an important) proof-of-concept and lacks the necessary *elegance* and *simplicity* to be of practical use. Specifically, it is not *like* the random-access algorithm. We take note of a natural extension of this *gossip* based approach for designing distributed cross-layer algorithm for the optimal control of the multi-hop wireless network by Eryilmaz, Ozdaglar, Shah and Modiano [7]. Again, it should be noted that this serves as an important proof-of-concept of such distributed algorithm, but far from being useful in practice.

In summary, none of the random access based algorithms that are studied in the literature is truly throughput-optimal, as either there is an assumption of saturated system, knowledge

of arrival rate is required, or the capacity region is not the largest possible. The distributed gossip implementation of the MW algorithm, though provides the proof-of-concept of existence of a distributed, simple and throughput optimal algorithm; it is not as elegant (and hence not practicable) as random access based algorithms.

1.2. *Contributions.* As the main contribution of this paper, we design a throughput-optimal or stable[2] random-access algorithm for multiple-access in a network of queues where contention is modeled through the independent set constraint. Our random access algorithm is elegant, simple and in our opinion, of great practical importance. And indeed, it achieves the desired throughput optimality property by making the random access probabilities time-varying and a function of the queue-size. The key to efficiency of our algorithm lies in the careful selection of this function. To this end, first we observe that if queue-sizes were *fixed* then one can use Metropolis-Hastings based sampling mechanism to sample independent sets so that the sampled independent sets are a good approximation of the MW algorithm. As explained later in detail (or an informed reader may gather from literature), the Metropolis-Hastings based sampling mechanism is essentially a continuous time random access protocol. Therefore, for our purposes the use of Metropolis-Hastings sampler would suffice only if queue-sizes were *fixed*. But queue-sizes change essentially at unit rate and the time for Metropolis-Hastings to reach "equilibrium" can be much longer. Therefore, in essence the Metropolis-Hastings mechanism may never reach "equilibrium" and hence such an algorithm may perform very poorly.

We make the following crucial observations to resolve this issue: (1) if queue-sizes are changing slowly then the Metropolis-Hastings based mechanism is likely to reach "equilibrium" as in the case when queue-sizes were fixed, and (2) the MW algorithm is stable even with weight selected as a monotonically increasing function of queue-size (in this paper, we use a function $f(x) \sim \log\log x$). Therefore, even though queue-sizes may change at unit rate, one can choose an appropriate weight function which changes very slowly (i.e. f' is small). These observations suggest that if we design Metropolis-Hastings sampling mechanism to sample independent sets with weights defined as this slowly changing function of queue-size then it is likely that our network will always be in a state so that the random access algorithm based on Metropolis-Hastings method is essentially sampling independent sets as per the "correct" distribution all the time. We indeed establish this non-trivial desirable result — this is very much like a "robust probabilistic" version of the standard adiabatic theorem [4, 13] which states that *if a system changes in a reversible manner at an infinitesimally small rate, then it always remains in its ground state* (see statement of Lemma 13 and section 5.7 for precise details). As a consequence, we obtain a random access based algorithm under which the network Markov process is positive Harris recurrent or the system is stable or throughput-optimal. We strongly believe that the algorithmic and analytic methods of this paper will be of much broader interest for system design and analysis in general.

2. Network model and performance metric.

[2]The notion of stability is defined as positive recurrence or positive Harris recurrence of network Markov process in this paper.

2.1. *Notation.* We start with basic notation that will be useful throughout the paper. In this paper, we will reserve bold letters for vectors. For example, $\mathbf{u} = [u_i]_{i=1}^d$ denotes a d-dimensional vector. The index t will be reserved for continuous time and τ for discrete time. We will reserve $\mathbf{1}$ and $\mathbf{0}$ for vector of all 1s and all 0s. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, define

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{d} u_i v_i.$$

Given a function $\phi : \mathbb{R} \to \mathbb{R}$, by $\phi(\mathbf{u})$ we mean application of ϕ to $\mathbf{u}$ componentwise, i.e. $\phi(\mathbf{u}) = [\phi(u_i)]$. Unless stated otherwise, $\log(\cdot)$ will be natural logarithm. For any vector $\mathbf{u} = [u_i]$, define $u_{\max} = \max_i u_i$ and $u_{\min} = \min_i u_i$. For a probability vector $\pi \in \mathbb{R}_+^d$ on d elements, we will use notation $\pi = [\pi(i)]$ where $\pi(i)$ is the probability of $i, 1 \le i \le d$.

2.2. *An abstract model.* Our network is a collection of n queues. Each queue has a dedicated exogeneous arrival process through which new work arrives in the form of unit sized packets. Each queue can be potentially serviced at unit rate resulting in departures of packets from it upon completion of their unit service requirement. The network will be assumed to be *single-hop*, i.e. once work leaves a queue, it leaves the network. At first glance, this appears to be a strong limitation. However, as we discuss later in Section 3, the results, in terms of algorithm design and analysis, of this paper naturally extend to the case of multi-hop setup. The arrival process is assumed to be discrete time process for convenience. However, service or scheduling decisions are not as per slotted time. Equivalently, scheduling decisions are assumed to totally asychronous, thus preserving notion of "distributed decision making" to the full extent.

Let $t \in \mathbb{R}_+$ denote the (continous) time and $\tau = \lfloor t \rfloor \in \mathbb{N}$ denote the corresponding discrete time slot. Let $Q_i(t) \in \mathbb{R}_+$ be the amount of work in the ith queue at time t. We will throughout assume that queues adopt the First-Come-First-Serve (FCFS) policy for servicing packets. Therefore, $Q_i(t)$ denotes the size of the queue i in terms of number of packets in the queue at time t. For example, $Q_i(t) = 2.7$ implies that the head-of-line packet has received 0.3 amount of service and 2 packets are waiting behind it. For convenience of notation, define $Q_i(\tau) = Q_i(\tau^+)$, i.e. the queue-size measured in the very beginning of the time slot τ. Let $\mathbf{Q}(t), \mathbf{Q}(\tau)$ denote the vector of queue sizes $[Q_i(t)]_{1 \le i \le n}$, $[Q_i(\tau)]_{1 \le i \le n}$ respectively. Initially, time $t = \tau = 0$ and the system starts empty. That is, $\mathbf{Q}(0) = \mathbf{0}$.

As noted earlier, arrival process is assumed to be discrete time for convenience. Specifically, arrivals happen in terms of packets, each requiring unit amount of service. Let $A_i(\tau)$ denote the cumulative arrival process for queue i in time interval $[0, \tau]$. Specifically, we assume that arrivals happen at the end in each time slot, i.e. arrivals in time slot τ happen at time $(\tau + 1)^-$ and are equal to $A_i(\tau + 1) - A_i(\tau)$ number of packets. For simplicity, we assume that the arrival processes are independent across queues and $A_i(\cdot)$ is a Bernoulli processes with parameter λ_i for each i. That is, $A_i(\tau + 1) - A_i(\tau) \in \{0, 1\}$ and $\Pr(A_i(\tau + 1) - A_i(\tau) = 1) = \lambda_i$ for all i and τ. Denote the arrival rate vector as $\lambda = [\lambda_i]_{1 \le i \le n}$.

The queues are offered service as per scheduling constraint. Let $\boldsymbol{\sigma}(t) = [\boldsymbol{\sigma}_i(t)]$ denote the scheduling decision at time $t \in \mathbb{R}_+$, where $\sigma_i(t)$ denotes the service rate that queue i receives at time t. For simplicity, we assume that each queue can be serviced at rate 0 or 1 at any given time,

i.e. $\sigma_i(t) \in \{0,1\}$. Thus, schedule at any time corresponds to a selection of subset of queues that are provided service at unit rate. The scheduling constraints require that only certain subsets of queues can be chosen to be served at unit rate at each time. Let $\mathcal{S} \subset \{0,1\}^n$ denote the set of all feasible schedules, or the set of all simultaneously schedulable queues. Under thus described setup, the schedule $\boldsymbol{\sigma}(t)$ at time t is such that $\boldsymbol{\sigma}(t) \in \mathcal{S} \subset \{0,1\}^n$.

The queuing dynamics induced under the above described model can be summarized by the following equation: for any $0 \leq s < t$ and $1 \leq i \leq n$,

$$Q_i(t) = Q_i(s) - \int_s^t \sigma_i(r)\mathbf{1}_{\{Q_i(r)>0\}}dr + A_i(s,t),$$

where $A_i(s,t)$ denotes the cumulative arrival to queue i in time interval $[s,t]$ and $\mathbf{1}_{\{x\}}$ denotes indicator function

$$\mathbf{1}_{\{x\}} = \begin{cases} 1 & \text{if } x \text{ is 'true'} \\ 0 & \text{otherwise} \end{cases}$$

Here, we implicitly assume that the choice of $\boldsymbol{\sigma}(t), t \in \mathbb{R}_+$ is Lebegue integrable. Indeed, we are interested in simple algorithms and hence it is unlikely to expect non-integrable scheduling decisions. Finally, define the cumulative departure process $\boldsymbol{D}(t) = [D_i(t)]$, where

$$D_i(t) = \int_0^t \sigma_i(r)\mathbf{1}_{\{Q_i(r)>0\}}dr.$$

In this paper, we will restrict the treatment to a special case of the above described general "switched network" model. Specifically, we will assume that the feasible set of schedules, $\mathcal{S}$ arises due to *independent set* constraints over interference network graph. However, as we remark in Section 3 the algorithm design and their properties naturally extend to the general model described above.

2.3. *Wireless network.* Consider a network of n wireless transmission capable devices with the queue $Q_i(\cdot)$ hosted at the device or node i. Under any reasonable model of communication deployed in practice (e.g. 802.11 standards), in essence if two devices are close to each other and share a common frequency to transmit at the same time, there will be interference and data is likely to be lost. If the devices are far away, they may be able to simultaneously transmit with no interference. Thus the scheduling constraint here is that no two devices that might interfere with each other can transmit at the same time. This can be naturally modeled as an *independent set* constraint on a graph (called the *intereference graph*), whose vertices correspond to the devices, and where two vertices share an edge if and only if the corresponding devices would interfere when simultaneously transmitting. Specifically, let $G = (V, E)$ denote the network interference graph with $V = \{1, \ldots, n\}$ representing n nodes and

$$E = \{(i,j) : i \text{ and } j \text{ interfere with each other}\}.$$

Let $\mathcal{N}(i) = \{j \in V : (i,j) \in E\}$ denote the neighbors of node i. We assume that if node i is transmitting, then all of its neighbors in $\mathcal{N}(i)$ can "listen" to it. Let $\mathcal{I}(G)$ denote the set of all

independent sets of G, i.e. subsets of V so that no two neighbors are adjacent to each other. Formally,

$$\mathcal{I}(G) = \{\boldsymbol{\sigma} = [\sigma_i] \in \{0,1\}^n : \sigma_i + \sigma_j \leq 1 \text{ for all } (i,j) \in E\}.$$

Under this setup, the set of feasible schedules $\mathcal{S} = \mathcal{I}(G)$.

2.4. *Scheduling algorithm, performance metric.* We need an algorithm to select schedule $\boldsymbol{\sigma}(t) \in \mathcal{I}(G)$ (or more generally, $\boldsymbol{\sigma}(t) \in \mathcal{S}$) for all $t \in \mathbb{R}_+$. Thus, a scheduling algorithm is equivalent to scheduling choices $\boldsymbol{\sigma}(t), t \in \mathbb{R}_+$. From the perspective of network performance, we would like the scheduling algorithm such that the queues in network remain as small as possible given the arrival process. From the implementation perspective, we wish that the algorithm be simple and distributed, i.e. perform constant number of logical operations at each node (or queue) per unit time, utilize information only available locally at the node or obtained through a neighbor and maintain as little data structure as possible at each node.

First, we formalize the notion of performance. In the setup described above, we define capacity region $\mathcal{C} \subset [0,1]^n$ as the convex hull of the feasible scheduling set $\mathcal{I}(G) = \mathcal{S}$, i.e.

$$\mathcal{C} = \left\{ \sum_{\boldsymbol{\sigma} \in \mathcal{S}} \alpha_{\boldsymbol{\sigma}} \boldsymbol{\sigma} : \sum_{\boldsymbol{\sigma} \in \mathcal{S}} \alpha_{\boldsymbol{\sigma}} = 1, \text{ and } \alpha_{\boldsymbol{\sigma}} \geq 0 \text{ for all } \boldsymbol{\sigma} \in \mathcal{I}(G) \right\}.$$

The intuition behind this definition of capacity region comes from the fact that any algorithm has to choose schedule from $\mathcal{I}(G)$ each time and hence the time average of the 'service rate' induced by any algorithm must belong to $\mathcal{C}$. Therefore, if arrival rates $\boldsymbol{\lambda}$ can be 'served' by any algorithm then it must belong to $\mathcal{C}$.

Motivated by this, we call an arrival rate vector $\boldsymbol{\lambda}$ admissible if $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, where

$$\boldsymbol{\Lambda} = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^n : \boldsymbol{\lambda} \leq \boldsymbol{\sigma} \text{ componentwise, for some } \boldsymbol{\sigma} \in \mathcal{C} \right\}.$$

We say that an arrival rate vector $\boldsymbol{\lambda}$ is strictly admissible if $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}^o$, where $\boldsymbol{\Lambda}^o$ is the interior of $\boldsymbol{\Lambda}$ formally defined as

$$\boldsymbol{\Lambda}^o = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^n : \boldsymbol{\lambda} < \boldsymbol{\sigma} \text{ componentwise, for some } \boldsymbol{\sigma} \in \mathcal{C} \right\}.$$

Equivalently, we may say that the network is *underloaded*. Now we are ready to define the performance metric for a scheduling algorithm.

Definition 1 (Throughput optimal) *We call a scheduling algorithm* throughput optimal *or providing* 100% throughput *or* stable*, if for any $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}^o$ the underlying network Markov process is* Positive Harris Recurrent*.*

2.4.1. *Positive Harris recurrence & its implications.* For completeness, we define the well known notion of positive Harris recurrence (e.g. see [5]). We also state its useful implications to explain its desirability. In this paper, we will be concerned with discrete-time, time-homogeneous Markov process or chain evolving over a complete, separable metric space X. Let $\mathcal{B}_{\mathsf{X}}$ denote the Borel σ-algebra on X. Let $X(\tau)$ denote the state of Markov chain at time $\tau \in \mathbb{N}$.

Consider any $A \in \mathcal{B}_{\mathsf{X}}$. Define stopping time $T_A = \inf\{\tau \geq 1 : X(\tau) \in A\}$. Then the set A is called Harris Recurrent if

$$\Pr_x(T_A < \infty) = 1, \qquad \text{for any } x \in \mathsf{X},$$

where $\Pr_x(\cdot) \equiv \Pr(\cdot|X(0) = x)$. A Markov chain is called Harris recurrent if there exists a σ-finite measure μ on $(\mathsf{X}, \mathcal{B}_{\mathsf{X}})$ such that whenever $\mu(A) > 0$ for $A \in \mathcal{B}_{\mathsf{X}}$, A is Harris recurrent. It is well known that if X is Harris recurrent then an essentially unique invariant measure exists (e.g. see Getoor [9]). If the invariant measure is finite, then it may be normalized to obtain a unique invariant probability measure (or stationary probability distribution); in this case X is called positive Harris recurrent.

Now we describe a useful implication of positive Harris recurrence. Let π be the unique invariant (or stationary) probability distribution of the positive Harris recurrent Markov chain X. Then the following ergodic property is satisfied: for any $x \in \mathsf{X}$ and non-negative measurable function $f : \mathsf{X} \to \mathbb{R}_+$,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\tau=0}^{T-1} f(X(\tau)) \to \mathbb{E}_\pi[f], \quad \Pr_x\text{-almost surely.}$$

Here $\mathbb{E}_\pi[f] = \int f(z)\pi(z)$. Note that $\mathbb{E}_\pi[f]$ may not be finite.

2.5. *A popular algorithm.* In this paper, our interest is in scheduling algorithms that utilize the network state, i.e. the queue-size $\mathbf{Q}(t)$, to obtain schedule. An important class of scheduling algorithms with this property is the well known *maximum-weight* scheduling algorithm which was first proposed by Tassiulas and Ephremides [33]. We describe the slotted time version of this algorithm. In this version, the algorithm changes decision in the beginning of every time slot using $\mathbf{Q}(\tau) = \mathbf{Q}(\tau^+)$. Specifically, the scheduling decision $\boldsymbol{\sigma}(\tau)$ remains the same for the entire time slot τ, i.e. $\boldsymbol{\sigma}(t) = \boldsymbol{\sigma}(\tau)$ for $t \in (\tau, \tau + 1]$, and it satisfies

$$\boldsymbol{\sigma}(\tau) \in \arg\max_{\sigma\in\mathcal{S}} \sum_i \sigma_i Q_i(\tau).$$

Thus, this maximum weight or MW algorithm chooses schedule $\boldsymbol{\sigma} \in \mathcal{S}$ (with $\mathcal{S} = \mathcal{I}(G)$ for wireless network) that has the maximum weight, where weight is defined as $\boldsymbol{\sigma} \cdot \mathbf{Q}(\tau) = \sum_{i=1}^n \boldsymbol{\sigma}_i Q_i(\tau)$.

A generalized version of the MW algorithm, denoted by MW-f, picks a schedule with the maximum weight, where the weight of queue i is $f(Q_i(\tau))$ at time τ for some non-negative increasing function f with $f(0) = 0$. That is, under MW-f,

$$\sigma(\tau) \in \arg\max_{\sigma\in\mathcal{S}} \sum_i \sigma_i f(Q_i(\tau)).$$

It is well-known that MW-f algorithm is throughput optimal as long as $f : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies the following properties: (a) $f(0) = 0$, (b) f is strictly increasing and (c) $f(rx) = g(r)f(x)$ for any $r, x > 0$, where $g(r)$ is some function of r. To learn details, for example, see Shah and Wischik [29, 30]. Examples of such functions include $f(x) = x^\alpha$ for any $\alpha > 0$.

3. Main result: an efficient algorithm. This section presents the main result of this paper in terms of the efficient distributed scheduling algorithm. In what follows, we begin by describing the algorithm. Our algorithm is designed with the aim of approximating the maximum weight in a distributed manner. For our distributed algorithm to be efficient (or throughput optimal), it has to be good approximation of the maximum weight. As we shall establish, such is the case when the selection of weight function is done carefully. Therefore, first we describe the algorithm for a generic weight function. Next, we formally state the efficiency of the algorithm for a specific weight function. This is followed by some details for distributed implementation. Finally, we discuss extension of the algorithm for the multi-hop setting as well as a conjecture.

3.1. *Algorithm description.* As before, let $t \in \mathbb{R}_+$ denote the time. Let $\boldsymbol{W}(t) = [W_i(t)] \in \mathbb{R}_+^n$ denote the vector of weights at the n queues at time t. As we shall see, $\boldsymbol{W}(t)$ will be certain function of the queue-sizes $\mathbf{Q}(t)$. For the purpose of the algorithm description, one may assume $\boldsymbol{W}(t)$ as given. The algorithm we describe is a continuous time algorithm that wishes to compute schedule $\boldsymbol{\sigma}(t) \in \mathcal{I}(G)$ in a distributed manner so as to have weight $\sum_i \sigma_i(t)W_i(\tau)$ as large as possible.

The algorithm is randomized and asynchronous. Each node has an independent Exponential clock of rate 1. To this end, let T_k^i be the time when the clock of node i ticks for the kth time. Initially, $k = 0$ and $T_0^i = 0$ for all i. The $T_i^{k+1} - T_i^k$ are i.i.d. and have Exponential distribution of mean 1. The nodes change their scheduling decisions only upon their clock ticks. That is, $\sigma_i(t)$ remains constant for $t \in (T_i^k, T_i^{k+1}]$. Note that due to the property of continuous random variables, no two clock ticks at different nodes will happen at the same time with probability 1.

Let the algorithm start with null-schedule, i.e. $\boldsymbol{\sigma}(0) = [0] \in \mathcal{I}(G)$. Consider time T_k^i, the kth clock tick of node i for $k > 0$. Clearly, only node i's clock will tick at this particular time with probability 1. Now node i makes the following scheduling decision at this particular time instance $t = T_k^i$.

- If $\sigma_i(t^-) = 1$, then $\sigma_i(t^+) = 1$ with probability $\exp(W_i(t))/(1+\exp(W_i(t)))$ and $\sigma_i(t^+) = 0$ otherwise. This randomized decision is done independently of everything else.
- If $\sigma_i(t^-) = 0$, then node i "listens" to the medium. If any neighbor is transmitting, then $\sigma_i(t^+) = 0$. Else, $\sigma_i(t^+) = 1$ with probability $\exp(W_i(t))/(1 + \exp(W_i(t)))$ and $\sigma_i(t^+) = 0$ otherwise. Again, randomized decision is done independently of everything else.

We assume that if $\sigma_i(t) = 1$, then node i will always transmit data irrespective of the value of $Q_i(t)$ so that the neighbors of node i, i.e. nodes in $\mathcal{N}(i)$, can infer $\sigma_i(t)$ by "listening" to the medium.

3.2. *Efficiency of algorithm.* We describe a specific choice of weight $\boldsymbol{W}(t)$ for which the above described algorithm is throughput optimal for any network graph G. In what follows, let $f(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly concave monotonically increasing function with $f(0) = 0$. We will be interested in functions growing much slower than $\log(\cdot)$ function. Specifically, we will use the function $f(x) = \log\log(x + e)$ in our algorithm. For defining the weight, we will utilize two given constants $\varepsilon > 0, B > 0$. Here we will be interested in small ε and large B. Finally,

let $Q_{\max}(t) = \max_i Q_i(t)$ and let $\widetilde{Q}_{\max,i}(t)$ be an estimation of $Q_{\max}(t)$ at node i at time t. A straightforward algorithm to compute $\widetilde{Q}_{\max}(t)$ is described in section 3.3. As will be established in Lemma 2, $Q_{\max}(t) - 2n \leq \widetilde{Q}_{\max,i}(t) \leq Q_{\max}(t)$ for all i and $t > 0$. Now define the weight at node i,

$$W_i(t) = \max\left\{f(Q_i(\lfloor t \rfloor)), \frac{\varepsilon}{n} f(\widetilde{Q}_{\max,i}(\lfloor t \rfloor)), B\right\}. \tag{1}$$

For such a choice of weight, we state the following throughput optimality property of the algorithm.

Theorem 1 *Consider any $\varepsilon > 0$ and large enough B.*[3] *Suppose the algorithm use weight as defined in* (1) *with $f(x) = \log\log(x + e)$ and $|\widetilde{Q}_{\max,i}(t) - Q_{\max}(t)|$ be uniformly bounded by a constant for all t. Then, for any $\boldsymbol{\lambda} \in (1 - 2\varepsilon)\boldsymbol{\Lambda}^o$, the (appropriately defined) network Markov process is positive Harris recurrent. Further with respect to its stationary distribution,*

$$\mathbb{E}[f(\mathbf{Q}) \cdot \mathbf{1}] < \infty.$$

3.3. *Distributed implementation.* The goal here is to design an algorithm that is truly distributed and simple. That is, each node makes only *constant* number of operations locally each time, communicates only *constant* amount of information to its neighbors, maintains only *constant* amount of data structure and utilizes only local information. Further, we wish to avoid algorithms that satisfy the above properties by collecting some information over time. In essence, we want simple "Markovian" algorithms.

The algorithm described above, given the knowledge of node weight $W_i(\cdot)$ at node i for all i, does have these properties. Now the weight $W_i(\cdot)$ as defined in (1) depends on $Q_i(\cdot)$ and $Q_{\max}(\cdot)$ (or its estimate $\widetilde{Q}_{\max,i}(\cdot)$). Trivially, the $Q_i(\cdot)$ is known at each node. However, the computation of $Q_{\max}(\cdot)$ requires global information. Next, we describe a simple scheme in which each node maintains an estimate $\widetilde{Q}_{\max,i}(\cdot)$ at node i. To keep this estimate updated, each node broadcasts exactly one number to all of its neighbor every time slot. And, using the information received from its neighbors each time, it updates its estimate. We do not discuss the precise implementation of the information exchange required by this algorithm for two reasons: (1) We are interested in providing an abstract description of the algorithm and one may imagine various ways to implement such an algorithm (like any other wireless network protocol) over a separate 'control channel'. (2) In section 3.5, we provide a conjecture (supported by some experimental results) that the algorithm without the term corresponding to $\widetilde{Q}_{\max,i}(t)$ in (1) should be throughput optimal. Thus, our algorithm for finding $\widetilde{Q}_{\max,i}(t)$ (and its utilization) is merely for proving that a totally distributed, simple and provably throughput optimal algorithm does exist. In practice, we recommend the algorithm that is conjectured in section 3.5.

Now, we state the precise procedure to compute $\widetilde{Q}_{\max,i}(t)$, the estimate of $Q_{\max}(t)$ at node i at time t. It is updated once every time slot. That is, $\widetilde{Q}_{\max,i}(t) = \widetilde{Q}_{\max,i}(\lfloor t \rfloor)$. Suppose $\widetilde{Q}_{\max,i}(\tau)$ be

[3]The precise "large enough" value of B depends on the weight function f and number of nodes n as stated in the Definition 5. It should be noted (and hopefully a careful reader will notice) that the use of B is merely for keeping the proof simpler and it is not crucial for the correctness of Theorem 1.

the estimate of node i at time slot $\tau \in \mathbb{N}$. Then node i broadcasts this estimate to its neighbors at the end of time slot τ. Let $\widetilde{Q}_{\max,j}(\tau)$ for $j \in \mathcal{N}(i)$ be the estimates received by node i at the end of time slot τ. Then, update

$$\widetilde{Q}_{\max,i}(\tau+1) = \max\left\{\left(\max_{j\in\mathcal{N}(i)} \widetilde{Q}_{\max,j}(\tau)\right) - 1, \widetilde{Q}_{\max,i}(\tau) - 1, Q_i(\tau+1)\right\}.$$

We state the following property of this estimation algorithm.

Lemma 2 *As long as graph G is connected, for all $\tau \geq 0$ and all i,*

$$Q_{\max}(\tau) - 2n \leq \widetilde{Q}_{\max,i}(\tau) \leq Q_{\max}(\tau).$$

Proof. First, the proof of upper bound which follows by induction on τ. For this, note that initially $\widetilde{Q}_{\max,i}(0) = 0$ for all i and hence the upper bound holds for $\tau = 0$. Suppose it is true up to τ for all i. Consider $\tau+1$. Now since for each j, $Q_j(\tau+1) \geq Q_j(\tau) - 1$, we have $Q_{\max}(\tau+1) \geq Q_{\max}(\tau) - 1$. By induction hypothesis $\widetilde{Q}_j(\tau) \leq Q_{\max}(\tau)$ and $\widetilde{Q}_i(\tau) \leq Q_{\max}(\tau)$. And of course, $Q_i(\tau+1) \leq Q_{\max}(\tau+1)$. Therefore, it follows that $\widetilde{Q}_i(\tau+1) \leq Q_{\max}(\tau+1)$.

Next, the proof of lower bound. For this, note that for $\tau < n$ the $Q_{\max}(\tau) \leq n$ since at most one arrival can happen per time slot. Therefore, the lower bound follows in a straightforward manner. For $\tau \geq n$, consider a "breadth-first search tree" grown starting from node i of depth n as follows. The root node corresponds to node i with value $\widetilde{Q}_i(\tau)$, its first level children nodes correspond to nodes $j \in \mathcal{N}(i)$ with each having value $\widetilde{Q}_j(\tau-1)-1$, and then recursively the ℓth level children nodes of a node at level $\ell-1$ correspond to its neighbor in G (excluding those that already appeared up to level $\ell-1$) and a node k at level ℓ has value associated $\widetilde{Q}_k(\tau-\ell)-\ell$. It can be easily checked that the thus grown tree has depth at most n and all n nodes in G are present in this tree as long as G is connected. Further, under the algorithm described above the $\widetilde{Q}_{\max,i}(\tau)$ is larger than the values associated to all nodes in this tree. Now suppose k^* be the node such that $Q_{k^*}(\tau) = Q_{\max}(\tau)$. Let node k^* appear at some level $\ell \leq n$ in the tree. Now

$$\widetilde{Q}_{k^*}(\tau-\ell) \geq Q_{k^*}(\tau-\ell) \geq Q_{k^*}(\tau) - \ell,$$

because $Q_k(\cdot)$ can change at most by unit amount in one time slot. Therefore, it follows that

$$\begin{aligned}
\widetilde{Q}_{\max,i}(\tau) &\geq \widetilde{Q}_{k^*}(\tau-\ell) - \ell \\
&\geq Q_{k^*}(\tau-\ell) - \ell \\
&\geq Q_{k^*}(\tau) - 2\ell \\
&\geq Q_{\max}(\tau) - 2n.
\end{aligned}$$

This completes the proof of lower bound and that of Lemma 2. □

3.4. *Extensions.* The algorithm described here is for the single hop network with exogenous arrival process. As a reader will find, the key reason behind the efficiency of the algorithm is similar to the reason behind the efficiency of the standard maximum weight scheduling (here, the weight is $\log\log(\cdot)$ function of the queue-size). The standard maximum weight algorithm

has a known version for a general multi-hop network with choice of routing by Tassiulas and Ephremides [33]. This is popularly known as *back pressure* algorithm, where weight of an action of transferring a packet from node i to node j is determined in terms of the difference of queue-sizes at node i and node j. Analogously, our algorithm can be modified for such a setup by using the weight of an action of transferring a packet from node i to node j as the difference of $\log\log(\cdot)$ of queue-sizes at node i and node j. The corresponding changes in algorithm described in section 3.1 is strongly believed to be efficient using the similar proof method as that in this paper. More generally, there has been clever utilization of such a back-pressure approach in designing congestion control and scheduling algorithm in a multi-hop wireless network. For example, see survey by Shakkottai and Srikant [31]. Again, we strongly believe that utilization of our algorithm with appropriate weight will lead to a complete solution for congestion control and scheduling in a multi-hop wireless network.

3.5. *A conjecture.* The algorithm described for the single hop network utilizes the weight $W_i(t)$ defined as (1). This weight $W_i(t)$ depends on the queue-size of node i, $Q_i(\lfloor t \rfloor)$; the 'universal' constant B and the estimate of $Q_{\max}(t)$, $\widetilde{Q}_{\max,i}(t)$. Among these, the use of constant B and $\widetilde{Q}_{\max,i}(t)$ are primarily for 'technical' reasons. While the algorithm described here provides a provably random access like algorithm, we strongly believe that the algorithm that operates without the use of $\widetilde{Q}_{\max,i}(\cdot)$ and constant B in the weight definition should be efficient. Formally, we state our conjecture.

Conjecture 3 *Consider the algorithm described in section 3.1 with weight of node i at time t as*

$$W_i(t) \quad = \quad f(Q_i(\lfloor t \rfloor)). \tag{2}$$

Then, this algorithm is Positive Harris Recurrent as long as $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}^o$ and $f(x) = \log\log(x + e)$.

This conjecture is found empirically true in the context of a specific class of network graph topologies (grid graph). However, such a verification can only be accepted with partial faith.

4. Technical preliminaries. We present some known results about stationary distribution and convergence time (or mixing time) to stationary distribution for a specific class of finite state Markov chains known as Glauber dynamics (or Metropolis-Hastings). As a reader will find, these results will play an important role in establishing positive Harris recurrence of network Markov chain (evolving over a Polish space).

4.1. *Finite state Markov chain.* Consider a time homogeneous Markov chain over a finite state space Ω. Let the $|\Omega| \times |\Omega|$ matrix P be its transition probability matrix. If P is irreducible and aperiodic, then the Markov chain has unique stationary distribution and it is ergodic in the sense that $\lim_{\tau\to\infty} P^\tau(j,i) \to \pi_i$ for any $i, j \in \Omega$. Here $\pi = [\pi_i]$ denotes the stationary distribution of the Markov chain. The adjoint of transition matrix P, also called the time-reversal of P, is denoted by P^* and defined as: for any $i, j \in \Omega$

$$\pi(i)P^*(i,j) = \pi(j)P(j,i).$$

By definition, P^* has π as its stationary distribution. If $P = P^*$ then P is called *reversible*.

Our interest is in a specific irreducible, aperiodic Markov chain on the finite space $\Omega = \mathcal{I}(G)$, the set of independent sets of a given network graph $G = (V, E)$. This is also known as Glauber dynamics (or Metropolis-Hastings). We define it next.

Definition 2 (Glauber dynamics) *Consider a node weighted graph $G = (V, E)$ with $\boldsymbol{W} = [W_i]_{i \in V}$ be the vector of node weights. Let $\mathcal{I}(G)$ denote the set of all independent sets of G. Then the Glauber dynamics on $\mathcal{I}(G)$ with weights given by $\boldsymbol{W}$, denoted by $GD(\boldsymbol{W})$ is the following Markov chain. Suppose Markov chain is at state $\boldsymbol{\sigma} = [\sigma_i]_{i \in V}$, then the next transition happens as follows:*

- *Pick a node $i \in V$ uniformly at random.*
- *If $\sigma_i = 1$, then*
$$\sigma_i = \begin{cases} 1 & \text{with probability } \frac{\exp(W_i)}{1+\exp(W_i)} \\ 0 & \text{otherwise} \end{cases}.$$
- *If $\sigma_i = 0$ and $\sigma_j = 0$ for all $j \in \mathcal{N}(i)$, then*
$$\sigma_i = \begin{cases} 1 & \text{with probability } \frac{\exp(W_i)}{1+\exp(W_i)} \\ 0 & \text{otherwise} \end{cases}.$$
- *Otherwise, $\sigma_i = 0$.*

As the reader will notice, our algorithm described in section 3 is effectively an asynchronous version of the above described Glauber dynamics with time-varying weights. In essence, we will be establishing that even with asynchronous time-varying weights, the behavior of our algorithm will be very close to that of the Glauber dynamics with fixed weight in its stationarity. To this end, next we state a property of this Glauber dynamics in terms of its stationary distribution.

Lemma 4 *Let π be stationary distribution of $GD(\boldsymbol{W})$ on the space of independent sets $\mathcal{I}(G)$ of graph $G = (V, E)$. Then,*

$$\pi(\boldsymbol{\sigma}) = \frac{1}{Z} \exp(\boldsymbol{W} \cdot \boldsymbol{\sigma}) \cdot \mathbf{1}_{\boldsymbol{\sigma} \in \mathcal{I}(G)},$$

where Z is the normalizing factor.

Proof. Under $GD(\boldsymbol{W})$, there is a positive transition probability from independent set $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma'}$ if and only if they differ in exactly one coordinate, i.e. $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$ are *neighbors*. The set of independent sets, a subset of $\{0,1\}^n$ for $n = |V|$, has monotone structure: if a set is independent set then so is any subset. Therefore, $\mathbf{0} \in \{0,1\}^n$ is an independent set and is reachable to and from all independent sets under $GD(\boldsymbol{W})$. Thus, $GD(\boldsymbol{W})$ is irreducible. By definition $GD(\boldsymbol{W})$ is aperiodic. Therefore, it is a finite state ergodic Markov chain with unique stationary distribution. Next we establish that the stationary distribution is indeed π.

To this end, consider transitions between any two neighboring independent sets $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$. Without loss of generality, let $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$ differ in the ith co-ordinate with $\boldsymbol{\sigma}_i = 0$ and $\boldsymbol{\sigma}'_i = 1$. Let P denote

the transition matrix of $GD(\boldsymbol{W})$. As per the definition, the transition probability from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma'}$ is

$$P_{\boldsymbol{\sigma}\boldsymbol{\sigma'}} = \Pr\left(i \text{ was picked}\right) \frac{\exp(W_i)}{1+\exp(W_i)} = \frac{1}{n}\frac{\exp(W_i)}{1+\exp(W_i)}\,,$$

while the probability of the transition $\boldsymbol{\sigma'}$ to $\boldsymbol{\sigma}$ is

$$P_{\boldsymbol{\sigma'}\boldsymbol{\sigma}} = \Pr\left(i \text{ was picked}\right) \frac{1}{1+\exp(W_i)} = \frac{1}{n}\frac{1}{1+\exp(W_i)}\,.$$

Hence

$$\frac{P_{\boldsymbol{\sigma}\boldsymbol{\sigma'}}}{P_{\boldsymbol{\sigma'}\boldsymbol{\sigma}}} = \exp(W_i) = \frac{\pi(\boldsymbol{\sigma})}{\pi(\boldsymbol{\sigma'})}.$$

The above is relation is called *detailed balance equation* and such Markov chains are know as reversible as noted earlier. In such scenario, it is well known that π is the unique stationary distribution. □

4.2. *Mixing time.* The Glauber dynamics as described above converges to its stationary distribution π starting from any initial condition. To establish our results, we will need quantifiable bounds on the time it takes for the Glauber dynamics to reach close to stationary distribution. Specifically, we wish to quantify a bound in terms of the number of nodes n and the weight vector $\boldsymbol{W}$. To this end, we start with definition of distance between probability distributions.

Definition 3 *(Distance of measures) Given two probability distributions μ and ν on a finite space Ω, we define the following two distances. The total variation distance, denoted as $\|\mu-\nu\|_{TV}$ is*

$$\|\mu-\nu\|_{TV} = \frac{1}{2}\sum_{i\in\Omega}\left|\mu(i)-\nu(i)\right|.$$

The χ^2 distance, denoted as $\left\|\frac{\nu}{\mu}-1\right\|_{2,\mu}$ is

$$\left\|\frac{\nu}{\mu}-1\right\|_{2,\mu}^2 = \|\nu-\mu\|_{2,\frac{1}{\mu}}^2 = \sum_{i\in\Omega}\mu(i)\left(\frac{\nu(i)}{\mu(i)}-1\right)^2.$$

More generally, for any two vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}_+^{|\Omega|}$, we define

$$\|\mathbf{v}\|_{2,\mathbf{u}}^2 = \sum_{i\in\Omega} u_i v_i^2.$$

We make note of the following relation between the above defined two distances: for any probability distributions μ,ν, using the Cauchy-Schwartz inequality we have

$$\begin{aligned}
\left\|\frac{\nu}{\mu}-1\right\|_{2,\mu} &= \sqrt{\sum_{i\in\Omega}\mu(i)\left(\frac{\nu(i)}{\mu(i)}-1\right)^2} \\
&= \sqrt{\sum_{i\in\Omega}\mu(i)}\sqrt{\sum_{i\in\Omega}\mu(i)\left(\frac{\nu(i)}{\mu(i)}-1\right)^2} \\
&\geq \sum_{i\in\Omega}\mu(i)\left|\frac{\nu(i)}{\mu(i)}-1\right| \;=\; \sum_{i\in\Omega}\left|\nu(i)-\mu(i)\right| \\
&= 2\left\|\nu-\mu\right\|_{TV}. \qquad (3)
\end{aligned}$$

Next, we define a matrix norm that will be useful in determining rate of convergence or mixing time of a finite state Markov chain.

Definition 4 (Matrix norm) *Consider an $|\Omega| \times |\Omega|$ non-negative valued matrix $A \in \mathbb{R}_+^{|\Omega|\times|\Omega|}$ and given vector $\mathbf{u} \in \mathbb{R}_+^{|\Omega|}$. Then, the matrix norm of A with respect to $\mathbf{u}$ is defined as follows:*

$$\|A\|_{\mathbf{u}} = \sup_{\mathbf{v}:\mathbb{E}_{\mathbf{u}}[\mathbf{v}]=0} \frac{\|A\mathbf{v}\|_{2,\mathbf{u}}}{\|\mathbf{v}\|_{2,\mathbf{u}}},$$

where $\mathbb{E}_{\mathbf{u}}[\mathbf{v}] = \sum_i u_i v_i$.

It can be easily checked that the above definition of matrix norm satisfies the following properties.

P1. For matrices $A, B \in \mathbb{R}_+^{|\Omega|\times|\Omega|}$ and $\pi \in \mathbb{R}_+^{|\Omega|}$

$$\|A + B\|_\pi \le \|A\|_\pi + \|B\|_\pi.$$

P2. For matrix $A \in \mathbb{R}_+^{|\Omega|\times|\Omega|}$, $\pi \in \mathbb{R}_+^{|\Omega|}$ and $c \in \mathbb{R}$,

$$\|cA\|_\pi = |c|\|A\|_\pi.$$

P3. Let A and B be transition matrices of reversible Markov chains, i.e. $A = A^*$ and $B = B^*$. Let both of them have π as their unique stationary distribution. Then,

$$\|AB\|_\pi \le \|A\|_\pi \|B\|_\pi.$$

For a probability matrix P, mostly in this paper we will be interested in the matrix norm of P with respect to its stationary distribution π, i.e. $\|P\|_\pi$. Therefore, unless stated otherwise if we use matrix norm for a probability matrix without mentioning the reference measure, then it is with respect to the stationary distribution. That is, in above example $\|P\|$ will mean $\|P\|_\pi$.

With these definitions and fact that P and P^* have the same stationary distribution, say π, it follows that for any distribution μ on Ω

(4)
$$\left\| \frac{\mu P}{\pi} - 1 \right\|_{2,\pi} \le \|P^*\| \left\| \frac{\mu}{\pi} - 1 \right\|_{2,\pi},$$

since $\mathbb{E}_\pi\left[\frac{\mu}{\pi} - 1\right] = 0$, with interpretation $\frac{\mu}{\pi} = [\mu(i)/\pi(i)]$. The Markov chain of our interest, Glauber dynamics, is reversible, and for a reversible Markov chain, $P = P^*$. Therefore, for a reversible Markov chain starting with initial distribution $\mu(0)$, the distribution $\mu(\tau)$ at time τ is such that

(5)
$$\left\| \frac{\mu(\tau)}{\pi} - 1 \right\|_{2,\pi} \le \|P\|^\tau \left\| \frac{\mu(0)}{\pi} - 1 \right\|_{2,\pi}.$$

Now starting from any state i, i.e. probability distribution with unit mass on state i, the initial distance $\left\| \frac{\mu(0)}{\pi} - 1 \right\|_{2,\pi}$ in the worst case is bounded above by $\sqrt{1/\pi_{\min}}$ where $\pi_{\min} = \min_i \pi_i$. Therefore, for any $\delta > 0$ we have $\left\| \frac{\mu(\tau)}{\pi} - 1 \right\|_{2,\pi} \le \delta$ for any τ such that

$$\tau \ge \frac{\log 1/\pi_{\min} + \log 1/\delta}{\log \|P\|} \;=\; \Theta\left(\frac{\log 1/\pi_{\min} + \log 1/\delta}{1 - \|P\|}\right).$$

This suggests that the "mixing time", i.e. time to reach (close to) stationary distribution of the Markov chain scales inversely with $1 - \|P\|$. Therefore, we will define the "mixing time" of a Markov chain with transition matrix P as $1/(1 - \|P\|)$. This also suggests that in order to bound the distance between a Markov chain's distribution after some steps and its stationary distribution, it is sufficient to obtain a bound on $\|P\|$. One such bound is stated below.

Lemma 5 *Let P be the transition matrix of the Glauber dynamics $GD(\boldsymbol{W})$ on graph $G = (V, E)$ of $n = |V|$ nodes. Then,*

$$\|P\| \leq 1 - \frac{1}{8n^2 \exp(2W_{\max})},$$

where $W_{\max} = \max_{i \in V} W_i$.

Proof. Since P is reversible and probability matrix, it is well known that it has real eigenvalues with values between $[-1, 1]$ with the largest eigenvalue equal to 1. Let them be denoted as $1 = \lambda_0 \geq \lambda_1 \cdots \geq \lambda_{N-1}$ with the corresponding left eigenvectors $\mathbf{u}_0 = \pi, \mathbf{u}_1, \ldots, \mathbf{u}_{N-1}$ and the corresponding right eigenvectors $\mathbf{v}_0 = \mathbf{1}, \ldots \mathbf{v}_{N-1}$. Here $N = |\mathcal{I}(G)|$, the size of the state space over which $GD(\boldsymbol{W})$ evolves. By the spectral theorem for reversible matrices, we can assume that the vectors v_i are orthonormal with respect to π, i.e.

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle_\pi = \sum_{k=1}^{N} v_{i_k} v_{j_k} \pi_k = \delta_{ij},$$

with the definition $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Therefore, any vector $\mathbf{x} \in \mathbb{R}^N$ can be written as

$$\mathbf{x} = \sum_{i=0}^{N-1} \alpha_i \mathbf{v}_i,$$

where $\alpha_i = \langle \mathbf{x}, \mathbf{v}_i \rangle_\pi$. By definition, $\alpha_0 = 0$ when $\mathbb{E}_\pi[\mathbf{v}] = 0$. Therefore,

$$\|P\| = \sup_{E_\pi[\mathbf{v}]=0} \frac{\|P\mathbf{v}\|_{2,\pi}}{\|\mathbf{v}\|_{2,\pi}} = \frac{\sqrt{\sum \alpha_i^2 \lambda_i^2}}{\sqrt{\sum \alpha_i^2}} \leq \lambda_{\max},$$

where $\lambda_{\max} = \max\{\lambda_1, |\lambda_{n-1}|\}$. By Cheeger's inequality, it is well known that $\lambda_{\max} \leq 1 - \frac{\Phi^2}{2}$ where Φ is the conductance of P, defined as

$$\Phi = \min_{S \subset \mathcal{I}(G) : \pi(S) \leq \frac{1}{2}} \frac{P(S, S^c)}{\pi(S)},$$

where $S^c = \mathcal{I}(G) \backslash S$, $P(S, S^c) = \sum_{\boldsymbol{\sigma} \in S, \boldsymbol{\sigma}' \in S'} P(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$. Now we have

$$\begin{aligned}
\Phi &\geq \min_{S \subset V} P(S, S') \geq \min_{P(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \neq 0} P(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \\
&\geq \min_i \frac{1}{n} \frac{1}{1 + \exp(W_i)} = \frac{1}{n} \frac{1}{1 + \exp(W_{\max})} \\
&\geq \frac{1}{2n \exp(W_{\max})}
\end{aligned}$$

Therefore $\|P\| \leq \lambda_{\max} \leq 1 - \frac{\Phi^2}{2} \leq 1 - \frac{1}{8n^2 \exp(2W_{\max})}$. □

5. Proof of Main Result: Theorem 1. This section presents the detailed proof of Theorem 1. We will present the sketch of the proof followed by details.

5.1. *Proof sketch.* We first introduce the necessary definition of the network Markov chain under our algorithm. As before, τ be the index for discrete time. Let $\mathbf{Q}(\tau) = [Q_i(\tau)]$ denote the vector of queue sizes at time τ; $\widetilde{\mathbf{Q}}(\tau) = [\widetilde{Q}_{\max,i}(\tau)]$ be the vector of estimates of $Q_{\max}(\tau)$ at time τ; $\boldsymbol{\sigma}(\tau) = [\sigma_i(\tau)]$ be the scheduling choices at the n nodes at time τ and $\boldsymbol{S}(\tau) = [S_i(\tau)]$ be the remaining service required for the head-of-line packets in queue i at time τ. Then it can be checked that the tuple $X(\tau) = (\mathbf{Q}(\tau), \widetilde{\mathbf{Q}}(\tau), \boldsymbol{\sigma}(\tau), \boldsymbol{S}(\tau))$ is the Markov state of the network operating under the algorithm. Note that $X(\tau) \in \mathsf{X}$ where $\mathsf{X} = \mathbb{R}_+^n \times \mathbb{R}_+^n \times \mathcal{I}(G) \times [0,1]^n$. Clearly, X is a Polish space endowed with the natural product topology. Let $\mathcal{B}_{\mathsf{X}}$ be the Borel σ-algebra of X with respect to this product topology. Let P denote the probability transition matrix of this discrete time X-valued Markov chain. We wish to establish that $X(\tau)$ is indeed positive Harris recurrent under this setup. For any $\mathbf{x} = (\mathbf{Q}, \widetilde{\mathbf{Q}}, \boldsymbol{\sigma}, \boldsymbol{S}) \in \mathsf{X}$, we define norm of $\mathbf{x}$ denoted by $|\mathbf{x}|$ as

$$|\mathbf{x}| = |\mathbf{Q}| + |\widetilde{\mathbf{Q}}| + |\boldsymbol{\sigma}| + |\boldsymbol{S}|,$$

where $|\mathbf{Q}|, |\widetilde{\mathbf{Q}}|$ and $|\boldsymbol{S}|$ denote the standard ℓ_1 norm while $|\boldsymbol{\sigma}|$ is defined as its index in $\{0, \ldots, |\mathcal{I}(G)| - 1\}$, which is assigned arbitrarily. Thus, $|\boldsymbol{S}|, |\boldsymbol{\sigma}|$ are always bounded. Further, by Lemma 2 we have $|\widetilde{\mathbf{Q}}| = \Theta(|\mathbf{Q}|)$ under the evolution of Markov chain. Therefore, in essence if $|\mathbf{x}| \to \infty$ then $|\mathbf{Q}| \to \infty$. Next, we present the proof based on a sequence of lemmas. The proofs will be presented subsequently.

We will need some definitions to begin with. Given a probability distribution (also called sampling distribution) a on $\mathbb{N}$, the a-sampled transition matrix of the Markov chain, denoted by K_a is defined as

$$K_a(\mathbf{x}, B) = \sum_{\tau \geq 0} a(\tau) P^\tau(\mathbf{x}, B), \quad \text{for any} \quad \mathbf{x} \in \mathsf{X}, \ B \in \mathcal{B}_{\mathsf{X}}.$$

Now, we define a notion of a *petite* set. A non-empty set $A \in \mathcal{B}_{\mathsf{X}}$ is called μ_a-*petite* if μ_a is a non-trivial measure on $(\mathsf{X}, \mathcal{B}_{\mathsf{X}})$ and a is a probability distribution on $\mathbb{N}$ such that for any $\mathbf{x} \in A$,

$$K_a(\mathbf{x}, \cdot) \geq \mu_a(\cdot).$$

A set is called *petite* set if it is μ_a-petite for some such non-trivial measure μ_a. A known sufficient condition to establish positive Harris recurrence of a Markov chain is to establish positive Harris recurrence of closed petite sets as stated in the following lemma. We refer an interested reader to the book by Meyn and Tweedie [24] or recent survey by Foss and Konstantopoulos [8] for details.

Lemma 6 *Let B be a closed petite set. Suppose B is Harris recurrent, i.e. $\Pr_{\mathbf{x}}(T_B < \infty) = 1$ for any $\mathbf{x} \in \mathsf{X}$. Further, let*

$$\sup_{\mathbf{x} \in B} \mathbb{E}_{\mathbf{x}}[T_B] < \infty.$$

Then the Markov chain is positive Harris recurrent.

Lemma 6 suggests that to establish the positive Harris recurrence of the network Markov chain, it is sufficient to find a closed petite set that satisfies the conditions of Lemma 6. To this end, we first establish that there exist closed sets that satisfy condition of Lemma 6. Later we will establish that they are indeed petite sets. This will conclude the proof of positive Harris recurrence of the network Markov chain.

Recall that the 'weight' function is $f(x) = \log\log(x+e)$. Define its integral, $F(x) = \int_0^x f(y)dy$. The system Lyapunov function, $L : \mathsf{X} \to \mathbb{R}_+$ is defined as

$$L(\mathbf{x}) = \sum_{i=1}^{n} F(q_i) \stackrel{\triangle}{=} F(\mathbf{q}) \cdot \mathbf{1}, \quad \text{where } \mathbf{x} = (\mathbf{Q}, \widetilde{\mathbf{Q}}, \boldsymbol{\sigma}, \boldsymbol{S}) \in \mathsf{X}.$$

We will establish the following whose proof of given in section 5.3.

Lemma 7 *Let $\lambda \in (1-2\varepsilon)\Lambda^o$. Then there exist functions $h, g : \mathsf{X} \to \mathbb{R}$ such that for any $\mathbf{x} \in \mathsf{X}$,*

$$\mathbb{E}\left[L(X(g(\mathbf{x})) - L(X(0))|X(0) = \mathbf{x}\right] \leq -h(\mathbf{x}),$$

and (a) $\inf_{\mathbf{x}\in\mathsf{X}} h(\mathbf{x}) > -\infty$, (b) $\liminf_{L(\mathbf{x})\to\infty} h(\mathbf{x}) > 0$, (c) $\sup_{L(\mathbf{x})\leq\gamma} g(\mathbf{x}) < \infty$ for all $\gamma > 0$ and (d) $\limsup_{L(\mathbf{x})\to\infty} g(\mathbf{x})/h(\mathbf{x}) < \infty$.

Now define $B_\kappa = \{\mathbf{x} : L(\mathbf{x}) \leq \kappa\}$ and $C_\kappa = \{\mathbf{x} : |\mathbf{x}| \leq \kappa\}$ for any $\kappa > 0$. It can be easily checked that for any $\kappa > 0$, there exists κ' such that for any $\mathbf{x} \in \mathsf{X}$,

$$L(\mathbf{x}) \leq \kappa \;\Rightarrow\; |\mathbf{x}| \leq \kappa'.$$

In above, one needs to use the fact that $\mathbf{x}$ always satisfies condition of Lemma 2. Similarly, for any $\kappa' > 0$ there exists κ'' such that

$$|\mathbf{x}| \leq \kappa' \;\Rightarrow\; L(\mathbf{x}) \leq \kappa''.$$

In summary, it follows that for any $\kappa > 0$, there exist $0 < \kappa_1(\kappa) \leq \kappa_2(\kappa)$ such that

$$B_{\kappa_1(\kappa)} \subset C_\kappa \subset B_{\kappa_2(\kappa)}.$$

Using this relation, arguments of Theorem 1 in the survey [8] and Lemma 7, it immediately follows that there exists a constant $\kappa_0 > 0$ such that for all $\kappa_0 < \kappa$, the following holds:

$$\mathbb{E}_{\mathbf{x}}\left[T_{C_\kappa}\right] < \infty, \qquad \text{for any } \mathbf{x} \in \mathsf{X} \tag{6}$$

$$\sup_{\mathbf{x}\in C_\kappa} \mathbb{E}_{\mathbf{x}}\left[T_{C_\kappa}\right] < \infty. \tag{7}$$

Now we are ready to state the final nugget required in proving positive Harris recurrence as stated below.

Lemma 8 *Consider any $\kappa > 0$. Then, the set $C_\kappa = \{\mathbf{x} \in \mathsf{X} : |\mathbf{x}| \leq \kappa\}$ is a closed petite set.*

The proof of Lemma 8 is presented in section 5.8. Thus, Lemmas 6, 7 and 8 imply that the network Markov chain is positive Harris recurrent. Finally, we state a corollary of Lemma 7.

Corollary 9 *Under the algorithm,*

$$\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}_{\mathbf{x}}\left[\sum_{\tau\le T} f(\mathbf{Q}(\tau))\cdot\mathbf{1}\right] = O(1), \quad \Pr_{\mathbf{x}}\text{-}\textit{almost surely.}$$

The proof of Corollary 9 is presented in section 5.4. The Corollary 9 establishes the uniform integrability of $(\sum_{\tau\le T} f(\mathbf{Q}(\tau))\cdot\mathbf{1})/T$. Therefore, by ergodicity (implied by the positive Harris recurrence of the network Markov chain as stated in section 2.4.1), it follows that

$$\mathbb{E}\left[f(\mathbf{Q})\cdot\mathbf{1}\right] = O(1).$$

This completes the proof of Theorem 1.

5.2. *Some preliminaries.* We define condition on constant $B > 0$ required for the efficiency of the algorithm for given $\varepsilon > 0$.

Definition 5 *Given $\varepsilon > 0$ (as in the statement of Theorem 1), B is a constant that is large enough such that it satisfies the following:*

$$f^{-1}(B) \ge \max\{2n, 7^7\} \quad \textit{and} \qquad \frac{256n^2\left(\log\left(x+e\right)\right)^2}{e^{(\log(x-2n+e))^{\frac{\varepsilon}{n}}} - e - 1} < \varepsilon, \quad \forall x \ge f^{-1}(B).$$

Indeed, there exists such a $B = B(n,\varepsilon)$ since $\lim_{x\to\infty} \frac{256n^2(\log(x+e))^2}{e^{(\log(x-2n+e))^{\frac{\varepsilon}{n}}}-e-1} = 0$ for any fixed $\varepsilon > 0$.

Now we relate our algorithm described in section 3.1 with appropriate continuous time version of Glauber dynamics described in section 4.1. To this end, recall that the algorithm changes its scheduling decision as per Exponential rate 1 clock ticks of nodes. Due to the property of Exponential distribution, no two nodes have clocks ticking at the same time. Now given a clock tick, it is equally likely to be any of the n nodes. The node whose clock ticks decides its transition based on probability prescribed by the Glauber dynamics $GD(\boldsymbol{W}(t))$ where recall that $\boldsymbol{W}(t)$ are determined based on $\mathbf{Q}(\lfloor t\rfloor), \tilde{\mathbf{Q}}(\lfloor t\rfloor)$. Thus the transition probabilities of Markov process determining the schedule $\boldsymbol{\sigma}(t)$ change every discrete time. Let $P(t)$ denote the transition matrix prescribed by the Glauber dynamics $GD(\boldsymbol{W}(t))$ and $\pi(t)$ denote its stationary distribution. Now the scheduling algorithm evolves the scheduling decision $\boldsymbol{\sigma}(\cdot)$ over time with time varying $P(t)$ as described before. Let $\mu(t)$ be the distribution of the schedule $\boldsymbol{\sigma}(t)$ at time t. The algorithm is essentially running $P(\lfloor t\rfloor)$ on $\mathcal{I}(G)$ when a clock ticks at time t. Since there are n clocks and each clock rate is 1, we have

$$\begin{aligned}\mu(t) &= \sum_{i=0}^{\infty}\Pr(\zeta = i)\mu(\lfloor t\rfloor)P(\lfloor t\rfloor)^i\\ &= \frac{1}{e^{n(t-\lfloor t\rfloor)}}\mu(\lfloor t\rfloor)e^{n(t-\lfloor t\rfloor)P(\lfloor t\rfloor)}\\ &= \mu(\lfloor t\rfloor)e^{n(t-\lfloor t\rfloor)(P(\lfloor t\rfloor)-I)},\end{aligned}$$

where ζ is the number of clock ticks in time $(\lfloor t\rfloor, t]$ and it is distributed as a Poisson random variable with mean $n(t-\lfloor t\rfloor)$. Thus, for any $\tau\in\mathbb{N}$,

$$\mu(\tau+1) \;=\; \mu(\tau)e^{n(P(\tau)-I)}. \tag{8}$$

In the above (and in the remaining of the paper) in the left multiplication of a vector with a matrix, the vector should be thought of as a row vector. The equation (8) gives the discrete time interpretation on μ, hence the mixing time based analysis on μ with the transition matrix $e^{n(P(\tau)-I)}$ becomes possible. The transition matrix $e^{n(P(\tau)-I)}$ has properties similar to that of $P(\tau)$ as stated below.

Lemma 10 *$e^{n(P(\tau)-I)}$ is reversible and its stationary distribution is $\pi(\tau)$. Furthermore, its matrix norm is bounded as*

$$\left\|e^{n(P(\tau)-I)}\right\| \leq 1 - \frac{1}{16n\exp(2W_{\max}(\tau))}.$$

Proof. We have that $\pi(\tau)$ is the stationary distribution of $P(\tau)$ and $P(\tau)$ is reversible. Given this, the reversibility of $e^{n(P(\tau)-I)}$ as well as $\pi(\tau)$ being its stationary distribution follows directly from the definition. Now, consider the following (using properties **P1**, **P2** and **P3** of matrix norm):

$$\begin{aligned}
\left\|e^{n(P-I)}\right\| &= \left\|e^{-n}\sum_{k=0}^{\infty}\frac{n^k P^k}{k!}\right\| \\
&\leq e^{-n}\sum_{k=0}^{\infty}\frac{n^k \|P\|^k}{k!} \\
&= e^{n(\|P\|-1)} \\
&\leq 1 - \frac{n(1-\|P\|)}{2}. \qquad (9)
\end{aligned}$$

In the last inequality, we have used the fact that $\|P\| < 1$ and $e^{-x} \leq 1 - x/2$ for all $x \in [0,1]$. Use of $P = P(\tau)$ in (9) and Lemma 5, we obtain

$$\left\|e^{n(P-I)}\right\| \leq 1 - \frac{1}{16n\exp(2W_{\max}(\tau))}.$$

This completes the proof of Lemma 10. □

5.3. *Proof of Lemma 7.* We have $\lambda \in (1-2\varepsilon)\Lambda^o$. That is, for some $\delta > 0$, $\boldsymbol{\lambda} \leq (1-2\varepsilon-\delta)\Lambda$. The proof of Lemma 7 crucially utilizes the following Lemma 11. We will prove Lemma 11 in section 5.5.

Lemma 11 *For given $\delta, \varepsilon > 0$, let $\boldsymbol{\lambda} \leq (1-2\varepsilon-\delta)\Lambda$. Given any starting condition $X(0) = \mathbf{x} = (\mathbf{Q}(0), \widetilde{\mathbf{Q}}(0), \boldsymbol{\sigma}(0), \boldsymbol{S}(0))$, there exists a constant $C \stackrel{\triangle}{=} C(\mathbf{Q}(0))$ such that for $T > C$,*

$$\mathbb{E}_{\mathbf{x}}[L(X(T)) - L(X(C))] \leq -\frac{\delta}{n}\sum_{\tau=C}^{T-1}\mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(\tau))\cdot\mathbf{1}] + (Bn+3n+\delta)(T-C),$$

with $C(\boldsymbol{Q}(0)) = O\left(\log^9 Q_{\max}(0)\right)$. Here, as usual $\mathbb{E}_{\mathbf{x}}[\cdot]$ denotes expectation with respect to the condition that $X(0) = \mathbf{x}$.

Now, we proceed towards the proof of Lemma 7. From Lemma 11, for $T > C = C(\mathbf{Q}(0))$,

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}}[L(X(T)) - L(X(0))] &\leq -\frac{\delta}{n}\sum_{\tau=C}^{T-1}\mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(\tau))\cdot\mathbf{1}] + (Bn+3n+\delta)(T-C) \\
&\quad + L(X(C)) - L(X(0)) \\
&\leq -\frac{\delta}{n}\sum_{\tau=C}^{T-1}\mathbb{E}_{\mathbf{x}}[f(Q_{\max}(\tau))] + (Bn+3n+\delta)(T-C) \\
&\quad + L(X(C)) \\
&\leq -\frac{\delta}{n}(T-C)f\left((Q_{\max}(0)-T)^+\right) + (Bn+3n+\delta)(T-C) \\
&\quad + L(X(C)) \\
&= -\frac{\delta}{n}(T-C)f\left((Q_{\max}(0)-T)^+\right) + (Bn+3n+\delta)(T-C) \\
&\quad + F(\boldsymbol{Q}(C))\cdot\mathbf{1} \\
&\leq -\frac{\delta}{n}(T-C)f\left((Q_{\max}(0)-T)^+\right) + (Bn+3n+\delta)(T-C) \\
&\quad + C\,n\,f(Q_{\max}(0)+C). \qquad (10)
\end{aligned}$$

Let $K = K(\delta,n) \stackrel{\triangle}{=} \lceil 2 + n^2/\delta\rceil$. Now for given $\mathbf{x} = (\mathbf{Q}, \widetilde{\mathbf{Q}}, \boldsymbol{\sigma}, \boldsymbol{S}) \in \mathsf{X}$, we define functions g and h as desired as follows:

$$\begin{aligned}
g(\mathbf{x}) &= K(\delta,n)C(\mathbf{Q}), \quad \text{and} \qquad (11) \\
h(\mathbf{x}) &= \frac{\delta}{n}\left(g(\mathbf{x}) - C(\mathbf{Q})\right) f\left((Q_{\max} - g(\mathbf{x}))^+\right) \\
&\qquad -((B+3)n+\delta)(g(\mathbf{x}) - C(\mathbf{Q})) - C(\mathbf{Q})nf(Q_{\max}+C(\mathbf{Q})) \\
&= \frac{\delta}{n}(K(\delta,n)-1)C(\mathbf{Q})f\left((Q_{\max} - K(\delta,n)C(\mathbf{Q}))^+\right) \\
&\qquad -((B+3)n+\delta)(K(\delta,n)-1)C(\mathbf{Q}) - C(\mathbf{Q})nf(Q_{\max}+C(\mathbf{Q})). \qquad (12)
\end{aligned}$$

It can be verified that by setting $T = g(\mathbf{x})$ in (10), we have that

$$\mathbb{E}[L(X(g(\mathbf{x}))) - L(X(0))|X(0) = \mathbf{x}] \leq -h(\mathbf{x}).$$

Now to complete the proof of Lemma 7 we need to verify that functions g, h satisfy conditions (a), (b), (c) and (d). For this, note that if $L(\mathbf{x}) \to \infty$ then it must be that $\|\mathbf{Q}\| \to \infty$. Using this and definitions of g, h as per (11) and (12) the conditions (a), (b) and (c) follow immediately. For condition (d), consider the following with short-hand notation of $K = K(\delta,n)$ and $C = C(\mathbf{Q})$

$$\begin{aligned}
\frac{g(\mathbf{x})}{h(\mathbf{x})} &= \frac{KC}{\frac{\delta}{n}(K-1)Cf\left((Q_{\max}-KC)^+\right) - ((B+3)n+\delta)(K-1)C - Cnf(Q_{\max}+C)} \\
&\leq \frac{1}{\frac{\delta}{n}(1-\frac{1}{K})f\left((Q_{\max}-KC)^+\right) - ((B+3)n+\delta)(1-\frac{1}{K}) - \frac{n}{K}f(Q_{\max}+C)}, \qquad (13) \\
&\stackrel{L(\mathbf{x})\to\infty}{\longrightarrow} 0,
\end{aligned}$$

since $\frac{\delta}{n}(1-\frac{1}{K}) > \frac{n}{K}$; $C = C(\mathbf{Q}) = O\left(\log^9 Q_{\max}\right)$ and

$$L(\mathbf{x}) \to \infty \quad \Rightarrow \quad (\|\mathbf{Q}\| \to \infty \Leftrightarrow Q_{\max} \to \infty),$$

imply that the denominator of (13) goes to ∞. This completes the verification of condition (d) and the proof of Lemma 7.

5.4. *Proof of Corollary 9.* From Lemma 11 it follows that starting with any initial state $X(0) = \mathbf{x}$, we have for T large enough

$$\mathbb{E}_{\mathbf{x}}[L(X(T)) - L(X(C))] \leq -\frac{\delta}{n}\sum_{\tau=C}^{T-1}\mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(\tau))\cdot\mathbf{1}] + (Bn+3n+\delta)(T-C), \tag{14}$$

with $C = C(\mathbf{Q}(0)) = O\left(\log^9 Q_{\max}(0)\right)$. Given the bound on C and the fact that the queue size can grow by at most a constant amount per unit time, it follows that $L(X(C))$ can be bounded as a function of $L(\mathbf{x})$. Therefore, starting with $X(0) = \mathbf{x}$

$$\lim_{T\to\infty}\frac{1}{T}L(X(C)) = 0.$$

Further L is non-negative valued. Therefore, from (14) we obtain that

$$\begin{aligned}\frac{1}{T}\sum_{\tau=C}^{T-1}\mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(\tau))\cdot\mathbf{1}] &\leq \frac{n}{\delta}\left[(B+3)n+\delta+\frac{1}{T}\mathbb{E}_{\mathbf{x}}[L(X(C))]\right]\\ &\overset{T\to\infty}{\longrightarrow} \frac{n}{\delta}\left[(B+3)n+\delta\right] = O(1).\end{aligned} \tag{15}$$

This completes the proof of Corollary 9.

5.5. *Proof of Lemma 11.* Here we prove Lemma 11 using the following two lemmas, which will be proven in later sections.

Lemma 12 *Consider a vector of queues $\mathbf{Q} \in \mathbb{R}_+^n$. Let the vector of estimation of $Q_{\max}$ be $\widetilde{\mathbf{Q}} \in \mathbb{R}_+^n$ satisfying the property of Lemma 2. Let weight vector $\boldsymbol{W}$ based on these queues be defined as per equation* (1)*. Consider the Glauber dynamics $GD(\boldsymbol{W})$ and let π denote its stationary distribution. If $\boldsymbol{\sigma}$ is distributed as per π then*

$$E_\pi[f(\boldsymbol{Q})\cdot\boldsymbol{\sigma}] \geq (1-\varepsilon)\left(\max_{\boldsymbol{\rho}\in\mathcal{I}(G)} f(\mathbf{Q})\cdot\boldsymbol{\rho}\right) - (B+2)n.$$

Before we state the next lemma, we define a transformation of queue-size vector: given vector of queue size $\mathbf{Q} \in \mathbb{R}_+^n$ and vector of estimation of $Q_{\max}$ denoted as $\widetilde{\mathbf{Q}}$, recall that the corresponding weight $\boldsymbol{W}$ was defined in (1) as

$$W_i(t) = \max\left\{f(Q_i(\lfloor t\rfloor)), \frac{\varepsilon}{n}f(\widetilde{Q}_{\max,i}(\lfloor t\rfloor)), B\right\}.$$

Define $\widehat{Q}_i = f^{-1}(W_i)$ and let $\widehat{\mathbf{Q}}$ denote thus transformed vector of queue sizes. We state the following important properties of $\widehat{Q}$ which will be used in the later analysis:

$$\widehat{Q}_i = f^{-1}(W_i) \geq f^{-1}(B) \geq 2n, \tag{16}$$

$$\widehat{Q}_{\max} = f^{-1}(W_{\max}) = \max\{Q_{\max}, f^{-1}(B)\}, \tag{17}$$

$$\widehat{Q}_{\min} \geq f^{-1}\left(\frac{\varepsilon}{n} f(\widehat{Q}_{\max} - 2n)\right). \tag{18}$$

Properties (16) and (17) follow from definition 5 in a straightforward manner. For (18), consider two cases. First, when $\widehat{Q}_{\max} = Q_{\max}$, consider the following:

$$\begin{aligned}
\widehat{Q}_{\min} &= f^{-1}(W_{\min}) \geq f^{-1}\left(\frac{\varepsilon}{n} f(\widetilde{Q}_{\max,i})\right) \\
&\geq f^{-1}\left(\frac{\varepsilon}{n} f(Q_{\max} - 2n)\right) \quad \text{(from Lemma 2)} \\
&= f^{-1}\left(\frac{\varepsilon}{n} f(\widehat{Q}_{\max} - 2n)\right).
\end{aligned}$$

In the second case, when $\widehat{Q}_{\max} = f^{-1}(B)$, by definition it must be that $\widehat{Q}_i = f^{-1}(B)$ for all i, i.e. $\widehat{Q}_{\max} = \widehat{Q}_{\min}$ and hence (18) follows trivially.

Lemma 13 (Network adiabatic Theorem) *For a given $\boldsymbol{Q}(0)$, as defined earlier let $\mu(t)$ be the distribution of the schedule over $\mathcal{I}(G)$ at time t and let $\pi(t)$ be the stationary distribution of Markov process over $\mathcal{I}(G)$ with respect to the probability transition matrix $P(t)$ as defined in section 5.2. Then, for $t > C_1(\widehat{\mathbf{Q}}(0))$*

$$\left\| \frac{\mu(t)}{\pi(t)} - 1 \right\|_{2,\pi(t)} < \varepsilon, \qquad \textit{with probability } 1,$$

where

$$C_1(\widehat{\mathbf{Q}}(0)) = \left\lceil 16^2 n^2 \log^4(\widehat{Q}_{\max}(0) + 1 + e) \log\left(\frac{2}{\varepsilon}\left(2\log(\widehat{Q}_{\max}(0) + e)\right)^{n/2}\right)\right\rceil^2 + 1.$$

Remark 1 *The statement of Lemma 13 suggests that for all (large enough) time, the distribution of schedules $\mu(t)$ is essentially close to the stationary distribution $\pi(t)$ at each time despite the fact that queue sizes (and hence weights) keep changing. In conjuction with Lemma 12 this adiabatic like result suggests that indeed the choice of schedule is such that at each time the weight of schedule is close to that of the maximum weight schedule as desired. This is the key property that establishes the efficiency (positive Harris recurrence) of the network Markov chain.*

Now we proceed towards proving Lemma 11. From Lemma 13 and relation (3), we have that for $t \geq C_1(\widehat{\mathbf{Q}}(0))$,

$$\begin{aligned}
\left| E_{\pi(t)}[f(\boldsymbol{Q}(t)) \cdot \boldsymbol{\sigma}] - E_{\mu(t)}[f(\boldsymbol{Q}(t)) \cdot \boldsymbol{\sigma}] \right| &\leq \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}(t)) \cdot \boldsymbol{\sigma} \right) \|\pi(t) - \mu(t)\|_{TV} \\
&\leq \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}(t)) \cdot \boldsymbol{\rho} \right) \left\| \frac{\mu(t)}{\pi(t)} - 1 \right\|_{2,\pi(t)} \\
&\leq \varepsilon \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}(t)) \cdot \boldsymbol{\rho} \right). \qquad (19)
\end{aligned}$$

Thus from Lemma 12,

$$E_{\mu(t)}[f(\boldsymbol{Q}(t))\cdot\boldsymbol{\sigma}] \geq E_{\pi(t)}[f(\boldsymbol{Q}(t))\cdot\boldsymbol{\sigma}] - \varepsilon\left(\max_{\boldsymbol{\rho}\in\mathcal{I}(G)} f(\boldsymbol{Q}(t))\cdot\boldsymbol{\rho}\right)$$
$$\geq (1-2\varepsilon)\left(\max_{\boldsymbol{\rho}\in\mathcal{I}(G)} f(\boldsymbol{Q}(t))\cdot\boldsymbol{\rho}\right) - n(B+2).$$

Now consider the difference between $L(X(\tau+1))$ and $L(X(\tau))$ as follows.

$$\begin{aligned}
L(X(\tau+1)) - L(X(\tau)) &= (F(\boldsymbol{Q}(\tau+1)) - F(\boldsymbol{Q}(\tau)))\cdot\mathbf{1}\\
&\leq f(\boldsymbol{Q}(\tau+1))\cdot(\boldsymbol{Q}(\tau+1)-\boldsymbol{Q}(\tau)), \qquad (\text{as } F \text{ is convex}),\\
&\leq f(\boldsymbol{Q}(\tau+1))\cdot\left(A(\tau,\tau+1) - \int_\tau^{\tau+1}\boldsymbol{\sigma}(r)\mathbf{1}_{\{Q_i(r)>0\}}\,dr\right)\\
&\leq \int_\tau^{\tau+1} f(\boldsymbol{Q}(\tau+1))\cdot\left(A(\tau,\tau+1) - \boldsymbol{\sigma}(r)\mathbf{1}_{\{Q_i(r)>0\}}\right)\,dr\\
&\leq \int_\tau^{\tau+1} f(\boldsymbol{Q}(r))\cdot\left(A(\tau,\tau+1) - \boldsymbol{\sigma}(r)\mathbf{1}_{\{Q_i(r)>0\}}\right)dr\\
&\quad + \int_\tau^{\tau+1}(f(\boldsymbol{Q}(\tau+1)) - f(\boldsymbol{Q}(r)))\cdot\left(A(\tau,\tau+1) - \boldsymbol{\sigma}(r)\mathbf{1}_{\{Q_i(r)>0\}}\right)\,dr\\
&\leq \int_\tau^{\tau+1} f(\boldsymbol{Q}(r))\cdot(A(\tau,\tau+1) - \boldsymbol{\sigma}(r)\mathbf{1}_{\{Q_i(r)>0\}})\,dr + n\\
&= \int_\tau^{\tau+1} f(\boldsymbol{Q}(r))\cdot(A(\tau,\tau+1) - \boldsymbol{\sigma}(r))\,dr + n, \qquad (\text{as } f(0)=0), \qquad (20)
\end{aligned}$$

where the second last inequality follows by the fact that f is 1-Lipschitz[4] and $\mathbf{Q}(\cdot)$ changes at unit rate. Taking expectation of (20) given initial state $X(0)=\mathbf{x}$ and $\tau \geq C_1(\widehat{\mathbf{Q}}(0))$ we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}}[L(X(\tau+1)) - L(X(\tau))] &\leq \int_\tau^{\tau+1}\left(\mathbb{E}_{\mathbf{x}}\left[f(\boldsymbol{Q}(r))\cdot A(\tau+1,\tau)\right] - \mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(r))\cdot\boldsymbol{\sigma}(r)]\right)\,dr + n\\
&\leq \int_\tau^{\tau+1}\left(\mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(r))]\cdot\lambda - (1-2\varepsilon)\mathbb{E}_{\mathbf{x}}\left[\max_{\boldsymbol{\rho}\in\mathcal{I}(G)} f(\boldsymbol{Q}(r))\cdot\boldsymbol{\rho}\right]\right)dr\\
&\quad +(B+3)n\\
&\overset{(a)}{\leq} -\int_\tau^{\tau+1}\delta\mathbb{E}_{\mathbf{x}}\left[\max_{\boldsymbol{\rho}\in\mathcal{I}(G)} f(\boldsymbol{Q}(r))\cdot\boldsymbol{\rho}\right]dr\\
&\quad +(B+3)n\\
&\overset{(b)}{\leq} -\int_\tau^{\tau+1}\frac{\delta}{n}\mathbb{E}_{\mathbf{x}}\left[f(\boldsymbol{Q}(r))\cdot\mathbf{1}\right]\,dr\\
&\quad +(B+3)n\\
&\leq -\frac{\delta}{n}\left(\mathbb{E}_{\mathbf{x}}\left[f(\boldsymbol{Q}(\tau))\cdot\mathbf{1}\right] - \int_\tau^{\tau+1}\mathbb{E}_{\mathbf{x}}\left[(f(\boldsymbol{Q}(r)) - f(\boldsymbol{Q}(\tau)))\cdot\mathbf{1}\right]\,dr\right)\\
&\quad +(B+3)n\\
&\leq -\frac{\delta}{n}\mathbb{E}_{\mathbf{x}}\left[f(\boldsymbol{Q}(\tau))\cdot\mathbf{1}\right] + (B+3)n + \delta, \quad (\text{as } f(\boldsymbol{Q}) \text{ is Lipshitz}). \qquad (21)
\end{aligned}$$

[4]A continuous function $f:\mathbb{R}\to\mathbb{R}$ is K-Lipschitz if $|f(x)-f(y)|\leq K|x-y|$ for all $x,y\in\mathbb{R}$.

In above, we justify (a) and (b) as follows. For (a), recall that $\boldsymbol{\lambda} \in (1-2\varepsilon-\delta)\boldsymbol{\Lambda}$ and hence $\boldsymbol{\lambda} \leq \sum_{\boldsymbol{\rho}} \alpha_{\boldsymbol{\rho}} \boldsymbol{\rho}$ with $\sum_{\boldsymbol{\rho}} \alpha_{\boldsymbol{\rho}} \leq 1-2\varepsilon-\delta$ and $\alpha_{\boldsymbol{\rho}} \geq 0$. Therefore,

$$\begin{aligned} f(\boldsymbol{Q}(r))] \cdot \lambda &\leq \sum_{\boldsymbol{\rho}} \alpha_{\boldsymbol{\rho}} f(\boldsymbol{Q}(r)) \cdot \boldsymbol{\rho} \\ &\leq \left(\sum_{\boldsymbol{\rho}} \alpha_{\boldsymbol{\rho}}\right) \left(\max_{\boldsymbol{\sigma} \in \mathcal{I}(G)} f(\boldsymbol{Q}(r)) \cdot \boldsymbol{\sigma}\right) \\ &\leq (1-2\varepsilon-\delta) \left(\max_{\boldsymbol{\sigma} \in \mathcal{I}(G)} f(\boldsymbol{Q}(r)) \cdot \boldsymbol{\sigma}\right). \end{aligned}$$

This will lead to inequality (a). For (b), note that for any graph G, $\mathbf{1}$ can be written as a convex combination of n singleton independent sets. Therefore, it follows that the

$$\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}(r)) \cdot \boldsymbol{\rho} \geq \frac{1}{n} f(\mathbf{Q}(r)) \cdot \mathbf{1}.$$

Therefore, by summing over τ from $C_1 = C_1(\widehat{\mathbf{Q}}(0))$ to $T-1$, we have

$$\mathbb{E}_{\mathbf{x}}\left[L(X(T)) - L(X(C_1))\right] \leq -\frac{\delta}{n} \sum_{\tau=C_1}^{T-1} \mathbb{E}_{\mathbf{x}}[f(\boldsymbol{Q}(\tau)) \cdot \mathbf{1}] + ((B+3)n+\delta)(T-C_1).$$

Finally, from (17), $\widehat{Q}_{\max} = \Theta(Q_{\max})$ for any queue-size vector $\mathbf{Q}$ and its transformation $\widehat{\mathbf{Q}}$ as defined earlier. From the statement of Lemma 13, it follows that $C_1(\widehat{\mathbf{Q}}(0)) = O\left(\log^9 \widehat{Q}_{\max}(0)\right) = O\left(\log^9 Q_{\max}(0)\right)$. Thus, by choice of $C(\mathbf{Q}(0)) = C_1(\widehat{\mathbf{Q}}(0))$ we obtain the desired result and complete the proof of Lemma 11.

5.6. *Proof of Lemma 12.* The proof of Lemma 12 is based on the known classical variational characterization of distribution in the exponential form. Specifically, we state the following proposition which is a direct adaptation of the known results in literature.

Proposition 14 *Let μ be a distribution on a discrete space $\mathcal{S} \subset \{0,1\}^N$ such that*

$$\mu(\boldsymbol{\rho}) \propto \exp\left(\sum_{i=1}^{N} H_i \rho_i\right), \quad \textit{for all} \quad \boldsymbol{\rho} = [\rho_i] \in \mathcal{S}.$$

Here $\mathbf{H} = [H_i] \in \mathbb{R}_+^N$ is a "weight" vector. Let $\boldsymbol{\sigma}$ be random variable with distribution μ. Then,

$$\mathbb{E}_{\mu}\left[\sum_{i=1}^{N} H_i \sigma_i\right] \geq \left(\max_{\boldsymbol{\rho} \in \mathcal{S}} \sum_{i=1}^{N} H_i \rho_i\right) - N.$$

Proof. Let Z be normalization constant of μ, that is

$$\mu(\boldsymbol{\rho}) = \frac{1}{Z} \exp\left(\mathbf{H} \cdot \boldsymbol{\rho}\right), \quad \text{for all} \quad \boldsymbol{\rho} \in \mathcal{S}.$$

Then by definition for any $\boldsymbol{\rho} \in \mathcal{S}$

$$\log Z + \log \mu(\boldsymbol{\rho}) \quad = \quad \mathbf{H} \cdot \boldsymbol{\rho}. \tag{22}$$

Let $\mathcal{M}(\mathcal{S})$ be the space of distributions on $\mathcal{S}$. Define a functional $F : \mathcal{M}(\mathcal{S}) \to \mathbb{R}$ as

$$F(\mu) = \mathbb{E}_\mu \left[\mathbf{H} \cdot \boldsymbol{\sigma}\right] + H_{ER}(\mu), \quad \text{for any} \quad \mu \in \mathcal{M}(\mathcal{S}),$$

where $H_{ER}(\mu)$ is the standard entropy function,

$$H_{ER}(\mu) = -\sum_{\boldsymbol{\rho} \in \mathcal{S}} \mu(\boldsymbol{\rho}) \log \mu(\boldsymbol{\rho}).$$

Then, using (22)

$$\begin{aligned}
F(\nu) &= \sum_{\boldsymbol{\rho} \in \mathcal{S}} \nu(\boldsymbol{\rho}) \left(\mathbf{H} \cdot \boldsymbol{\rho} - \log \nu(\boldsymbol{\rho})\right) \\
&= \sum_{\boldsymbol{\rho} \in \mathcal{S}} \nu(\boldsymbol{\rho}) \left(\log Z + \log \frac{\mu(\boldsymbol{\rho})}{\nu(\boldsymbol{\rho})}\right) \\
&= \log Z + \sum_{\boldsymbol{\rho} \in \mathcal{S}} \nu(\boldsymbol{\rho}) \log \frac{\mu(\boldsymbol{\rho})}{\nu(\boldsymbol{\rho})} \\
&\leq \log Z + \log \left(\sum_{\boldsymbol{\rho} \in \mathcal{S}} \nu(\boldsymbol{\rho}) \frac{\mu(\boldsymbol{\rho})}{\nu(\boldsymbol{\rho})}\right) \quad \text{(from Jensen's inequality)}, \\
&= \log Z.
\end{aligned}$$

In above the equality holds iff $\nu = \mu$. Thus F is maximized at $\mu \in \mathcal{M}(\mathcal{S})$. Now define $\nu(\boldsymbol{\rho}) = \delta_{\boldsymbol{\rho}=\boldsymbol{\rho}^*}$ where

$$\boldsymbol{\rho}^* \in \arg\max_{\boldsymbol{\rho} \in \mathcal{S}} \mathbf{H} \cdot \boldsymbol{\rho}.$$

Then

$$\left(\max_{\boldsymbol{\rho} \in \mathcal{S}} \mathbf{H} \cdot \boldsymbol{\rho}\right) = F(\nu) \;\leq\; F(\mu) = \mathbb{E}_\mu \left[\mathbf{H} \cdot \boldsymbol{\sigma}\right] + H_{ER}(\mu) \leq \mathbb{E}_\mu \left[\mathbf{H} \cdot \boldsymbol{\sigma}\right] + \log |\mathcal{S}|.$$

Since $\mathcal{S} \subset \{0,1\}^N$, the above sequence of inequalities leads to the desired result. This completes the proof of Proposition 14. □

Now we prove Lemma 12 using Proposition 14. Recall that the stationary distribution, say π, of the Glauber dynamics $GD(\boldsymbol{W})$ is indeed of the exponential form with for any $\boldsymbol{\rho} \in \mathcal{I}(G)$,

$$\pi(\boldsymbol{\rho}) \propto \exp\left(\boldsymbol{W} \cdot \boldsymbol{\rho}\right).$$

Therefore, by Proposition 14 we have

$$\mathbb{E}_\pi \left[\boldsymbol{W} \cdot \boldsymbol{\sigma}\right] \;\geq\; \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} \boldsymbol{W} \cdot \boldsymbol{\rho}\right) - n, \tag{23}$$

where $\boldsymbol{\sigma}$ has distribution π and $\log |\mathcal{I}(G)| \leq n$. Next, we relate $\boldsymbol{W} \cdot \boldsymbol{\rho}$ with $f(\mathbf{Q}) \cdot \boldsymbol{\rho}$ to reach the desired conclusion as claimed in Lemma 12.

To this end, let $\boldsymbol{W}$ be weights defined as per (1) based on $\mathbf{Q}, \widetilde{\mathbf{Q}}$ and B. For any coordinate W_i, we have $W_i(t) \geq f(Q_i(\lfloor t \rfloor)) \geq f(Q_i(t)) - 1$. Further, $f(\cdot) = \log\log(\cdot + e)$ is non-negative and 1-Lipschitz. Therefore, we have

$$\begin{aligned} -1 &\leq W_i(t) - f(Q_i)(t) \leq \max\left\{\frac{\varepsilon}{n} f(\widetilde{Q}_{\max,i}(\lfloor t \rfloor)), B\right\} \\ &\leq \max\left\{\frac{\varepsilon}{n}\left(f(Q_{\max}(t)) + 1\right), B\right\} \quad \text{(from Lemma 2 and } Q \text{ is 1-Lipschitz)} \\ &\leq \frac{\varepsilon}{n} f(Q_{\max}(t)) + B. \end{aligned}$$

Therefore, for any $\sigma \in \mathcal{I}(G)$,

$$\begin{aligned} -n &\leq \boldsymbol{W} \cdot \sigma - f(\mathbf{Q}) \cdot \sigma = (\boldsymbol{W} - f(\mathbf{Q})) \cdot \sigma \\ &\leq \|\sigma\|_1 \|\boldsymbol{W} - f(\mathbf{Q})\|_\infty \\ &\leq n\left(\frac{\varepsilon}{n} f(Q_{\max}) + B\right) \\ &= \varepsilon f(Q_{\max}) + nB \\ &\leq \varepsilon\left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\mathbf{Q}) \cdot \boldsymbol{\rho}\right) + nB, \end{aligned} \tag{24}$$

where the last inequality is due to the structure of independent sets $\mathcal{I}(G)$ and it can be easily argued that

$$f(Q_{\max}) \leq \max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\mathbf{Q}) \cdot \boldsymbol{\rho}.$$

Using (23) and (24) we obtain

$$\begin{aligned} \mathbb{E}_\pi\left[f(\boldsymbol{Q}) \cdot \boldsymbol{\sigma}\right] &\geq \mathbb{E}_\pi\left[\boldsymbol{W} \cdot \boldsymbol{\sigma}\right] - \varepsilon\left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}) \cdot \boldsymbol{\rho}\right) - nB \\ &\geq \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} \boldsymbol{W} \cdot \boldsymbol{\rho}\right) - n - \varepsilon\left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}) \cdot \boldsymbol{\rho}\right) - nB \\ &\geq \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\mathbf{Q}) \cdot \boldsymbol{\rho}\right) - n - n - \varepsilon\left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\boldsymbol{Q}) \cdot \boldsymbol{\rho}\right) - nB \\ &= (1 - \varepsilon)\left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} f(\mathbf{Q}) \cdot \boldsymbol{\rho}\right) - (B + 2)n. \end{aligned}$$

This completes the proof of Lemma 12.

5.7. *Network adiabatic Theorem: Proof of Lemma 13.* This section establishes the proof of Lemma 13. In words, Lemma 13 states that the observed distribution of schedules is essentially the same as the desired stationary distribution for all (large enough) time despite the fact that the weights (or queue-sizes) keep changing. In a nutshell, by selection of weight function $f(\cdot) = \log\log(\cdot + e)$ the dynamics of weights become "slow enough", thus allowing for distribution

of scheduling decisions to remain close to the desired stationary distribution at all times. This is analogous to the classical adibatic theorem which states that *if the system is changed gradually (slowly) in a reversible manner and if the system starts in the ground states then it remains in the ground state.*

5.7.1. *Two useful results.* We state two Lemmas that will be useful for establishing Lemma 13.

Lemma 15 *Given $\tau \in \mathbb{N}$, define*

$$\alpha_\tau = \left(f'(\widehat{\boldsymbol{Q}}(\tau)) + f'(\widehat{\boldsymbol{Q}}(\tau+1))\right) \cdot \mathbf{1}.$$

Then the following holds:

1. *For any $\boldsymbol{\rho} \in \mathcal{I}(G)$,*

$$\exp\left(-\alpha_\tau\right) \quad \leq \quad \frac{\pi(\tau+1)(\boldsymbol{\rho})}{\pi(\tau)(\boldsymbol{\rho})} \leq \exp\left(\alpha_\tau\right). \tag{25}$$

2. *And,*

$$\|\pi(\tau+1) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau+1)}} \quad \leq \quad 2\alpha_\tau. \tag{26}$$

Proof. Consider any $\boldsymbol{\rho} \in \mathcal{I}(G)$. Then, from the definition of the stationary distributions $\pi(\tau)$ and $\pi(\tau+1)$,

$$\begin{aligned} \pi(\tau)(\boldsymbol{\rho}) &= \frac{1}{Z(\tau)} \exp\left(\boldsymbol{W}(\tau) \cdot \boldsymbol{\rho}\right) \\ \pi(\tau+1)(\boldsymbol{\rho}) &= \frac{1}{Z(\tau+1)} \exp\left(\boldsymbol{W}(\tau+1) \cdot \boldsymbol{\rho}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{Z(\tau+1)}{Z(\tau)} &= \frac{\sum_{\boldsymbol{\rho} \in \mathcal{I}(G)} \exp\left(\boldsymbol{W}(\tau+1) \cdot \boldsymbol{\rho}\right)}{\sum_{\boldsymbol{\rho} \in \mathcal{I}(G)} \exp\left(\boldsymbol{W}(\tau) \cdot \boldsymbol{\rho}\right)} \\ &\leq \left(\max_{\boldsymbol{\rho} \in \mathcal{I}(G)} \exp\left(\boldsymbol{W}(\tau+1) - \boldsymbol{W}(\tau)\right) \cdot \boldsymbol{\rho}\right). \end{aligned}$$

Recall that by definition (1) and our notation, $\boldsymbol{W}(\cdot) = f(\widehat{\mathbf{Q}}(\cdot))$. Using the fact that for the concave function f, $f(b) - f(a) \leq f'(a)(b-a)$, for any $\boldsymbol{\rho} \in \mathcal{I}(G)$

$$\begin{aligned} \left(\boldsymbol{W}(\tau+1) - \boldsymbol{W}(\tau)\right) \cdot \boldsymbol{\rho} &= \left(f(\widehat{\boldsymbol{Q}}(\tau+1)) - f(\widehat{\boldsymbol{Q}}(\tau))\right) \cdot \boldsymbol{\rho} \\ &\leq f'(\widehat{\boldsymbol{Q}}(\tau)) \cdot \boldsymbol{\rho}. \end{aligned} \tag{27}$$

In above we have used the fact that $\widehat{\mathbf{Q}}(\cdot)$ is 1-Lipschitz. Using (27), we obtain

$$\begin{aligned} \frac{Z(\tau+1)}{Z(\tau)} &\leq \max_{\boldsymbol{\rho} \in \mathcal{I}(G)} \exp\left(f'(\widehat{\boldsymbol{Q}}(\tau)) \cdot \boldsymbol{\rho}\right) \\ &\leq \exp\left(f'(\widehat{\boldsymbol{Q}}(\tau)) \cdot \mathbf{1}\right). \end{aligned} \tag{28}$$

Using a similar argument, we obtain

$$\frac{Z(\tau)}{Z(\tau+1)} \leq \exp\left(f'(\widehat{\boldsymbol{Q}}(\tau+1))\cdot\mathbf{1}\right). \tag{29}$$

From (27) and (29), it follows that for any $\boldsymbol{\rho}\in\mathcal{I}(G)$

$$\begin{aligned}\frac{\pi(\tau+1)(\boldsymbol{\rho})}{\pi(\tau)(\boldsymbol{\rho})} &= \frac{Z(\tau)}{Z(\tau+1)}\exp\left((\boldsymbol{W}(\tau+1)-\boldsymbol{W}(\tau))\cdot\boldsymbol{\rho}\right)\\ &\leq \exp\left(\left(f'(\widehat{\boldsymbol{Q}}(\tau))+f'(\widehat{\boldsymbol{Q}}(\tau+1))\right)\cdot\mathbf{1}\right). \qquad (30)\end{aligned}$$

Similarly,

$$\frac{\pi(\tau)(\boldsymbol{\rho})}{\pi(\tau+1)(\boldsymbol{\rho})} \leq \exp\left(\left(f'(\widehat{\boldsymbol{Q}}(\tau+1))+f'(\widehat{\boldsymbol{Q}}(\tau))\right)\cdot\mathbf{1}\right). \tag{31}$$

Thus (30) and (31) imply the first claim of Lemma 15:

$$\exp\left(-\alpha_\tau\right) \leq \frac{\pi(\tau+1)(\boldsymbol{\rho})}{\pi(\tau)(\boldsymbol{\rho})} \leq \exp\left(\alpha_\tau\right).$$

Here, α_τ can be bounded as follows:

$$\begin{aligned}\alpha_\tau &= \Delta t\left(f'(\widehat{\boldsymbol{Q}}(\tau))+f'(\widehat{\boldsymbol{Q}}(\tau+1))\right)\cdot\mathbf{1}\\ &\leq n\max_i\left(\frac{1}{\left(\widehat{Q}_i(\tau)+e\right)\log\left(\widehat{Q}_i(\tau)+e\right)}+\frac{1}{\left(\widehat{Q}_i(\tau+1)+e\right)\log\left(\widehat{Q}_i(\tau+1)+e\right)}\right)\\ &\leq \frac{2n}{(2n+e)\log(2n+e)} \quad \text{(from (16))}\\ &< 1, \qquad (32)\end{aligned}$$

Using the facts that $1-x\leq e^{-x}$, $e^x\leq 1+2x$ for all $x\in[0,1]$, we have that for any $\boldsymbol{\rho}\in\mathcal{I}(G)$

$$-\alpha_\tau \leq \frac{\pi(\tau)(\boldsymbol{\rho})}{\pi(\tau+1)(\boldsymbol{\rho})}-1 \leq 2\alpha_\tau.$$

Therefore,

$$\left(\frac{\pi(\tau)(\boldsymbol{\rho})}{\pi(\tau+1)(\boldsymbol{\rho})}-1\right)^2 \leq 4\alpha_\tau^2\ .$$

This implies that

$$\begin{aligned}\|\pi(\tau+1)-\pi(\tau)\|_{2,\frac{1}{\pi(\tau+1)}} &= \sqrt{\sum_{\boldsymbol{\rho}\in\mathcal{I}(G)}\pi(\tau+1)(\boldsymbol{\rho})\left(\frac{\pi(\tau)(\boldsymbol{\rho})}{\pi(\tau+1)(\boldsymbol{\rho})}-1\right)^2}\\ &\leq 2\alpha_\tau.\end{aligned}$$

This completes the proof of the second claim of Lemma 15. □

Next, we state and prove a lemma that states that the change in $\pi(\cdot)$ is "small" compared to the "mixing time" of the Glauber dynamics. It will play crucial role in establishing Lemma 13.

Lemma 16 *Given $\varepsilon \in (0,1)$, for any $\tau \in \mathbb{N}$*

$$T_{\tau+1}\alpha_\tau \quad \leq \quad \frac{\varepsilon}{8},$$

where T_τ is the mixing time of the transition matrix $e^{n(P(\tau)-I)}$ defined as

$$T_\tau = \frac{1}{1 - \|e^{n(P(\tau)-I)}\|}.$$

Proof. From Lemma 10, we have that

$$\begin{aligned} T_{\tau+1} &\leq 16n \exp\left[2f(\widehat{Q}_{\max}(\tau+1))\right] \\ &\leq 16n \left(\log\left(\widehat{Q}_{\max}(\tau+1)+e\right)\right)^2. \end{aligned}$$

Now $f'(x) = \frac{1}{(x+e)\log(x+e)} < \frac{1}{x}$. This leads to the following bound:

$$\begin{aligned} T_{\tau+1}\alpha_{\tau,1} &\leq 16n \left(\log\left(\widehat{Q}_{\max}(\tau+1)+e\right)\right)^2 \left[\left(f'(\widehat{\boldsymbol{Q}}(\tau)) + f'(\widehat{\boldsymbol{Q}}(\tau+1))\right)\cdot \mathbf{1}\right] \\ &\leq 16n \left(\log\left(\widehat{Q}_{\max}(\tau+1)+e\right)\right)^2 \left(\frac{n}{\widehat{Q}_{\min}(\tau)} + \frac{n}{\widehat{Q}_{\min}(\tau+1)}\right) \\ &\leq \frac{32n^2 \left(\log\left(\widehat{Q}_{\max}(\tau+1)+e\right)\right)^2}{\widehat{Q}_{\min}(\tau+1) - 1} \\ &\leq \frac{32n^2 \left(\log\left(\widehat{Q}_{\max}(\tau+1)+e\right)\right)^2}{f^{-1}\left(\frac{\varepsilon}{n} f\left(\widehat{Q}_{\max}(\tau+1) - 2n\right)\right) - 1} \qquad \text{(from (18))} \\ &\leq \frac{32n^2 \left(\log\left(x+e\right)\right)^2}{e^{(\log(x-2n+e))^{\frac{\varepsilon}{n}}} - e - 1}, \end{aligned} \tag{33}$$

where $x := \widehat{Q}_{\max}(\tau+1) \geq f^{-1}(B)$. By the definition of B (see Definition 5), the right hand side of (33) is bounded above by $\varepsilon/8$. This completes the proof of Lemma 16. □

5.7.2. *Proof of Lemma 13.* We wish to establish that for $t > C_1(\widehat{\mathbf{Q}}(0))$,

$$\left\|\frac{\mu(t)}{\pi(t)} - 1\right\|_{2,\pi(t)} \quad < \quad \varepsilon. \tag{34}$$

It is enough to show that for $\tau = \lfloor t \rfloor > C_1(\widehat{\mathbf{Q}}(0))$,

$$\left\|\frac{\mu(\tau)}{\pi(\tau)} - 1\right\|_{2,\pi(\tau)} < \varepsilon,$$

since

$$\begin{aligned} \left\|\frac{\mu(t)}{\pi(t)} - 1\right\|_{2,\pi(t)} &\overset{(a)}{=} \left\|\frac{\mu(t)}{\pi(\lfloor t \rfloor)} - 1\right\|_{2,\pi(\lfloor t \rfloor)} \leq \left\|e^{n(t-\lfloor t \rfloor)(P(\tau+1)-I)}\right\| \left\|\frac{\mu(\lfloor t \rfloor)}{\pi(\lfloor t \rfloor)} - 1\right\|_{2,\pi(\lfloor t \rfloor)} \\ &\leq \left\|\frac{\mu(\lfloor t \rfloor)}{\pi(\lfloor t \rfloor)} - 1\right\|_{2,\pi(\lfloor t \rfloor)} < \varepsilon. \end{aligned}$$

In above, for (a), note that $\pi(t) = \pi(\lfloor t \rfloor)$ since $\pi(t)$ is the stationary distribution of $P(t)$ which depends only on $\mathbf{Q}(\lfloor t \rfloor)$ and $\widetilde{\mathbf{Q}}(\lfloor t \rfloor)$. Now we first show that for any $\tau \in \mathbb{N}$ with $\tau \geq C_1(\widehat{\mathbf{Q}}(0)) - 1$

$$\left\| \frac{\mu(\tau+1)}{\pi(\tau)} - 1 \right\|_{2,\pi(\tau)} < \varepsilon/2. \tag{35}$$

To this end, suppose (35) is correct. Then, for $\tau \geq C_1(\widehat{\mathbf{Q}}(0))$,

$$\begin{aligned}
\left\| \frac{\mu(\tau)}{\pi(\tau)} - 1 \right\|_{2,\pi(\tau)} &= \|\mu(\tau) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau)}} \\
&\leq (e^{\alpha_{\tau-1}/2}) \|\mu(\tau) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau-1)}} \qquad \text{(from Lemma 15(1))}, \\
&\leq (1+\alpha_{\tau-1}) \|\mu(\tau) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau-1)}} \\
&\leq (1+\alpha_{\tau-1}) \left(\|\mu(\tau) - \pi(\tau-1)\|_{2,\frac{1}{\pi(\tau-1)}} + \|\pi(\tau-1) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau-1)}} \right) \\
&\leq (1+\alpha_{\tau-1}) \left(\frac{\varepsilon}{2} + 2\alpha_{\tau-1} \right) \qquad \text{(from Lemma 15(2))} \\
&\leq \left(1 + \frac{\varepsilon}{8}\right) \left(\frac{\varepsilon}{2} + \frac{\varepsilon}{4} \right) \qquad \text{(from Lemma 16)} \\
&\leq \varepsilon.
\end{aligned}$$

Therefore, it suffices to establish (35) for completing the proof of Lemma 13. To this end, for simplicity of notation define

$$a_\tau \stackrel{\triangle}{=} \left\| \frac{\mu(\tau+1)}{\pi(\tau)} - 1 \right\|_{2,\pi(\tau)}.$$

Consider the following recursive relation for a_τ:

$$\begin{aligned}
a_{\tau+1} &= \left\| \frac{\mu(\tau+2)}{\pi(\tau+1)} - 1 \right\|_{2,\pi(\tau+1)} \\
&\leq \left\| e^{n(P(\tau+1)-I)} \right\| \left\| \frac{\mu(\tau+1)}{\pi(\tau+1)} - 1 \right\|_{2,\pi(\tau+1)} \qquad \text{(from (8))} \\
&= \left(1 - \frac{1}{T_{\tau+1}}\right) \|\mu(\tau+1) - \pi(\tau+1)\|_{2,\frac{1}{\pi(\tau+1)}} \qquad \text{(from definition of } T_{\tau+1}) \\
&\leq \left(1 - \frac{1}{T_{\tau+1}}\right) \left(\|\mu(\tau+1) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau+1)}} + \|\pi(\tau) - \pi(\tau+1)\|_{2,\frac{1}{\pi(\tau+1)}} \right) \\
&\leq \left(1 - \frac{1}{T_{\tau+1}}\right) \left(\|\mu(\tau+1) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau+1)}} + 2\alpha_\tau \right) \quad \text{(from Lemma 15(2))} \\
&\leq \left(1 - \frac{1}{T_{\tau+1}}\right) \left(e^{\alpha_\tau/2} \|\mu(\tau+1) - \pi(\tau)\|_{2,\frac{1}{\pi(\tau)}} + 2\alpha_\tau \right) \quad \text{(from Lemma 15(1))} \\
&= \left(1 - \frac{1}{T_{\tau+1}}\right) \left(e^{\alpha_\tau/2} a_\tau + 2\alpha_\tau \right) \\
&\leq \left(1 - \frac{1}{T_{\tau+1}}\right) \left((1+\alpha_\tau) a_\tau + 2\alpha_\tau \right) \quad \text{(as } \alpha_\tau < 1 \text{ from (32))}.
\end{aligned} \tag{36}$$

From (36), if we have $a_\tau < \varepsilon/2$ then

$$\begin{aligned} a_{\tau+1} &< \left(1 - \frac{1}{T_{\tau+1}}\right)(\varepsilon/2 + (2+\varepsilon/2)\alpha_\tau) \\ &\leq \left(1 - \frac{1}{T_{\tau+1}}\right)\left(\varepsilon/2 + \frac{\varepsilon}{2T_{\tau+1}}\right) \qquad \text{(from Lemma 16 and } \tfrac{\varepsilon}{8} < \tfrac{\varepsilon}{4+\varepsilon}) \\ &< \varepsilon/2. \end{aligned} \tag{37}$$

Hence, for establishing (35) it is enough to show that there exists a C such that $a_{C-1} < \varepsilon/2$ and $C \leq C_1(\widehat{\boldsymbol{Q}}(0))$. To this end, fix τ and assume $a_s \geq \varepsilon/2$ for all integers $s \leq \tau$. Then, from recursive relation (36) it follows that for $s \leq \tau$,

$$\begin{aligned} a_{s+1} &\leq \left(1 - \frac{1}{T_{s+1}}\right)((1+\alpha_s)\, a_s + 2\alpha_s) \\ &\leq \left(1 - \frac{1}{T_{s+1}}\right)\left((1+\alpha_s)\, a_s + 4\alpha_s \frac{a_s}{\varepsilon}\right) \qquad (\text{as } \ a_s \geq \varepsilon/2) \\ &\leq \left(1 - \frac{1}{T_{s+1}}\right)\left(1 + \left(1 + \frac{4}{\varepsilon}\right)\alpha_s\right) a_s \\ &\leq \left(1 - \frac{1}{T_{s+1}}\right)\left(1 + \frac{1}{T_{s+1}}\right) a_s \quad \text{(from Lemma 16 and } \tfrac{\varepsilon}{8} < \tfrac{\varepsilon}{4+\varepsilon}) \\ &= \left(1 - \frac{1}{T_{s+1}^2}\right) a_s \\ &< e^{-\frac{1}{T_{s+1}^2}} a_s. \end{aligned} \tag{38}$$

Using the (38) for all $s \leq \tau$, we obtain the following:

$$a_{\tau+1} < e^{-\sum_{s=1}^{\tau+1} \frac{1}{T_s^2}} a_0. \tag{39}$$

Now, consider the following.

$$\begin{aligned} \sum_{s=1}^{\tau} \frac{1}{T_s^2} &\geq \sum_{s=1}^{\tau} \frac{1}{16^2 n^2 e^{4f\left(\widehat{Q}_{\max}(s)\right)}} \qquad \text{(from Lemma 10)} \\ &= \frac{1}{16^2 n^2} \sum_{s=1}^{\tau} \frac{1}{e^{4\log\log\left(\widehat{Q}_{\max}(s)+e\right)}} \\ &= \frac{1}{16^2 n^2} \sum_{s=1}^{\tau} \left(\frac{1}{\log\left(\widehat{Q}_{\max}(s)+e\right)}\right)^4 \\ &\geq \frac{1}{16^2 n^2} \sum_{s=1}^{\tau} \left(\frac{1}{\log\left(\widehat{Q}_{\max}(0)+s+e\right)}\right)^4 \\ &> \frac{\tau}{16^2 n^2 \left(\log\left(\widehat{Q}_{\max}(0)+\tau+e\right)\right)^4} \end{aligned}$$

$$\stackrel{(a)}{\geq} \frac{\tau}{16^2 n^2 \sqrt{\tau}\left(\log\left(\widehat{Q}_{\max}(0)+1+e\right)\right)^4}$$
$$= \frac{\sqrt{\tau}}{16^2 n^2 \left(\log\left(\widehat{Q}_{\max}(0)+1+e\right)\right)^4},$$

where (a) follows from the fact that $\tau \geq 1, \widehat{Q}_{\max}(0) \geq f^{-1}(B) \geq 7^7$ and

$$\sqrt{x} \geq \left(\frac{\log(x+y)}{\log(1+y)}\right)^4, \quad \forall x \geq 1, y \geq 7^7.$$

Finally, a_0 is bounded above as

$$\begin{aligned} a_0 &= \left\| \frac{\mu(1)}{\pi(0)} - 1 \right\|_{2,\pi(0)} \leq \sqrt{\frac{1-\pi_{\min}(0)}{\pi_{\min}(0)}} \\ &< \sqrt{\frac{1}{\pi_{\min}(0)}} \qquad < \sqrt{Z(0)} \\ &\leq \sqrt{2^n e^{nf(\widehat{Q}_{\max}(0))}} \\ &= \left(2\log(\widehat{Q}_{\max}(0)+e)\right)^{n/2}. \end{aligned}$$

Now if we choose

$$C = \left\lceil 16^2 n^2 \log^4(\widehat{Q}_{\max}(0)+1+e) \log\left(\frac{2}{\varepsilon}\left(2\log(\widehat{Q}_{\max}(0)+e)\right)^{n/2}\right)\right\rceil^2 + 1,$$

it can be checked that $e^{-\sum_{i=1}^{C-1}\frac{1}{T_i^2}} a_0 < \varepsilon/2$. Therefore, from (39), if $a_s \geq \varepsilon/2$ for all $s < C-1$, $a_{C-1} < e^{-\sum_{i=1}^{C}\frac{1}{T_i^2}} a_0 < \varepsilon/2$. Otherwise, there exists $C' < C-1$ such that $a_{C'} < \varepsilon/2$, which also implies $a_{C-1} < \varepsilon/2$ from (37). In either case, $a_{C-1} < \varepsilon/2$ and it completes the proof of (35) and hence the proof of Lemma 13.

5.8. *Proof of Lemma 8.* We wish to establish that set $C_\kappa = \{\mathbf{x} \in \mathsf{X} : |\mathbf{x}| \leq \kappa\}$ is a closed petit set. By definition, it is closed. To establish that it is a petit set, we need to find a non-trivial measure μ on $(\mathsf{X}, \mathcal{B}_{\mathsf{X}})$ and sampling distribution a on $\mathbb{N}$ so that for any $\mathbf{x} \in C_\kappa$,

$$K_a(\mathbf{x}, \cdot) \geq \mu(\cdot).$$

To construct such a measure μ, we shall use the following Lemma (its proof is presented later).

Lemma 17 *Let network Markov chain $X(\cdot)$ start with state $\mathbf{x} \in C_\kappa$ at time 0, $X(0) = \mathbf{x}$. Then, there exists $T_\kappa \geq 1$ and $\gamma_\kappa > 0$ such that*

$$\sum_{\tau=1}^{T_\kappa} \Pr_{\mathbf{x}}(X(\tau) = \mathbf{0}) \geq \gamma_\kappa, \quad \forall \mathbf{x} \in C_\kappa.$$

Here $\mathbf{0} = (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}) \in \mathsf{X}$ denote the state where all components of $\mathbf{Q}, \tilde{\mathbf{Q}}, \boldsymbol{S}$ are 0 and the schedule is the empty independent set.

In what follows, Lemma 17 will be used to complete the proof of Lemma 8 followed by the proof of Lemma 17. To this end, consider Geometric(1/2) as the sampling distribution a, i.e.

$$a(\ell) = 2^{-\ell}, \quad \ell \geq 1.$$

Let $\boldsymbol{\delta_0}$ be the delta distribution on element $\mathbf{0} \in \mathsf{X}$. Then, define μ as

$$\mu = 2^{-T_\kappa}\gamma_k \boldsymbol{\delta_0}, \quad \text{that is} \quad \mu(\cdot) = 2^{-T_\kappa}\gamma_k \boldsymbol{\delta_0}(\cdot).$$

Clearly, μ is non-trivial measure on $(\mathsf{X}, \mathcal{B}_\mathsf{X})$. With these definitions of a and μ, Lemma 17 immediately implies that for any $\mathbf{x} \in C_\kappa$,

$$K_a(\mathbf{x}, \cdot) \geq \mu(\cdot).$$

This establishes that set C_κ is a closed petit set and this completes the proof of Lemma 8.

5.8.1. *Proof of Lemma 17.* Consider any $\mathbf{x} \in C_\kappa$. By definition total amount of work in each queue is no more than $\kappa + 1$. Consider some large enough (soon to be determined) T_κ. By the property of Bernoulli arrival process, there is a positive probability $\theta^0_\kappa > 0$ of no arrivals happening to the system in time T_κ. Assuming no arrivals happen, we will show that in large enough time t^1_κ, with probability $\theta^1_\kappa > 0$ each queue receives at least $\kappa + 1$ amount of service; and after that in additional time t^2 with positive probability $\theta^2 > 0$ the empty set schedule is reached. This will imply that by defining $T_\kappa \stackrel{\triangle}{=} t^1_\kappa + t^2$ the state $\mathbf{0} \in \mathsf{X}$ is reached with probability at least

$$\gamma_\kappa \stackrel{\triangle}{=} \theta^0_\kappa \theta^1_\kappa \theta^2 > 0.$$

And this will immediately imply the desired result of Lemma 17. To this end, we need to show existence of $t^1_\kappa, \theta^1_\kappa$ and t^2, θ^2 with properties stated above to complete the proof of Lemma 17.

First, existence of $t^1_\kappa, \theta^1_\kappa$. For this, note that the Markov chain corresponding to the scheduling algorithm has time varying transition probabilities and is irreducible over the space of all independent sets, $\mathcal{I}(G)$. If there are no new arrivals and initial $\mathbf{x} \in C_\kappa$, then clearly queue-sizes are uniformly bounded (with bound dependent on κ). Therefore, the transition probabilities of all feasible transitions for this time varying Markov chain is uniformly lower bounded by a strictly positive constant (dependent on κ, n). It can be easily checked that the transition probability induced graph on $\mathcal{I}(G)$ has diameter at most $2n$ and Markov chain transits as per Exponential clock of overall rate n. Therefore, it follows that starting from any initial scheduling configuration, there exists finite time $\widehat{t}_\kappa$ such that a schedule is reached so that any given queue i is scheduled for at least unit amount of time with probability at least $\widehat{\theta}_\kappa > 0$. Here, both $\widehat{t}_\kappa, \widehat{\theta}_\kappa$ depend on n, κ. Therefore, it follows that in time $t^1_\kappa \stackrel{\triangle}{=} (\kappa+1)n\widehat{t}_\kappa$ all queues become empty with probability at least $\theta^1_\kappa \stackrel{\triangle}{=} \left(\widehat{\theta}_\kappa\right)^{n(\kappa+1)}$. Next, to establish existence of t^2, θ^2 as desired, observe that once the system reaches empty queues, it follows that in the absence of new arrivals the empty schedule $\mathbf{0}$ is reached after some finite time t^2 with probability $\theta^2 > 0$ by similar properties of the Markov chain on $\mathcal{I}(G)$ when all queues are 0. Here t^2 and θ^2 are dependent on n only. This completes the proof of Lemma 17.

6. Conclusion. In this paper, we resolved a long-standing and an important question of designing efficient random access algorithm for contention resolution in a network of queues. Our algorithm is essentially a random access based implementation, inspired by Metropolis-Hasting's sampling method, of the classical maximum weight algorithm with "weight" being an appropriate function ($f(x) = \log\log(x+e)$) of the queue size. The key ingredient in establishing the efficiency of the algorithm is a novel *adiabatic*-like theorem for the underlying queueing network. We strongly believe that this *network adiabatic theorem* in particular and methods of this paper in general will be of interest in understanding effect of dynamics in networked system.

REFERENCES

[1] N. Abramson and F. Kuo (Editors). The aloha system. *Computer-Communication Networks*, 1973.

[2] D. J. Aldous. Ultimate instability of exponential back-off protocol for acknowledgement-based transmission control of random access communication channels. *IEEE Transactions on Information Theory*, 33(2):219–223, 1987.

[3] C. Bordenave, D. McDonald, and A. Proutiere. Performance of random medium access - an asymptotic approach. In *Proceedings of ACM Sigmetrics*, 2008.

[4] M. Born and V. A. Fock. Beweis des adiabatensatzes. *Zeitschrift fr Physik a Hadrons and Nuclei*, 51(3-4):165180, 1928.

[5] J. G. Dai. Stability of fluid and stochastic processing networks. *Miscellanea Publication*, (9), 1999.

[6] A. Ephremides and B. Hajek. Information theory and communication networks: an unconsummated union. *IEEE Transactions on Information Theory*, 44(6):2416–2432, 1998.

[7] A. Eryilmaz, A. Ozdaglar, D. Shah, and E. Modiano. Distributed cross-layer algorithms for the optimal control of multi-hop wireless networks. *submitted to IEEE/ACM Transactions on Networking*, 2008.

[8] S. Foss and Takis Konstantopoulos. An overview of some stochastic stability methods. *Journal of Operations Research, Society of Japan*, 47(4), 2004.

[9] R. K. Getoor. Transience and recurrence of markov processes. *In Azma, J. and Yor, M., editors, Sminaire de Probabilits XIV*, page 397409, 1979.

[10] L.A. Goldberg, M. Jerrum, S. Kannan, and M. Paterson. A bound on the capacity of backoff and acknowledgement-based protocols. *Research Report 365, Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK*, January 2000.

[11] Leslie Ann Goldberg. Design and analysis of contention-resolution protocols, epsrc research grant gr/l60982. http://www.csc.liv.ac.uk/ leslie/contention.html, Last updated, Oct. 2002.

[12] A. G. Greenberg, P. Flajolet, and R. E. Ladner. Estimating the multiplicities of conflicts to speed their resolution in multiple access channels. *Journal of the ACM*, 34(2):289–325, 1987.

[13] D. J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Prentice Hall, 2005.

[14] P. Gupta and A. L. Stolyar. Optimal throughput allocation in general random-access networks. In *Proceedings of 40th Annual Conf. Inf. Sci. Systems, IEEE, Princeton, NJ*, pages 1254–1259, 2006.

[15] Johan Hastad, Tom Leighton, and Brian Rogoff. Analysis of backoff protocols for multiple access channels. *SIAM J. Comput*, 25(4), 1996.

[16] L. Jiang and J. Walrand. A distributed csma algorithm for throughput and utility maximization in wireless networks. In *Proceedings of 46th Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL*, 2008.

[17] J.Liu and A. L. Stolyar. Distributed queue length based algorithms for optimal end-to-end throughput allocation and stability in multi-hop random access networks. In *Proceedings of 45th Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL*, 2007.

[18] F. P. Kelly. Stochastic models of computer communication systems. *J. R. Statist. Soc B*, 47(3):379–395, 1985.

[19] F.P. Kelly and I.M. MacPhee. The number of packets transmitted by collision detect random access schemes. *The Annals of Probability*, 15(4):1557–1568, 1987.

[20] I.M. MacPhee. On optimal strategies in stochastic decision processes, d. phil. thesis, university of cambridge, 1989.

[21] P. Marbach. Distributed scheduling and active queue management in wireless networks. In *Proceedings of IEEE INFOCOM, Minisymposium*, 2007.

[22] P. Marbach, A. Eryilmaz, and A. Ozdaglar. Achievable rate region of csma schedulers in wireless networks with primary interference constraints. In *Proceedings of IEEE Conference on Decision and Control*, 2007.

[23] R. Metcalfe and D. Boggs. Distributed packet switching for local computer networks. *Comm. ACM*, 19:395–404, 1976.

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

[25] E. Modiano, D. Shah, and G. Zussman. Maximizing throughput in wireless network via gossiping. In *ACM SIGMETRICS/Performance*, 2006.

[26] J. Mosely and P.A. Humblet. A class of efficient contention resolution algorithms for multiple access channels. *IEEE Transactions on Communications*, 33(2):145–151, 1985.

[27] Microsoft research lab. self organizing neighborhood wireless mesh networks. http://research.microsoft.com/mesh/.

[28] D. Shah. Gossip algorithms. http://web.mit.edu/devavrat/www/gossipbook.pdf, 2008.

[29] D. Shah and D. J. Wischik. Optimal scheduling algorithm for input queued switch. In *Proceeding of IEEE INFOCOM*, 2006.

[30] D. Shah and D. J. Wischik. Heavy traffic analysis of optimal scheduling algorithms for switched networks. *Submitted*, 2007.

[31] S. Shakkottai and R. Srikant. *Network Optimization and Control*. Foundations and Trends in Networking, NoW Publishers, 2007.

[32] A. L. Stolyar. Dynamic distributed scheduling in random access networks. *Journal of Applied Probabability*, 45(2):297–313, 2008.

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

[34] B.S. Tsybakov and N. B. Likhanov. Upper bound on the capacity of a random multiple-access system. *Problemy Peredachi Informatsii*, 23(3):64–78, 1987.