

# On the Asymptotic Validity of the Decoupling Assumption for Analyzing 802.11 MAC Protocol

Jeong-woo Cho, Jean-Yves Le Boudec, *Fellow, IEEE*, and Yuming Jiang

**Abstract**—Performance evaluation of the 802.11 MAC protocol is classically based on the decoupling assumption, which hypothesizes that the backoff processes at different nodes are independent. This decoupling assumption results from mean field convergence and is generally true in transient regime in the asymptotic sense (when the number of wireless nodes tends to infinity), but, contrary to widespread belief, may not necessarily hold in stationary regime. The issue is often related with the existence and uniqueness of a solution to a fixed point equation; however, it was also recently shown that this condition is not sufficient; in contrast, a sufficient condition is a global stability property of the associated ordinary differential equation. In this paper, we give a simple condition that establishes the asymptotic validity of the decoupling assumption for the homogeneous case. We also discuss the heterogeneous and the differentiated service cases and formulate a new ordinary differential equation. We show that the uniqueness of a solution to the associated fixed point equation is not sufficient; we exhibit one case where the fixed point equation has a unique solution but the decoupling assumption is not valid in the asymptotic sense in stationary regime.

**Index Terms**—Mean field theory, ordinary differential equation, fixed point equation, 802.11, decoupling assumption.

## I. INTRODUCTION

THE Wireless LAN standard is evolving towards higher and higher aggregate throughput. The increased maximum bit rate of 802.11n, 600Mbps, along with its easy deployability, suggests the potential use of an 802.11n access point as an wireless router transacting a huge amount of data of many nodes. In this work, we focus on the performance evaluation of 802.11 under the many-node regime as the population size (the number of wireless nodes)  $N$  tends to infinity.

Most existing work on performance evaluation of the 802.11 MAC protocol [2], [11], [12], [18] relies on the “decoupling assumption” which was first adopted in the seminal work by Bianchi [2]. Though having been defined in various ways, it essentially assumes that all the nodes in the same network experience the same *time-invariant* collision probability, with

This work was supported in part by “Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence” appointed by The Research Council of Norway, and funded by The Research Council, NTNU and UNINETT. A part of this work was done when J. Cho was with EPFL, Switzerland.

J. Cho is with the School of Information and Communication Technology at KTH Royal Institute of Technology, Stockholm, Sweden (email: jwcho@kth.se).

J.-Y. Le Boudec is with École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (email: jean-yves.leboudec@epfl.ch).

Y. Jiang is with the Department of Telematics, Norwegian University of Science and Technology (NTNU), N-7491, Trondheim, Norway. (email: jiang@item.ntnu.no).

the direct consequence that the backoff processes are *independent*<sup>1</sup>. This assumption is unavoidable primarily because the stationary distribution of the original Markov chain cannot be explicitly written due to the irreversibility of the chain [12] even for small number of backoff stages, *i.e.*, 3 and 4, unless the population of the network is very small. A similar point was stressed by P. R. Kumar in an interview with Science Watch Newsletter [13]:

*“A good analogy is in thermodynamics. Instead of trying to study the behavior of just three or four molecules and how they move around, you study the behavior of billions and trillions of molecules. ... Similarly, we want to see what you can say about wireless networks in the aggregate.”*

which suggests an analogy of the intractable small-scale problems in different areas. If we *liken* each wireless node to a particle in a physical system, which condition would suffice for every particle being absolutely *decoupled* from the rest?

Once we assume that the decoupling assumption holds, the analysis of the 802.11 MAC protocol leads to a *fixed point equation* (FPE) [12], also called Bianchi’s formula. Kumar *et al.* [12] revisited the FPE and made several remarkable observations, advancing the state of the art to more systematic models and paving the way for more comprehensive understanding of 802.11. Above all, one of the key findings of [12], already adopted in the field [15], [18], is that the full interference model, also called the single-cell model [12] and the main focus of our work, leads to the *backoff synchrony property* [17] which implies the backoff process can be completely separated and analyzed solely through the FPE technique.

This decoupling assumption can be formally justified as a consequence of convergence to mean field and of Sznitman’s result [23]; it can thus only be asymptotically true as the population  $N$  goes to infinity. However, it is recently pointed out by Benaïm and Le Boudec [1, Section 8.2] that Sznitman’s result and convergence to mean field imply the asymptotic validity of the decoupling assumption only in the transient regime, *i.e.*, over a finite horizon, and given some initial conditions. In stationary regime, there may be no decoupling assumption even in the limit of large population size  $N$ . This may happen for example when the ordinary differential equation (ODE) that defines the mean field limit has a limit cycle. In such a case, nodes are asymptotically independent only conditional to the state of the fluid limit. In contrast, if the ODE satisfies a strong global stability property, namely, it has

<sup>1</sup>The meaning of “to decouple” in the literature as well as in our work is an abuse of terminology, in the sense that it has implied not only ‘to decouple nodes’ (independence) but also ‘to have a time-invariant collision probability’.

a unique stationary point to which all trajectories converge, then the decoupling assumption is also valid in stationary regime [1]. For the case of the 802.11 MAC protocol, the stationary points of the ODE are the solutions of the FPE mentioned above. However, existence and uniqueness of a solution to the FPE does not guarantee that all trajectories of the ODE converge to the unique fixed point; in [1], there is a simple example of mean field limit where the FPE has a unique solution but trajectories of the ODE do not converge, in general, to this unique fixed point. Therefore, though the decoupling assumptions that underly Bianchi's formula is plausible and intuitive, the question of its validity can be asked. The main purpose of this paper is to provide an answer to the following question.

“Under which conditions is the decoupling assumption for the model of the 802.11 MAC asymptotically valid?”

To put it another way, we ask whether the FPE method and Bianchi's assumption are valid. To this end, we use mean field theoretic results [1], [7], [22] which state that, as  $N$  tends to infinity, a scaled version of the original Markov chain model of the backoff process in 802.11 MAC protocol converges to a nonlinear *ordinary differential equation* (ODE) so that the asymptotic validity of the decoupling assumption and thus of the FPE boils down to the stability of this ODE. Denoting by  $p_k$  the attempt probability of each wireless node at each time-slot in backoff stage  $k \in \{0, 1, \dots, K\}$ , we assume in what follows that our mean field models are derived when  $K$  is finite and fixed while the number of nodes  $N$  goes to infinity.

In connection with the mean field models, it is worth while to clarify why the relevant works [1], [7], [22] have used a specific intensity scaling regime, under which the activity of each node in backoff stage  $k$  is scaled as follows:

$$p_k := \epsilon(N) \cdot q_k \quad (q_k \text{ is a constant.}) \quad (1)$$

where  $q_k$  is called the *scaled* attempt rate throughout this paper. It is natural to assume that  $\epsilon(N)$  is *vanishing*, i.e.,  $\lim_{N \rightarrow \infty} \epsilon(N) = 0$ . Otherwise, the collision probability between wireless nodes converges to one as  $N$  goes to infinity. More importantly, we have to use an appropriate form of intensity scaling  $\epsilon(N)$  in order to avoid exceptional cases. For example, if  $\epsilon(N)$  decreases faster than  $1/N$  (e.g.,  $\epsilon(N) = 1/N^2$ ), it can be easily seen that the collision probability vanishes as  $N$  tends to infinity, irrespective of whichever backoff stage each node belongs to (we refer to Section II-E for a formal argument). In other words, each node is completely decoupled from the rest. On the other hand, if  $\epsilon(N)$  decreases slower than  $1/N$  (e.g.,  $\epsilon(N) = 1/\sqrt{N}$ ), the collision probability becomes one as  $N$  goes to infinity. That is to say,  $\epsilon(N) = 1/N$  is the **only** intensity scaling regime (up to a constant factor) that deserves to be analyzed.

Under the intensity scaling regime  $\epsilon(N) = 1/N$ , Bordenave *et al.* in [7, Theorem 5.4] studied the **homogeneous** case (all nodes have the same per-stage backoff probabilities) for the case when the number of backoff stages is infinite. They found the following sufficient condition for global stability of the ODE, hence for the asymptotic validity of the decoupling

assumption:

$$q_0 < \ln 2 \quad \text{and} \quad q_{k+1} = q_k/2, \quad \forall k \geq 0 \quad (\text{BMP})$$

where  $q_k$  is the scaled attempt rate in (1) for a node in backoff stage  $k$ . In this paper, we focus on the case where the total number of backoff stages  $K + 1$  is finite, as this is true in practice and in Bianchi's formula. Sharma *et al.* [22] obtained a result for  $K = 1$  and mentioned the difficulty to go beyond. A comprehensive summary of the literature and the outstanding questions raised therein has been recently made by Duffy [9].

We find that not only (i) the monotonicity ((MONO) in Section II) but also (ii) the mild intensity of scaled attempt rates ((MINT) in Section II) imply the uniqueness of a solution to the FPE, which is naturally a necessary condition for stability. Moreover, we prove that the latter (MINT) guarantees the global stability of the ODE. Thus the condition that the attempt rate is *upper-bounded* by the reciprocal of the population, namely  $q_k \leq 1$  for all  $k$ , suffices for the validity of the decoupling assumption. Moreover, for the familiar parameter setting  $q_k = q_0/m^k$  where  $m \geq 1$ , the condition (MINT) suffices for maximizing the aggregate throughput of the network, hence it is a practical condition.

In order to offer various services to higher priority users with additional performance requirements, 802.11e standard introduced the enhanced distributed channel access (EDCA) functionality that has three mechanisms to differentiate the per-class settings of (i) channel holding time, (ii) contention window (CW), and (iii) idle time after each transmission, where the first one has no effect on the backoff processes. Since the second one, CW differentiation, necessarily implies there are two or more classes, we call the corresponding system **heterogeneous**. The third one, called AIFS differentiation, imposes an additional complexity on the Markov chain analysis because whether the users of a class may attempt transmission at each time-slot depends on the type of the current time-slot, which again depends on the activity of the users in the previous time-slot. This mutual interaction of the two evolutions substantially complicates the analysis. As of now, there is *no* ODE in the literature which models AIFS differentiation using an appropriate formalism.

To tackle this problem, it is of importance to observe that the stage evolution of all nodes (or stage density) is much *slower* than the evolution of the type of time-slots under the AIFS differentiation. Thus the former can be taken to be constant by the latter. An application of mean field theoretic result [1, Theorems 1 & 2], formalized based on the same observation, yields an *extended ODE* model of the backoff processes in EDCA-enabled 802.11 networks. We also **formulate** an *extended FPE* on the basis of this ODE, which is satisfied by the equilibrium points of the ODE. It is remarkable that this FPE coincides with that proposed in [12, Section VI].

The versatility of the ODE model is demonstrated by investigating some selected counterexamples. In the first example, we consider a homogeneous system where all nodes use the same parameters and show that the system is *bistable* in that the backoff process, after whirling closely around an equilibrium for a very long time, suddenly jumps into another

equilibrium, and *vice versa*. The FPE model is only capable of identifying three equilibrium points as its solutions, whereas the ODE model is further capable of classifying the two of them into locally stable points and the other into unstable point, accurately reflecting the multistability. The trajectories of the ODE constitute a separatrix which divides the initial condition space into two regions. We also consider a heterogeneous system where the set of nodes are divided into two classes. A delicate determination of the parameters renders the system *oscillatory* such that all trajectories converge to a stable limit cycle formed around an unstable unique equilibrium point where the limit cycle is as determined by the extended ODE. This example also serves as an illustration of the fact that there may be a unique solution to the fixed point equation whereas the decoupling assumption does *not* hold in the asymptotic sense. We also stress that the stability condition established in this work for the first time has been tantalizing other researchers as well, *e.g.*, [22, Appendix B].

The rest of the paper is organized as follows. In Section II, we present a brief overview of recent advances in mean field theory and introduce the associated ordinary differential equation thereof. In Section III, we prove a global stability condition of the ODE, which is in turn shown to be capable of optimizing the throughput. In Section IV, we elaborate on another complexities arising from EDCA and derive its corresponding ODE model. Some counterexamples in Section V illustrate the utility of the ODE models. Concluding remarks and an outstanding problem are given in Section VI.

## II. MEAN FIELD TECHNIQUE REVISITED

To begin with, it should be noted that our analytical model of 802.11 MAC protocol is different from the original one. Thus we first briefly describe the original operation of 802.11 MAC in Section II-A and explore the differences between our model and the real 802.11 MAC protocol in Section II-B.

If the duration of per-stage backoff is taken to be geometric (which is uniform in the standard), the backoff process in 802.11 is governed by a few rules: (i) every node in backoff stage  $k$  attempts transmission with probability  $p_k$  for every time-slot; (ii) if it succeeds,  $k$  changes to 0; (iii) otherwise,  $k$  changes to  $(k + 1) \bmod (K + 1)$  where  $K$  is the index of the highest backoff stage. Markov chain models, which have been widely used in describing complex systems including 802.11, however, very often lead to excessive complications as discussed in Section I. In this section, we present a surrogate tool for the analysis, *mean field theory*. It is noteworthy that the rules used in 802.11, *i.e.*, (i)–(iii), closely resemble the mean field equations laid out below.

### A. Basic Operation of DCF Mode

Time is slotted. Since our analysis is mainly focused on the backoff procedure of 802.11 distributed coordination function (DCF), we call the standardized time interval in the backoff procedure of the 802.11 standard *time-slot*<sup>2</sup> for brevity. The

durations of frames, packets, and inter-frame spaces used in the other procedures are generally different from that of a time-slot.

Each node follows the randomized access procedure of 802.11 DCF. To begin with, each node generates a *backoff value* if it has a data packet to send. Since the backoff procedure of each node is controlled by *inter-frame spaces* that fill in spaces between frames and packets, we introduce them here to help to understand the basic operation of DCF mode. Two types of inter-frame spaces are used in 802.11 DCF, namely, Short Inter-Frame Space (SIFS) and Distributed Inter-Frame Space (DIFS). Each node freezes (stops) the countdown procedure *as soon as* the medium becomes busy. On the other hand, only when the medium is idle for the duration of a Distributed Inter-Frame Space (DIFS), a node may unfreeze (start) its countdown procedure of the backoff and decrements the backoff by one per every time-slot. If the backoff reaches zero, the sender transmits an RTS (ready to send) frame, followed by a CTS (clear to send) from the receiver, a data packet from the sender and an ACK packet from the receiver if RTS/CTS mechanism is switched on. Note that SIFS is smaller than DIFS so that no node is allowed to interrupt a sequence of frames and packets which are spaced out SIFS apart.

There exist  $K + 1$  backoff stages whose indices belong to the set  $\{0, 1, \dots, K\}$  where  $K > 0$ . If a node has not attempted transmission for a data packet yet, the node is supposed to be in the initial backoff stage where the backoff value is drawn uniformly from  $\{0, 1, \dots, 2b_0 - 1\}$  (or  $\{1, 2, \dots, 2b_0\}$ ). Here  $2b_0$  is the contention window that serves as the initial value of a backoff countdown. If two or more wireless nodes finish their countdowns at the same time-slot, there occurs a collision between RTS frames if the RTS/CTS mechanism is switched on, otherwise two or more data packets collide with each other. If there is a collision, each node who participated in the collision multiplies its contention window by the multiplicative factor  $m = 2$ . In other words, each node changes its backoff stage index  $k$  to  $k + 1$  and adopts a new contention window  $2b_{k+1} = 2m^{k+1}b_0$ . If  $k + 1$  is greater than the index of the highest backoff stage number,  $K$ , the node steps back into the initial backoff stage and the contention window is set to  $2b_0$ .

Let  $L$  and  $L_c$  denote the average duration of a successful packet transmission and the fixed duration of a collision, expressed in terms of backoff time-slot. Note here that the length of data packets can be arbitrary random values. Also the fixed overhead for each successful transmission is denoted by  $L_o$ . Note that  $L$ ,  $L_c$  and  $L_o$  do not need to be integer numbers but can be arbitrary positive *real* numbers. In 802.11 DCF, if the RTS/CTS mechanism is used,  $L$  represents the time to transmit an RTS frame, a CTS frame, a data packet, and an ACK packet plus inter-frame spaces, *i.e.*, SIFS and DIFS, where  $L_o$  is  $L$  minus the time to transmit a data packet. The duration  $L_c$ , much smaller than  $L$ , is the time to transmit an RTS frame plus one DIFS.

### B. Differences between Our Model and 802.11 DCF Mode

Our analysis is made tractable by a number of differences between our model used in this paper and the original operation of 802.11 DCF mode. First of all, we take the duration of

<sup>2</sup>This is equivalent to *slot* in the work by Kumar *et al.* [12] (*e.g.*, 20 $\mu$ s in IEEE 802.11b).

per-stage backoff to be geometric as we did at the beginning of Section II. Secondly, the parameter set is fixed for each version of the standard whereas each parameter in our model may be an arbitrary number. For example, in the IEEE 802.11b standard,  $m = 2$ ,  $K = 6$  (7 attempts per packet), and  $2b_0 = 32$  are used.

We also make a few assumptions for tractable analysis.

- **Single-cell assumption:** Most importantly, this work focuses on the performance of *single-cell* 802.11 networks in which all 802.11-compliant nodes are within such a distance from each other that a node can hear whatever the other nodes transmit. Since all nodes freeze their backoff countdown during channel activity, the total time spent in backoff countdowns up to any time is the same for all nodes. Therefore, it is *sufficient* to analyze the backoff process in order to investigate the performance of single-cell networks. This technique has been adopted in many works including [1], [7], [12], [18].
- **Greedy-node assumption:** Secondly, we only consider the case of greedy wireless nodes that persistently contend for the wireless medium.
- **Error-free channel assumption:** Lastly, we assume that the wireless channels are error-free so that failed transmissions are caused only by collisions between RTS frames (for the case of RTS/CTS) or data packets.

### C. Bianchi's Formula

In performance analysis of 802.11, Bianchi's formula and its many variants are probably the most known [2], [8], [11], [12], [15], [16], [18], [20]. Assuming that there are  $N$  nodes, Bianchi's formula can be written compactly in a more general *fixed point equation* (FPE) form:

$$\bar{p} = \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^K \frac{\gamma^k}{p_k}}, \quad (2)$$

$$\gamma = 1 - (1 - \bar{p})^{N-1} \quad (3)$$

where  $\bar{p}$  and  $\gamma$  respectively designate the average attempt probability and collision probability of every node at each time-slot. The attempt probability in backoff stage  $k$  is denoted by  $p_k$  and defined as the inverse of the mean contention window, *i.e.*,  $p_k = 1/(b_k - 1/2)$ . Note that, as long as the backoff stage  $k = 0$  follows backoff stage  $k = K$  for any attempts, the statistics like  $\bar{p}$  and  $\gamma$  are not affected by whether attempts in the highest backoff stage  $K$  are successful or not.

The FPE model has been used as a *de facto* principal tool for the analysis of the 802.11 MAC Protocol. The weak point of the FPE model is that it cannot be concluded entirely from the form of FPE whether its solution (even if it is unique) might be a good first-order approximation of  $\bar{p}$  and  $\gamma$ .

Exactly under which condition the FPE holds is recently being investigated with rigorous mathematical arguments [1], [7], [22], called *mean field independence*. This fundamental approach was originally developed in the two works by Bordenave *et al.* [5] and Sharma *et al.* [21] where the *first* mean field analyses of the 802.11 MAC protocol were performed<sup>3</sup>.

<sup>3</sup>The conference versions of the two works were *submitted* at roughly the same times.

In the rest of the paper, we will refer to their journal versions [7], [22]. Remarkably, Bordenave *et al.* [7] provided a broader mean field framework which extends to multiple-cell networks (*cf.* single-cell assumption in Section II-B) and supports the notion of 'resource'. The particle interaction model proposed in [1] overcomes some limitations and broadens applicability of the model proposed in [7]. The three works [1], [7], [22] have found that, as the number of particles goes to infinity, *i.e.*,  $N \rightarrow \infty$ , the stage distribution of every node evolves according to a set of  $K + 1$  dimensional nonlinear ordinary *differential* equations (ODE) under an appropriate scaling of time.

### D. The Mean Field ODE model

Let us dive into the details of the mean field interaction model for 802.11 used in [1], [7], [22] and how the Markov chain of the model converges to the associated ordinary differential equation.

**Model description:** In our version of 802.11 DCF mode under the assumption made in Section II-B, there are  $N$  wireless nodes evolving in a finite state space  $\{0, \dots, K\}$  at discrete time-slots  $t \in \{0, 1, \dots\}$ . Denoting by  $X_n(t) \in \{0, \dots, K\}$  the backoff stage (the state) of node  $n \in \{1, \dots, N\}$  at time-slot  $t$ , we collect the observations  $X_n(t)$ , for all  $n \in \{1, \dots, N\}$  and compute the relative frequencies, which is called the occupancy measure (or empirical measure). Formally, the occupancy measure in backoff stage  $k$  at (discrete) time-slot  $t$  is defined as

$$\Phi_k(t) := \frac{1}{N} \sum_{n=1}^N 1_{\{X_n(t)=k\}} \quad (4)$$

where  $1_{\{\cdot\}}$  is the indicator function. Let  $\mathbf{A}^T$  be the transpose of a matrix  $\mathbf{A}$ . It can be readily observed that the occupancy measure vector  $\Phi(t) := (\Phi_0(t) \dots \Phi_K(t))^T$  possesses the *Markovian* property because all nodes in the same backoff stage are exchangeable under the greedy-node assumption in Section II-B. Thus the system can be described by  $K$ -dimensional vector  $\Phi(t)$  rather than  $N$ -dimensional vector  $\mathbf{X}(t) := (X_1(t) \dots X_N(t))^T$  though the nodes are not distinguishable any more.

This Markov chain (discrete-time Markov Process) is in fact analogous to the special continuous-time Markov process, *i.e.*, *density dependent population process*, that was used in the seminal work by Kurtz [10, Chapter 11]. Basically, the three works, [1], [7], [22], are nontrivial extensions of the result in [10] to Markov chain version by means of the following scaling technique.

**Key scalings:** Unlike the density dependent population process in [14], our Markov chain in (4) cannot converge to an ODE as  $N \rightarrow \infty$  because a Markov chain evolves at discrete time-slots  $t \in \{0, 1, \dots\}$ . The ODE is derived by means of the following two key scalings.

- **Intensity scaling** is to *slow down* the evolution of each node by a factor of  $\epsilon(N)$ , such that each node in backoff stage  $k$  attempts transmission with probability  $p_k = \epsilon(N) \cdot q_k$ .

- **Time acceleration** is to *accelerate* the evolution of time-slots by  $1/\epsilon(N)$ , such that a variable at  $t$  before this operation is translated into another variable at  $t \cdot \epsilon(N)$ .

The main purpose of using the intensity scaling  $\epsilon(N)$  is to make sure that the **intensity**, defined as the number of state (backoff stage) transitions *per node* per time-slot, vanishes, *i.e.*, converges to 0 as  $N \rightarrow \infty$ . In our context, the intensity is  $p_k = \epsilon(N) \cdot q_k$ , and thus we require

$$\lim_{N \rightarrow \infty} \epsilon(N) = 0.$$

In Section II-E, the implications of intensity scaling will be explored in detail in conjunction with the collision probability and its *physical* meaning.

Since the intensity in the above vanishes, the number of state transitions of *all nodes* per time-slot is order of  $N \cdot \epsilon(N)$  which is dominated by  $N$ . That is, the expected change of  $\Phi_k(t)$  over two consecutive time-slots is order of  $\epsilon(N)$  which tends to zero as  $N \rightarrow \infty$ . However, if we accelerate the evolution of time-slots by  $1/\epsilon(N)$ , the change of  $\Phi_k(t)$  becomes order of one, and thus the time-slots get closer, hence the time continuity. The limit variables which we obtain by applying the time acceleration and the limit operation  $N \rightarrow \infty$  are dubbed *mean field limits (MFL)* in this paper.

To avoid notational confusion, we use capital Greek letters,  $\Phi_k(\cdot)$  (or  $\Phi(\cdot)$ ) and  $\Gamma$ , to denote the original variables and lower-case letters,  $\phi_k$  (or  $\phi(\cdot)$ ) and  $\gamma$  to denote their MFLs.

**The ODE:** The scaled version of the Markov chain converges to an ODE system as  $N \rightarrow \infty$ . It is shown in [7] that, as  $N$  tends to infinity,  $\Phi(t/\epsilon(N)) = (\Phi_0(t/\epsilon(N)) \cdots \Phi_K(t/\epsilon(N)))^T$  converges in probability to  $\phi(t) := (\phi_0(t) \cdots \phi_K(t))^T$  which is the solution of the ODE:

$$\frac{d\phi_0}{dt}(t) = \bar{q}(t) (1 - \gamma(t)) - q_0 \phi_0(t) + \underbrace{q_K \phi_K(t) \gamma(t)}_{\text{inflow from } K} \quad (5)$$

which is the differential equation with respect to  $\phi_0(t)$  and

$$\frac{d\phi_k}{dt}(t) = q_{k-1} \phi_{k-1}(t) \gamma(t) - q_k \phi_k(t), \quad (6)$$

which is the differential equation for  $k \in \{1, \dots, K\}$ . Note that we denote by

$$\bar{q}(t) := \sum_{k=0}^K q_k \phi_k(t) \quad (7)$$

the MFL of the average attempt rate and  $\gamma(t)$  is the MFL of the collision probability to be defined very soon. It is important to note that the above system is *degenerate*<sup>4</sup>, because we also have a *manifold* relation  $\phi_0(t) \equiv 1 - \sum_{k=1}^K \phi_k(t)$ , which can be plugged into (6) to eliminate (5), whereupon we only need to consider the  $K$ -dimensional system (6). We will use the reduced version (6) throughout this work to simplify the exposition. This system (6) will be called **homogeneous** because all nodes adopt the same parameter set  $q_k$  and  $K$ .

The differential equation (6) can be intuitively understood. For example, the first term and second term on the right-hand

side in (6) are respectively the inflow caused by collisions in the  $(k-1)$ th backoff stage and the outflow caused by any attempts in the  $k$ th backoff stage. Note that the underbraced term in (5) was not considered in [7], and exists only in networks with finite backoff stages.

**Collision probability:** The full derivation of the ODE is omitted due to the space limit and a detailed one can be found in the works by Sharma *et al.* [22, Section III] and Bordenave *et al.* [7, Section 5]. However, we believe that the readers can grasp the main idea by looking into the derivation of the MFL of collision probability in the following.

Pick a backoff stage  $k' \in \{0, \dots, K\}$ . For any node in backoff stage  $k'$ , the collision probability of the node at time-slot  $t$  is given by

$$\Gamma(t, k') := 1 - (1 - \epsilon(N) q_{k'})^{-1} \prod_{k=0}^K (1 - \epsilon(N) \cdot q_k)^{N \Phi_k(t)}. \quad (8)$$

This is the probability that at least one other node attempts transmission at time-slot  $t$ . Here we can see that the term  $\epsilon(N) q_{k'}$  vanishes as  $N \rightarrow \infty$ . Thus we can define the MFL of collision probability as follows:

$$\gamma(t) := \lim_{N \rightarrow \infty} \Gamma(t/\epsilon(N), k').$$

We assume the following special intensity scaling regime throughout the rest part of this paper:

$$\epsilon(N) = \frac{1}{N}.$$

It follows from the definition of exponential function

$$\lim_{N \rightarrow \infty} (1 - x/N)^N = \exp(-x) \quad (9)$$

and the definition of  $\bar{q}(t)$  in (7) that

$$\gamma(t) = 1 - e^{-\sum_{k=0}^K q_k \phi_k(t)} = 1 - e^{-\bar{q}(t)} \quad (10)$$

Also, remark that  $\Gamma(t, k)$  depends on backoff stage  $k$ , whereas its MFL  $\gamma(t)$  is common to all nodes.

### E. The Intensity Scaling Regime

Here we expatiate upon our discussion on the intensity scaling  $\epsilon(N)$  in Section I. Note that the expression (10) holds if and only if  $\epsilon(N) \in \Theta(1/N)$ . When  $\epsilon(N) \notin \Theta(1/N)$ , *e.g.*,  $\epsilon(N) = 1/N^2$  or  $\epsilon(N) = 1/\sqrt{N}$ , we can see from the forms of (8) and (9) that  $\gamma(t)$  becomes either zero or one because  $K$  is a finite constant. Summing up, if we consider  $\epsilon(N) \notin \Theta(1/N)$ , the decoupling assumption is asymptotically valid, which is in line with what intuition tells us.

In connection with the above discussion, the intensity scaling technique can be construed as an essential property that must be imposed upon all practical systems where particles (or nodes) *share* a common resource of fixed capacity [7]. If  $\epsilon(N)$  decreases faster than  $1/N$ , *e.g.*,  $\epsilon(N) = 1/N^2$ , the common resource is not used at all as population tends to infinity, hence no collision. On the other hand, if  $\epsilon(N)$  decreases slower than  $1/N$ , *e.g.*,  $\epsilon(N) = 1/\sqrt{N}$ , the common resource is utterly squandered in *attempting* transmission as  $N$  tends to infinity, ending up with collisions all the time. Therefore, the physical meaning of  $\epsilon(N) = 1/N$  is crystal clear.

<sup>4</sup>A degenerate system has a *singular* Jacobian matrix which means that its linearization cannot determine the local stability of the system.

### F. Equilibrium Points

Equating the right-hand sides of (5) and (6) to zero yields the following equilibrium points:

$$\phi_k = \frac{q_0}{q_k} \gamma^k \phi_0, \text{ and } \phi_0 = \frac{\bar{q}}{q_0 \sum_{k=0}^K \gamma^k}$$

whereupon the backoff stage distribution of every node at the equilibrium can be computed as:

$$\phi_k = \frac{\gamma^k}{q_k \sum_{j=0}^K \frac{\gamma^j}{q_j}}.$$

By plugging the manifold relation  $\sum_{k=0}^K \phi_k(t) \equiv 1$  into the above, we can get the following fixed point equation in the stationary regime:

$$\bar{q} = \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^K \frac{\gamma^k}{q_k}}, \quad (11)$$

$$\gamma = 1 - e^{-\bar{q}}. \quad (12)$$

Note that, under the intensity scaling regime, *i.e.*,  $p_k = q_k/N$  and  $\bar{p} = \bar{q}/N$ , (3) becomes

$$\lim_{N \rightarrow \infty} 1 - (1 - \bar{p})^{N-1} = \lim_{N \rightarrow \infty} 1 - \left(1 - \frac{\bar{q}}{N}\right)^{N-1} = 1 - e^{-\bar{q}}$$

which is identical to (12). That is, we do not need to distinguish between (12) and (3) under the intensity scaling regime.

### III. ASYMPTOTIC VALIDATION OF DECOUPLING

The theoretical limit of mean field analysis represented by (6) needs to be clearly understood. The nonlinear ODE model only implies that *any* node will be in backoff stage  $k$  with the common probability  $\phi_k(t)$  under the asymptotic regime. The component ratio  $\phi(t) = (\phi_0(t) \cdots \phi_K(t))^T$ , in general a time-varying solution of (6), is not guaranteed to be constant. Bordenave *et al.* [7, Theorem 5.4] studied its global stability of the asymptotic case when  $K = \infty$ , but the more practical case for finite  $K$  remains to be proved. The need of a proof for finite  $K$  is stressed in [1, pp.833] due to its practical implication. In line with this, by appealing to a Lyapunov function, Sharma *et al.* proved this for the case  $K = 1$  where there are only two backoff stages [22, Lemma 3].

#### A. Main Results

Before presenting the result for finite  $K$  in Theorem 1, we describe *two* different sufficient conditions for the uniqueness of the equilibrium. To simplify the exposition, we first define two conditions:

$$q_k \text{ is nonincreasing in } k. \quad (\text{MONO})$$

$$(11)-(12) \text{ has a unique solution.} \quad (\text{UNIQ})$$

The following lemma holds as long as the right-hand side of (12) is increasing in  $\bar{q}$ . That is, the lemma does not fully exploit the exponential form of (12). It is remarkable that Lemma 1 was originally established by Kumar *et al.* [12, Theorem 5.1]. We give a *simpler* alternative proof in Appendix A based on the method of mathematical induction.

#### Lemma 1 (Monotonicity Implies Uniqueness)

(MONO) implies (UNIQ).

To present the second sufficient condition for the uniqueness of the equilibrium, we define another condition:

$$\bar{q}(t) \leq 1, \quad \forall t \geq 0. \quad (13)$$

As we are interested in *global* stability, we need to show that the solutions of (6) with *any* initial condition converge to the unique equilibrium. Recall that  $\bar{q}(t) = \sum_{k=0}^K q_k \phi_k(t)$ , from the form of which it is clear that (13) holds for any initial condition  $\phi(0)$  *if and only if*

$$q_k \leq 1, \quad \forall k. \quad (\text{MINT})$$

We call this condition (MINT) which is an acronym for ‘Mild INTensity’. In a sense, we can interpret the intensity scaling regime  $p_k = q_k/N$  as a way of weakening the node activity. From this point of view, (MINT) implies we impose an additional constraint upon the node activity.

Interestingly, the above upper bound on the scaled attempt rates also implies (UNIQ). This *intermediate* result is presented here to shorten the proof of Theorem 1, which is the final form of the result.

#### Lemma 2 (Mild Intensity Implies Uniqueness)

(MINT) implies (UNIQ).

*Proof:* Putting  $q_{\max} := \max_{k \in \{0, \dots, K\}} q_k$ , it is clear that (MINT) is equivalent to  $q_{\max} \leq 1$ . First, we have

$$\bar{q} = \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^K \frac{\gamma^k}{q_k}} \leq \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^K \frac{\gamma^k}{q_{\max}}} = q_{\max} \leq 1. \quad (14)$$

Multiplying the both sides of (11) by  $e^{-\bar{q}}$  yields:

$$\bar{q} e^{-\bar{q}} = \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^K \frac{\gamma^k}{q_k}} \cdot e^{-\bar{q}} = \frac{1}{\sum_{k=0}^K \frac{\gamma^k}{q_k}} \cdot \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^{\infty} \gamma^k} \quad (15)$$

where the last equality follows from (12), *i.e.*,  $e^{-\bar{q}} = 1 - \gamma = 1 / \sum_{k=0}^{\infty} \gamma^k$ . The second factor of the last equation of (15) can be rearranged as

$$\begin{aligned} \sum_{k=0}^K \gamma^k (1 - \gamma) &= (1 + \cdots + \gamma^K) - (\gamma + \cdots + \gamma^{K+1}) \\ &= 1 - \gamma^{K+1} = 1 - (1 - e^{-\bar{q}})^{K+1} \end{aligned}$$

which is a decreasing function of  $\bar{q}$ . As the first factor of the last equation of (15) is also a decreasing function of  $\bar{q}$ , (15) is decreasing in  $\bar{q}$ . On the other hand,  $\bar{q} e^{-\bar{q}}$  is increasing in  $\bar{q} \in [0, 1]$  and the range of  $\bar{q} e^{-\bar{q}}$  is  $[0, e^{-1}]$ . Since (15) decreases from  $q_0$  at  $\bar{q} = 0$  to

$$\frac{\sum_{k=0}^K (1 - e^{-1})^k}{\sum_{k=0}^K \frac{(1 - e^{-1})^k}{q_k}} \cdot e^{-1}$$

at  $\bar{q} = 1$ , it suffices to show that the above is less than or equal to  $e^{-1}$ . In the meantime, (MINT) implies that the above is less than or equal to  $e^{-1}$ . Therefore, (11) and (12) have a unique solution. ■

Unlike Lemma 1, the forms of both (11) and (12) are fully exploited for the proof of Lemma 2. Specifically, the fact that  $\bar{q}(1 - \gamma)$  is an increasing function in  $\bar{q}$  over the interval  $[0, 1]$  is used for the proof of Lemma 2.

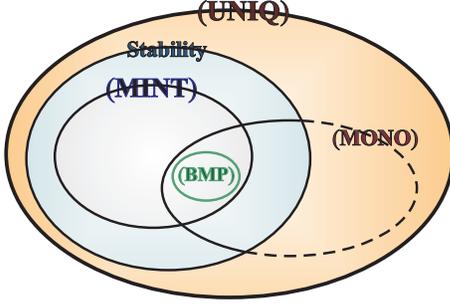


Fig. 1. Logical relations between conditions.

So far we have shown that there are two sufficient conditions, (MONO) and (MINT), for the uniqueness of the equilibrium, (UNIQ), which is naturally a necessary condition for the global stability. We now show that one of them implies the global stability in the following theorem, which also *completes* the logical relations between (MONO), (UNIQ), (MINT), and the global stability, as shown in the Venn diagram in Fig. 1. Since it is not yet clear if there exists any case when (MONO) holds at the same time as the associated ODE is unstable, a part of the set (MONO) is depicted by the dashed line in Fig. 1.

It is remarkable that Lemma 2 is now rendered *obsolete* by the following theorem because the global stability of (6) automatically implies (UNIQ), as clearly depicted in Fig. 1.

**Theorem 1 (Stability Condition)**

(MINT) implies the global stability of (6).

*Proof:* Because  $\bar{q}(t)$  is bounded, there exist  $\bar{q}^l$  and a sequence  $\{\tau_i\}$  such that

$$\liminf_{t \rightarrow \infty} \bar{q}(t) = \bar{q}^l, \quad \lim_{\tau_i \rightarrow \infty} \bar{q}(\tau_i) = \bar{q}^l.$$

Since  $\phi(t)$  is a probability measure on a finite sample space  $\{0, \dots, K\}$ ,  $\phi(t)$  is tight [4]. Appealing to this, we can pick a convergent *subsequence*  $\{t_i\}$  such that  $\lim_{t_i \rightarrow \infty} \phi_k(t_i) = \phi_k(\infty)$  exists.

Defining  $\nu(t) = \inf_{s \geq t} \bar{q}(s)$ , we necessarily have  $\nu(t) \leq \bar{q}(t)$ ,  $\forall t \geq 0$  and  $\lim_{t \rightarrow \infty} \nu(t) = \bar{q}^l$ . Consider the *degenerate* version of (6) which has one additional equation with respect to  $\frac{d\phi_0}{dt}(t)$ . By replacing  $\bar{q}(t)$  with  $\nu(t)$ , we get the following modified ODE:

$$\begin{aligned} \frac{d\phi_0}{dt}(t) &= \nu(t)e^{-\nu(t)} - q_0\phi_0(t) + q_K\phi_K(t)(1 - e^{-\nu(t)}), \\ \frac{d\phi_k}{dt}(t) &= q_{k-1}\phi_{k-1}(t)(1 - e^{-\nu(t)}) - q_k\phi_k(t). \end{aligned}$$

Since  $\nu(t)$  becomes a constant for  $t = \infty$ , this ODE reduces to a linear ODE as  $t \rightarrow \infty$  whose coefficient matrix takes the following form:

$$\begin{pmatrix} -q_0 & 0 & 0 & \dots & 0 & q_K\gamma^l \\ q_0\gamma^l & -q_1 & 0 & \dots & 0 & 0 \\ 0 & q_1\gamma^l & -q_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -q_{K-1} & 0 \\ 0 & 0 & 0 & \dots & q_{K-1}\gamma^l & -q_K \end{pmatrix}$$

where we used  $\gamma^l := \lim_{t \rightarrow \infty} (1 - e^{-\nu(t)})$  for notational simplicity. Applying Gershgorin's circle theorem to the transpose of this coefficient matrix shows that all the eigenvalues are negative hence that  $\varphi(t)$  converges as  $t \rightarrow \infty$ . Thus  $\varphi(\infty)$  should satisfy

$$q_0\varphi_0(\infty) = \bar{q}^l e^{-\bar{q}^l} + q_K\varphi_K(\infty) (1 - e^{-\bar{q}^l}), \quad (16)$$

$$q_k\varphi_k(\infty) = q_{k-1}\varphi_{k-1}(\infty) (1 - e^{-\bar{q}^l}), \quad (17)$$

for  $k \in \{1, \dots, K\}$  because  $\lim_{t \rightarrow \infty} \nu(t) = \bar{q}^l$ . Plugging (17) into (16) yields

$$q_k\varphi_k(\infty) = \bar{q}^l (1 - e^{-\bar{q}^l})^k \left/ \sum_{j=0}^K (1 - e^{-\bar{q}^l})^j \right. . \quad (18)$$

Suppose the initial condition  $\varphi_k(0) = \phi_k(0)$ ,  $\forall k \in \{0, \dots, K\}$ . We have the following equations from the modified ODE:

$$\begin{aligned} \varphi_0(t) &= e^{-q_0 t} \phi_0(0) \\ &+ \int_0^t e^{q_0(s-t)} \{ \nu(s)e^{-\nu(s)} + q_K\varphi_K(s) (1 - e^{-\nu(s)}) \} ds, \end{aligned} \quad (19)$$

$$\begin{aligned} \varphi_k(t) &= e^{-q_k t} \phi_k(0) \\ &+ \int_0^t e^{q_k(s-t)} q_{k-1}\varphi_{k-1}(s) (1 - e^{-\nu(s)}) ds, \end{aligned} \quad (20)$$

where  $k \in \{1, \dots, K\}$ . First we have  $\nu(t) \leq 1$  from the assumption (MINT). Since  $1 - e^{-x}$  and  $xe^{-x}$  terms in the above equations are increasing functions when  $x \in [0, 1]$  and  $\nu(t) \leq \bar{q}(t)$ , it can be checked by plugging (20) into (19)  $K$  times that  $\varphi_0(t) \leq \phi_0(t)$  and hence  $\varphi_k(t) \leq \phi_k(t)$ ,  $\forall t \geq 0$  and  $\forall k \in \{0, \dots, K\}$ . That is,  $\phi_k(t)$  is *lower-bounded* by  $\varphi_k(t)$ .

From (18) and the definition of the subsequence  $\{t_i\}$ , we have the following relation:

$$\sum_{k=0}^K q_k\varphi_k(\infty) = \bar{q}^l = \sum_{k=0}^K q_k\phi_k(\infty).$$

where we recall  $\phi_k(\infty)$  was defined as  $\lim_{t_i \rightarrow \infty} \phi_k(t_i) = \phi_k(\infty)$ . This result taken together with  $\varphi_k(t) \leq \phi_k(t)$  proves  $\varphi_k(\infty) = \phi_k(\infty)$ ,  $\forall k \in \{0, \dots, K\}$ , and therefore,  $\sum_{k=0}^K \varphi_k(\infty) = 1$ . Then it necessarily follows that  $\bar{q}^l$  should satisfy (11) and (12) which have a unique solution by Lemma 2. This implies  $\bar{q}^l = \bar{q}$ .

Note that we can also prove  $\bar{q}^u = \bar{q}$  in a similar way by defining  $\bar{q}^u$  and  $\{t_i\}$  such that  $\limsup_{t \rightarrow \infty} \bar{q}(t) = \bar{q}^u$ ,  $\lim_{t_i \rightarrow \infty} \bar{q}(t_i) = \bar{q}^u$  and  $\lim_{t_i \rightarrow \infty} \phi_k(t_i) = \phi_k(\infty)$ . This will show  $\lim_{t \rightarrow \infty} \bar{q}(t) = \bar{q}$ . That is, there is only one limit point for  $\bar{q}(t)$ .

Finally, we can pick a *new* sequence  $\{\tau_i\}$  such that

$$\liminf_{t \rightarrow \infty} \phi_k(t) = \phi_k^l, \quad \lim_{\tau_i \rightarrow \infty} \phi_k(\tau_i) = \phi_k^l,$$

for all  $k \in \{0, \dots, K\}$ . Using the fact  $\lim_{t \rightarrow \infty} \bar{q}(t) = \bar{q}$ , it can be easily proven that  $\lim_{t \rightarrow \infty} \phi_k(t) = \phi_k$ ,  $\forall k \in \{0, \dots, K\}$ , in a similar way. This establishes that  $(\phi, \bar{q}, \gamma)$  is globally stable. ■

**Remark 1** This result gives an answer to the question raised in Section I and **justifies** the FPE approach used in [2], [8], [11], [12], [15], [16], [18], [20] under a special scaling

regime. That is, the decoupling assumption is validated in the asymptotic sense, as long as the scaled attempt rates are mild, *i.e.*, (MINT).

The result of [7, Theorem 5.4] implies that, for the case  $K = \infty$ , a set of strong conditions is required for the global stability of the ODE; a *monotonicity* condition along with a condition on attempt rate in backoff stage  $k = 0$ , as shown in the condition (BMP) in Section I. These strong conditions, designated also by (BMP) in Fig. 1, were proven to prevent wireless node to escape to infinite backoff stage. We can see that they correspond to a proper subset of the intersection of (MINT) and (MONO). As compared with [7, Theorem 5.4], Theorem 1 is a stronger yet more practical argument due to finite  $K$ .

As shown in Fig. 1, while the monotonicity (MONO) implies only the uniqueness (UNIQ) which is not a decisive factor, (MINT) implies both (UNIQ) and the global stability, assuring the asymptotic validity of the decoupling assumption. It is still *open* whether (MONO) implies the global stability or not.

Informally, the proof of Theorem 1 follows from the fact that the solution  $\phi(t)$  cannot have more than one limit point. The *key observation* underlying its proof is that there exists a stable differential equation which becomes asymptotically linear as  $t \rightarrow \infty$  at the same time as its solution  $\varphi(t)$  lower-bounds  $\phi(t)$  such that  $\phi(t)$  is *squeezed* into  $\phi$  as  $t \rightarrow \infty$ .

It is an intriguing fact that the above theorem may be restated in terms of  $\gamma(t)$  rather than  $q_k$ , hence an alternative interpretation of the theorem: the ODE is globally stable if the collision probability  $\gamma(t) \leq 1 - e^{-1} = 0.632$  for any initial condition  $\phi(0) = (\phi_0(0) \cdots \phi_K(0))^T$ . This interpretation means that if the collision probability is small enough, then the decoupling assumption is asymptotically valid, which appears to be in best agreement with our intuition.

As was mentioned at the beginning of this paper, there is only one intensity scaling regime  $\epsilon(N) = 1/N$  which deserves to be analyzed because, under this regime, it is not clear whether the collision probability would converge to a unique equilibrium point and would stay around there forever. Though we have also shown that (MINT) is a sufficient condition for the asymptotic validity of the decoupling assumption, one may ask in return whether there exist *any* examples where (MINT) does not hold and the decoupling assumption is not asymptotically valid. Yes, there is. We will show in Section V-B that the collision probability may oscillate between two values as time goes if (MINT) is violated.

## B. Achievable Throughput

Recall that  $L$  and  $L_c$  denote the average duration of a successful packet transmission and the fixed duration of a collision, expressed in terms of backoff time-slot. The fixed overhead for each successful transmission is denoted by  $L_o$ . In what follows, we make a mild assumption that  $L_c \geq 1$  which means that the duration of a collision is no less than that of a single backoff time-slot. As we explained in Section II-A,  $L_c$  is an RTS frame plus a DIFS, both of which is larger

than a backoff time-slot in all versions of IEEE 802.11 MAC, regardless of the usage of the RTS/CTS mechanism.

Assuming that  $\bar{q}(t) \rightarrow \bar{q}$  as  $N$  tends to infinity, we can define the achievable throughput or alternatively the MFL of the aggregate throughput, as in [6, Section 5]:

$$\Omega(\bar{q}) := \frac{P_1(\bar{q}) \cdot L}{P_1(\bar{q}) \cdot (L + L_o) + P_0(\bar{q}) + P_c(\bar{q}) \cdot L_c} \quad (21)$$

where  $P_1(\bar{q}) := \bar{q}e^{-\bar{q}}$ ,  $P_0(\bar{q}) := e^{-\bar{q}}$ , and  $P_c(\bar{q}) := 1 - P_1(\bar{q}) - P_0(\bar{q})$  are the MFLs of the probabilities at each time-slot that only one node attempts transmission, none of the users attempts transmission, and at least two users attempt transmissions, respectively. Derivations of these MFLs are similar to that of (8) and thus omitted.

Since (21) holds on the condition that (6) is *globally stable* such that  $\lim_{t \rightarrow \infty} \bar{q}(t) = \bar{q}$ , we can use (21) so long as (MINT) holds. Then the result of Theorem 1 poses another question:

*“Is there  $q_k$  satisfying (MINT) and maximizing (21) as well?”*

Dividing the denominator of (21) by its nominator, we can see that maximizing (21) is equivalent to minimizing

$$\frac{1}{\bar{q}}(1 - L_c) + \frac{e^{\bar{q}}}{\bar{q}}L_c.$$

Differentiating this expression shows that the global maximum of (21) is at the solution of the following equation:

$$\frac{1}{L_c} - 1 = (\bar{q} - 1)e^{\bar{q}} \quad (22)$$

whose left-hand side is monotonically decreasing in  $L_c$  over the domain  $(0, \infty)$  and whose right-hand side is monotonically increasing in  $\bar{q}$  over the same domain. Also both sides have the same range, *i.e.*,  $(-1, \infty)$ . This implies, for each value of  $L_c \in (0, \infty)$ , there exists a unique solution to (22), which is from now on denoted by  $\bar{q} = \bar{q}^*$ . If  $L_c = 1$ , the solution is  $\bar{q}^* = 1$ . Putting these facts together, we can see that, if  $L_c \geq 1$ , there exists a solution  $\bar{q}^* \leq 1$  to (22) which maximizes (21).

It is more important that  $q_k$  satisfies (MINT) because (MINT) is a sufficient condition (and the only sufficient one we know) for the throughput equation (21) to hold. To this aim, we show here that there are infinitely many constructions  $q_k$  what satisfy (MINT) and (MONO) and maximize (21) at the same time. For  $q_k = q_0/m^k$  and  $\bar{q} = \bar{q}^*$ , plugging (12) into (11) yields:

$$\frac{\bar{q}^*}{q_0} = \frac{\sum_{k=0}^K (1 - e^{-\bar{q}^*})^k}{\sum_{k=0}^K (1 - e^{-\bar{q}^*})^k m^k}. \quad (23)$$

The right-hand side of (23) is decreasing in  $m \in (0, \infty)$ . For given optimal solution  $\bar{q}^*$ , one can use (23) to find  $q_0$  and  $m$  which satisfy (MINT) and maximize (21) at the same time. For instance, in order to obtain nonincreasing  $q_k$ , one can simply set  $q_0 = 1$  and compute  $m$  from (23) where  $m \geq 1$  is warranted because the left-hand side of (23) is no greater than 1 and the right-hand side of (23) decreases from 1 at  $m = 1$  to 0 at  $m = \infty$ .

To sum up, for every  $q_0 \in [\bar{q}^*, 1]$  where  $\bar{q}^*$  is the solution to (22), the construction  $q_k = q_0/(m^*)^k$ , where  $m^*$  is the solution to (23), maximizes the aggregate throughput (21) as well as guarantees the global stability of (6).

#### IV. MEAN FIELD WITH SERVICE DIFFERENTIATION

So far the discussion has centered on the homogeneous system where all nodes have the same parameter set. Now we turn to the heterogeneous case arising from the service differentiation mechanisms defined in 802.11e standard. In addition, a special kind of coupling caused by one of the mechanisms necessitates formulating a new ODE model.

##### A. Prioritization Mechanisms

Although three prioritization mechanisms are provided by enhanced distributed channel access (EDCA) functionality, one of which, called transmission opportunity (TXOP) [3], exerts its influence only on time-slots when all nodes are frozen (See Section II-A), hence no need for making an analysis of it. The other two mechanisms are to differentiate per-class settings of

- contention window (CW),
- arbitration interframe space (AIFS).

The first mechanism, CW differentiation, in the present context amounts to per-class setting of  $q_0$  and  $K$ , on the assumption that  $q_k = q_0/2^k$  for  $k \in \{0, \dots, K\}$ . We extend this feature by allowing per-class setting of  $K$  and  $q_k$  for any  $k \in \{0, \dots, K\}$  for the sake of generality and notational aesthetics. Since CW differentiation implies that there are two or more classes, the corresponding system will be called **heterogeneous**, whether the following differentiation is enabled or not.

The second, called AIFS differentiation, is to offer a *soft* non-preemptive prioritization to a certain class by holding back other classes from attempting transmissions for a few time-slots. This prioritization is effectuated by *idling* nodes for different durations, *i.e.*, AIFS, after every transmission. In other words, AIFS differentiation *reserves* a few time-slots for high-priority classes.

The analysis here is presented for the case where there are two classes, *i.e.*, Class H (high) and Class L (low), only to simplify the exposition, but can be extended to arbitrary number of classes. Let us call the time-slots reserved for Class H *reserved* slots, which will correspond to the superscript R. We call the remaining slots following reserved slots *common* slots, corresponding to the superscript C. Note that *both* Class H and Class L users can access the channel during common slots, whereas the backoff procedures of Class L users are suspended during reserved slots. The per-class parameters and occupancy measures are denoted by  $q_k^H, q_k^L, K^H, K^L, \Phi_k^H(t)$  and  $\Phi_k^L(t)$ .

There are two kinds of couplings caused by the above-mentioned prioritization mechanisms.

- **Inter-class coupling:** As compared with the analysis carried out in Section II-C where the stage evolution of nodes depends only on their own stage density, *i.e.*, the occupancy measure  $\Phi(t) = (\Phi_0(t) \dots \Phi_K(t))^T$ , the performance analysis of 802.11 in the presence of CW differentiation is complicated by the very fact that two-class users mutually interact with each other through  $\Phi^H(t) := (\Phi_0^H(t) \dots \Phi_{K^H}^H(t))^T$  and  $\Phi^L(t) :=$

$(\Phi_0^L(t) \dots \Phi_{K^L}^L(t))^T$ . Fortunately, it turns out not very difficult to incorporate this complication into the ODE model in the previous section because we are simply dealing with two evolutions of the same kind.

- **Coupling between two kinds of evolutions:** However, when it comes to AIFS differentiation, the issue is involved by the fact that the stage distribution of nodes in the previous time-slot affects the type of the current time-slot, and besides, the type of the current time-slot also affects the stage distribution of nodes in the next time-slot. That is, there are now two *different* kinds of evolutions, stage evolution of nodes and slot type evolution of time-slots, the latter of which adds a new type of state variable to the Markov chain model in [2]. An interesting point to note is that Sharma *et al.* [22, Section IV] in a similar context also reckoned this difficulty though they have not solved it.

Among many related works for modeling the AIFS differentiation, the works by Robinson and Randhawa [19] and Ramaiyan *et al.* [18] have taken the approach particularly relevant to our work. These works conducted their analyses under the assumption that per-class collision probabilities (or per-class average attempt rates) are constant over all time-slots. In what follows, we first analyze in Section IV-B a Markov chain model of slot type evolution based on the intuition that slot type evolves *much faster* than per-class collision probabilities do and demonstrate in Section IV-C that the result based on this analysis can be validated under the mean field regime by the result of [1], thereby a new variant of ODE model emerges.

##### B. Markov Model for the Evolution of Slot Type

To avoid notational confusion, we use only the original occupancy measures in discrete time, *i.e.*,  $\Phi_k^H(t)$  and  $\Phi_k^L(t)$  in this subsection. The MFLs of these variables will be defined in the next subsection. We first divide the population into two classes such that

$$N^H + N^L = N, \quad \sigma^H := \frac{N^H}{N}, \quad \sigma^L := \frac{N^L}{N}.$$

Without loss of generality, the sets of nodes of Class H and Class L are denoted by  $\mathbb{N}^H := \{1, \dots, N^H\}$  and  $\mathbb{N}^L := \{N^H + 1, \dots, N\}$ . Thus we define the occupancy measures as

$$\Phi_k^H(t) := \frac{1}{N} \sum_{n \in \mathbb{N}^H} 1_{\{X_n(t)=k\}}, \quad \Phi_k^L(t) := \frac{1}{N} \sum_{n \in \mathbb{N}^L} 1_{\{X_n(t)=k\}}$$

so that we have  $\sigma^H = \sum_{k=0}^{K^H} \Phi_k^H(t)$  and  $\sigma^L = \sum_{k=0}^{K^L} \Phi_k^L(t)$ . Since there is no inter-class transition of users,  $\sigma^H$  and  $\sigma^L$  are constant and satisfy the relation  $\sigma^H + \sigma^L = 1$ . In this setting, the probability that one or more nodes attempt transmission at time-slot  $t$  of slot type R or C is given as follows:

$$\Gamma^R(t) := 1 - \prod_{k=0}^{K^H} (1 - \epsilon(N)q_k^H)^{N \cdot \Phi_k^H(t)}, \quad (24)$$

$$\Gamma^C(t) := 1 - (1 - \Gamma^R(t)) \prod_{k=0}^{K^L} (1 - \epsilon(N)q_k^L)^{N \cdot \Phi_k^L(t)}. \quad (25)$$

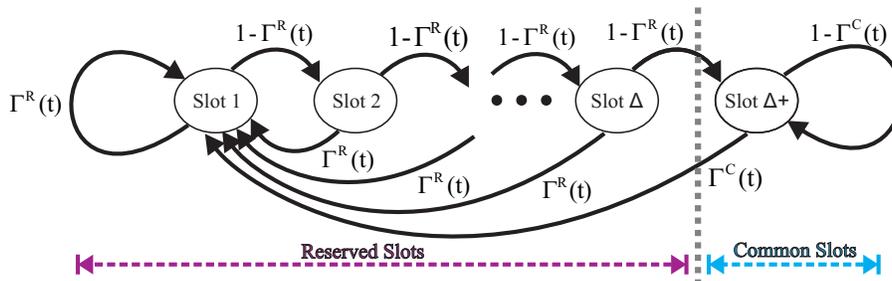


Fig. 2. Evolution of slot type follows a nonhomogeneous Markov chain.

Here we intentionally abuse the notation  $\Gamma$  which is the same as the collision probability in (8) because in the mean field limit the additional term  $q_{k'}/N$  in (8) vanishes, and thus the MFLs of the above equations and (8) are much alike.

Now recall our assumption,  $\epsilon(N) = 1/N$ , which was made in Section II-D. From the viewpoint of each individual node, we can describe AIFS differentiation by only three rules: (i) after any transmission attempt which is either successful or a failure, AIFS procedure is initialized, *i.e.*, a counter value is set to zero; (ii) if the current time-slot is idle, the counter value is incremented by one; (iii) if the counter value reaches its designated per-class AIFS value, the node may attempt transmission with its per-stage probabilities, *i.e.*,  $q_{X_n(t)}^H/N$  and  $q_{X_n(t)}^L/N$ .

Denoting the difference of the two per-class AIFS values by  $\Delta \geq 0$ , we can see that the transition structure based on the aforementioned rules are illustrated by the *nonhomogeneous* Markov chain in Fig. 2, where we used the non-idle probabilities, *i.e.*,  $\Gamma^R(t)$  and  $\Gamma^C(t)$ , and the idle probabilities, *i.e.*,  $1 - \Gamma^R(t)$  and  $1 - \Gamma^C(t)$ , as well. Here in Fig. 2 reserved time-slots and common time-slots are respectively denoted by the notations ‘Slot 1’–‘Slot  $\Delta$ ’ and ‘Slot  $\Delta+$ ’. Note that  $\Delta+$  means that, after any  $\Delta$  or *more* consecutive *idle* backoff time-slots, the corresponding slot-type must be C. It should be clear in Fig. 2 that not only slot-type but also the backoff stages of nodes, *i.e.*,  $\Phi^H(t)$  and  $\Phi^L(t)$ , are also changing over time-slots, hence  $\Gamma^R(t)$  and  $\Gamma^C(t)$  are.

The simplification in the analysis is made based on the following intuition that will be proven correct in Section IV-C:

“As population grows, the stage distribution (density) varies much slower than the type of time-slots.”

This observation follows essentially from the intensity scaling,  $\epsilon(N) = 1/N$ , which leads to separation of time scales. This **timescale decomposition** implies that slot type evolves on a relatively fast time scale, as compared with that of the evolution of occupancy measures. Formally speaking, the occupancy measures,  $\Phi_k^H(t)$  and  $\Phi_k^L(t)$ , evolve at a rate of  $\Theta(1/N)$  which ultimately vanishes as  $N \rightarrow \infty$ , whereas the probability that the slot-type changes for each time-slot does not vanish and remains strictly positive. Therefore, we can analyze the evolution of slot type **as if** the occupancy measures were constant. Working out the balance equations as if the nonhomogeneous Markov chain were homogeneous yields the

following stationary distributions for each slot type:

$$\Pi^R(t) = \frac{\sum_{i=0}^{\Delta-1} (1 - \Gamma^R(t))^i}{\left\{ \sum_{i=0}^{\Delta-1} (1 - \Gamma^R(t))^i \right\} + \frac{(1 - \Gamma^R(t))^\Delta}{\Gamma^C(t)}},$$

$$\Pi^C(t) = \frac{\frac{(1 - \Gamma^R(t))^\Delta}{\Gamma^C(t)}}{\left\{ \sum_{i=0}^{\Delta-1} (1 - \Gamma^R(t))^i \right\} + \frac{(1 - \Gamma^R(t))^\Delta}{\Gamma^C(t)}}$$

which satisfy  $\Pi^R(t) + \Pi^C(t) \equiv 1$ . Note however that in general it is impossible to derive the stationary distribution of nonhomogeneous Markov chains where the transition probabilities are *time-varying*.

### C. Extended ODE Model with Prioritization Mechanisms

The MFLs of  $\Phi_k^H(t)$  and  $\Phi_k^L(t)$  are denoted by  $\phi_k^H(t)$  and  $\phi_k^L(t)$  as in Section II-D. As  $N$  tends to infinity, by manipulation akin to (10), we can show that the collision probabilities for the different types of time-slots, *i.e.*, (24) and (25), become in the mean field regime:

$$\gamma^R(t) := \lim_{N \rightarrow \infty} \Gamma^R(t/\epsilon(N)) = 1 - e^{-\bar{q}^H(t)},$$

$$\gamma^C(t) := \lim_{N \rightarrow \infty} \Gamma^C(t/\epsilon(N)) = 1 - e^{-\bar{q}^H(t) - \bar{q}^L(t)},$$

where the MFLs of the per-class average attempt rates are defined as

$$\bar{q}^H(t) := \sum_{k=0}^{K^H} q_k^H \phi_k^H(t) \quad \text{and} \quad \bar{q}^L(t) := \sum_{k=0}^{K^L} q_k^L \phi_k^L(t).$$

Now we derive the final extended ODE by using the  $\gamma^R(t)$  and  $\gamma^C(t)$ . Let  $\mathbf{g}^R(t)$  and  $\mathbf{g}^C(t)$  denote the rates of change of  $(\phi^H(t), \phi^L(t))$  when all time-slots are of slot type R or C, respectively. As we did in Section II-D, we eliminate manifolds by using the equations  $\phi_0^H(t) \equiv \sigma^H - \sum_{k=1}^{K^H} \phi_k^H(t)$  and  $\phi_0^L(t) \equiv \sigma^L - \sum_{k=1}^{K^L} \phi_k^L(t)$  and hence we consider  $(K^H + K^L)$ -dimensional ODE. Since Class H users are allowed to attempt transmission only at time-slots of slot-type R, whereas Class L users are allowed to do so at time-slots of any slot-type, the

rates of change can be expressed as follows:

$$\mathbf{g}^R(t) = \begin{pmatrix} q_{K^H-1}^H \phi_{K^H-1}^H(t) \gamma^R(t) - q_{K^H}^H \phi_{K^H}^H(t) \\ \vdots \\ \text{---} q_0^H \phi_0^H(t) \gamma^R(t) - q_1^H \phi_1^H(t) \text{---} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$\mathbf{g}^C(t) = \begin{pmatrix} q_{K^H-1}^H \phi_{K^H-1}^H(t) \gamma^C(t) - q_{K^H}^H \phi_{K^H}^H(t) \\ \vdots \\ \text{---} q_0^H \phi_0^H(t) \gamma^C(t) - q_1^H \phi_1^H(t) \text{---} \\ \text{---} q_{K^L-1}^L \phi_{K^L-1}^L(t) \gamma^C(t) - q_{K^L}^L \phi_{K^L}^L(t) \text{---} \\ \vdots \\ q_0^L \phi_0^L(t) \gamma^C(t) - q_1^L \phi_1^L(t) \end{pmatrix}.$$

It is remarkable that all the expressions intuitively derived based on the timescale decomposition in Section IV-B can be formally **justified** by the result of Benaïm and Le Boudec [1]. That is, the adopted simplification can be regarded as a natural consequence of the limiting regime. Formally speaking, we can apply [1, Theorems 1 & 2]<sup>5</sup> to show that the resultant derivatives of  $(\phi^H(t), \phi^L(t))$  become:

$$\frac{d}{dt} \begin{pmatrix} \phi^H(t) \\ \phi^L(t) \end{pmatrix} = \mathbf{g}^R(t) \cdot \pi^R(t) + \mathbf{g}^C(t) \cdot \pi^C(t) \quad (26)$$

where  $\pi^R(t)$  and  $\pi^C(t)$  take the following forms

$$\pi^R(t) = \frac{\sum_{i=0}^{\Delta-1} (1 - \gamma^R(t))^i}{\left\{ \sum_{i=0}^{\Delta-1} (1 - \gamma^R(t))^i \right\} + \frac{(1 - \gamma^R(t))^\Delta}{\gamma^C(t)}},$$

$$\pi^C(t) = \frac{\frac{(1 - \gamma^R(t))^\Delta}{\gamma^C(t)}}{\left\{ \sum_{i=0}^{\Delta-1} (1 - \gamma^R(t))^i \right\} + \frac{(1 - \gamma^R(t))^\Delta}{\gamma^C(t)}}.$$

This result has the implication that we can derive (26) as a **linear combination** of the MFL vectors, *i.e.*,  $\mathbf{g}^R(t)$  and  $\mathbf{g}^C(t)$ , with the coefficients, *i.e.*,  $\pi^R(t)$  and  $\pi^C(t)$ , which can be defined as

$$\pi^R(t) := \lim_{N \rightarrow \infty} \Pi^R(Nt), \quad \pi^C(t) := \lim_{N \rightarrow \infty} \Pi^C(Nt)$$

where  $\Pi^R(\cdot)$  and  $\Pi^C(\cdot)$  are the stationary distributions we computed from Fig. 2 in Section IV-B as if the nonhomogeneous Markov chain were homogeneous.

Finally, after some manipulation of (26), we have the following enhanced ordinary differential equation:

$$\frac{d\phi_k^H}{dt}(t) = q_{k-1}^H \phi_{k-1}^H(t) \gamma^H(t) - q_k^H \phi_k^H(t), \quad (27)$$

$$\frac{d\phi_k^L}{dt}(t) = \pi^C(t) \{ q_{k-1}^L \phi_{k-1}^L(t) \gamma^C(t) - q_k^L \phi_k^L(t) \}, \quad (28)$$

where (27) and (28) respectively hold for  $k \in \{1, \dots, K^H\}$  and  $k \in \{1, \dots, K^L\}$ . Here we use the following shorthand notation:

$$\gamma^H(t) = \pi^R(t) \gamma^R(t) + \pi^C(t) \gamma^C(t)$$

whose form is obvious from (26).

In the stationary regime, we can get the following fixed point equation:

$$\bar{q}^H = \sigma^H \frac{\sum_{k=0}^{K^H} (\gamma^H)^k}{\sum_{k=0}^{K^H} \frac{(\gamma^H)^k}{q_k^H}}, \quad (29)$$

$$\bar{q}^L = \sigma^L \frac{\sum_{k=0}^{K^L} (\gamma^C)^k}{\sum_{k=0}^{K^L} \frac{(\gamma^C)^k}{q_k^L}}, \quad (30)$$

$$\gamma^H = \pi^R \left( 1 - e^{-\bar{q}^H} \right) + \pi^C \left( 1 - e^{-\bar{q}^H - \bar{q}^L} \right), \quad (31)$$

$$\gamma^C = 1 - e^{-\bar{q}^H - \bar{q}^L}. \quad (32)$$

**Remark 2** It is remarkable that the extended ODE model laid out in (27) and (28) *encompasses* the homogeneous system in Section II, and the heterogeneous system in Section IV as well, which has the two prioritization functionalities.

For instance, if  $\Delta = \infty$ , we have  $\pi^R(t) = 1$  and the ODE model reduces to the homogeneous system (6). On the other hand, if  $\Delta = 0$ , we have  $\pi^C(t) = 1$  and the ODE model reduces to a purely heterogeneous system, implying that the AIFS differentiation is disabled.

What is the most surprising is that the FPE (29)-(32) coincides with that proposed in [18, Section VI], which was derived rather intuitively.

In the following, we introduce three new conditions akin to those in Section II-D.

$$q_k^H \text{ and } q_k^L \text{ are nonincreasing in } k, \quad (33)$$

$$(29)-(32) \text{ has a unique solution,} \quad (34)$$

$$q_k^H \leq 1 \text{ and } q_k^L \leq 1, \quad \forall k. \quad (35)$$

By adopting the above conditions we present two lemmas.

**Lemma 3 (Monotonicity Implies Uniqueness)**

(33) implies (34).

**Lemma 4 (Mild Intensity Implies Uniqueness)**

(35) implies (34).

Proofs of the above two lemmas are in Appendix. It is of importance to note that these lemmas are of even *greater* generality because they hold for all  $\Delta \geq 0$  and  $\Delta = \infty$  as well, implying that Lemmas 1 and 2 respectively correspond to the special cases of Lemmas 3 and 4, *i.e.*, the case  $\Delta = \infty$ .

We are not able to prove the equivalent of Theorem 1 for the case where there are more than one class due to the inter-class coupling arising from CW differentiation. This coupling makes it technically challenging to find a stable ODE, which would bound the solution of the ODE as in the proof of Theorem 1. In the meantime, the other coupling induced by AIFS differentiation does not seem to cause a major technical difficulty. As of now, we have to be content with having stated the problem precisely with its inherent technical difficulty.

<sup>5</sup>The corresponding assumptions can be easily checked.

## V. SELECTED COUNTEREXAMPLES

Before proceeding to selected examples, the gap between the ODE model and the backoff processes in 802.11 must be bridged. This gap emerged right on applying the intensity scaling in Section II-D. The scaling relation  $p_k = q_k/N$  suggests to us that replacing  $q_k$  by  $Np_k$  should yield a reasonable approximation if  $p_k$  is small. Plugging this approximation and removing the time acceleration from (6), we have

$$\frac{d\phi_k}{dt}(t) = p_{k-1}\phi_{k-1}(t)\gamma(t) - p_k\phi_k(t) \quad (36)$$

where  $\gamma(t) := 1 - e^{-N\bar{p}(t)}$  and  $\bar{p}(t) := \sum_{k=0}^K p_k\phi_k(t)$ .

In this section, we provide two counterexamples which will demonstrate versatility of the ODE model. In particular, two goals of the simulation are as follows:

- By comparing the trajectories of the ODE model (36) with the simulation result of the corresponding Discrete Time Markov Chain (DTMC), we show that the simplistic ODE model is accurate enough to provide us insights into the formidably complex DTMC.
- For the case the stability condition (MONO) is violated, the examples illustrate two major unstable behavior patterns of the system.

Note that all Markov chains used for the simulation are ergodic: statistically speaking, the two systems will forget their initial states after evolving for a long enough time. To this aim, we have run each of the DTMC simulations for 120,000,000 backoff time-slots. It is remarkable that we must run the DTMC simulations for a very long duration because the DTMC in Section V-A exhibits a phenomenon called bistability, which can be observed by running simulations for a relatively long time (*e.g.*, see Fig. 3(b)).

To obtain the short-term average statistics, the entire duration of each simulation is divided into disjoint intervals of 2,000 time-slots and each short-term average data point was calculated over one of the disjoint intervals. We assume that all wireless nodes are in backoff stage 0 at the initial time-slot.

### A. Example 1: Multistability

Consider the **homogeneous** system (36). Plugging (2) into (3) yields:

$$f(\gamma) := 1 - \exp\left(-N \frac{\sum_{k=0}^K \gamma^k}{\sum_{k=0}^K \frac{\gamma^k}{p_k}}\right) - \gamma = 0 \quad (37)$$

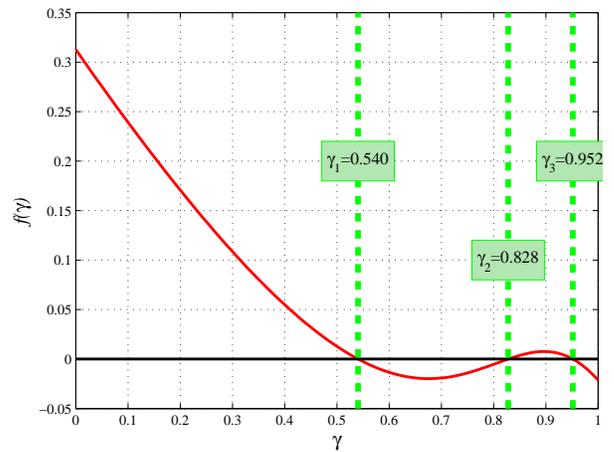
which is a function of only  $\gamma$ . Consider the following multistability example where there are  $N = 1200$  nodes and  $K + 1 = 13$  backoff stages. The attempt probability at each backoff stage  $p_k$  is

$$(p_0, p_1, \dots, p_{12}) = \left(\frac{1}{3200}, \frac{1}{160}, \frac{m}{160}, \dots, \frac{m^{11}}{160}\right)$$

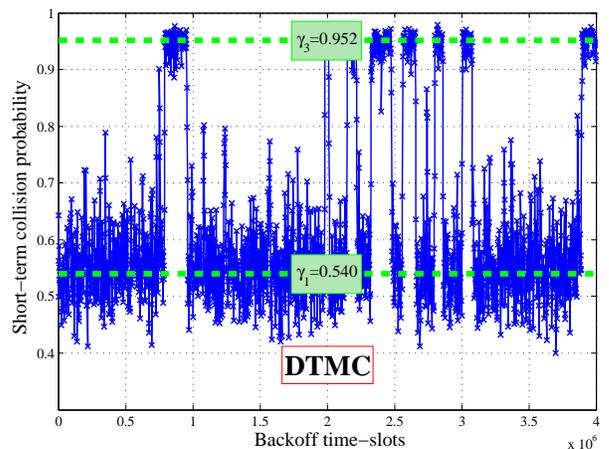
where  $m = 6/5 = 1.2$ . The roots of (37) can be computed from Fig. 3(a) as

$$(\gamma_1, \gamma_2, \gamma_3) = (0.540, 0.828, 0.952).$$

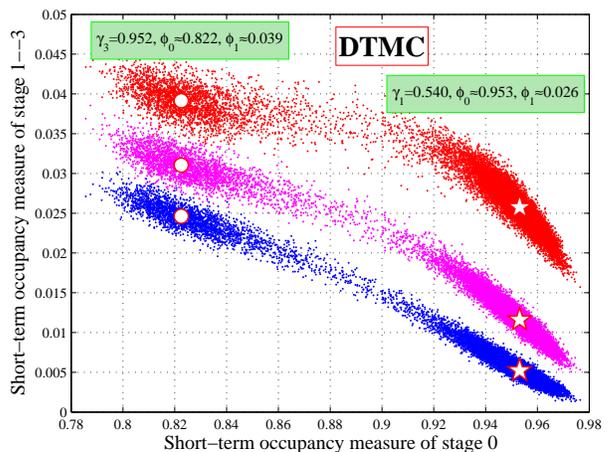
The instantaneous collision probability for each 2000 backoff time-slots is shown in Fig. 3(b), which tends to concentrate



(a)  $f(\gamma)$  versus  $\gamma$



(b) Short-term average collision probability vs. backoff time-slots



(c) Short-term average occupancy measure in backoff stages,  $(\phi_0(t), \phi_1(t))$ ,  $(\phi_0(t), \phi_2(t))$  and  $(\phi_0(t), \phi_3(t))$ . Stars (\*) and circles (o): mean field limits; dots: DTMC simulation.

Fig. 3. Bistability Example: There are three solutions to the fixed point equation, two of which ( $\gamma_1$  and  $\gamma_3$ ) are stable and the other one ( $\gamma_2$ ) is unstable. Short-term average statistics measured for each 2000 backoff time-slots suggest bistability.

around  $\gamma_1 = 0.540$  and  $\gamma_3 = 0.952$ . Note that the average collision probability for the entire duration of the simulation is 0.832 that is neither  $\gamma_1$  nor  $\gamma_3$ . Recall that  $\phi_k(t)$  denotes the fraction of nodes in backoff stage  $k$ . Fig. 3(c) shows the

short-term average of the fraction of nodes in backoff stage  $k \in \{1, 2, 3\}$  versus that in backoff stage 0. From the top to the bottom, the short-term occupancy measures of stages 1 – 3 are shown in order, where the two kinds of markers, *i.e.*, circle ( $\circ$ ) and star ( $\star$ ), stand for the occupancy measures at two equilibria,  $\gamma_3$  and  $\gamma_1$ , which are computed from (36). The *bistability* of this system is precisely predicted from either two modes of behavior of (36) or the eigenvalues of Jacobian matrices at the three equilibrium points.

### B. Example 2: Stable Oscillation

We have managed to discover a rare example by delving into the **heterogeneous** system, without AIFS differentiation, *i.e.*,  $\Delta = 0$ , which in turn leads to  $\pi^R = 0$  and  $\pi^C = 1$ . Suppose there are two classes H and L such that population of each class is  $N^H = N^L = 640$ . The numbers of backoff stages are assumed to be equal, *i.e.*,  $K^H + 1 = K^L + 1 = 21$ . The attempt probability at each backoff stage is:

$$(p_0^H, p_1^H, \dots, p_{20}^H) = \left( \frac{1}{2400}, \frac{1}{480}, \frac{m}{40}, \dots, \frac{m^{19}}{40} \right)$$

$$(p_0^L, p_1^L, \dots, p_{20}^L) = \left( \frac{1}{3840}, \frac{1}{64}, \frac{1}{64}, \dots, \frac{1}{64} \right)$$

where  $m = 4/5$ . It is easy to verify that the corresponding fixed point equation takes the following form:

$$f(\gamma) := 1 - \prod_{X \in \{H,L\}} \exp \left( -N^X \frac{\sum_{k=0}^{K^X} \gamma^k}{\sum_{k=0}^{K^X} \frac{\gamma^k}{p_k^X}} \right) - \gamma = 0$$

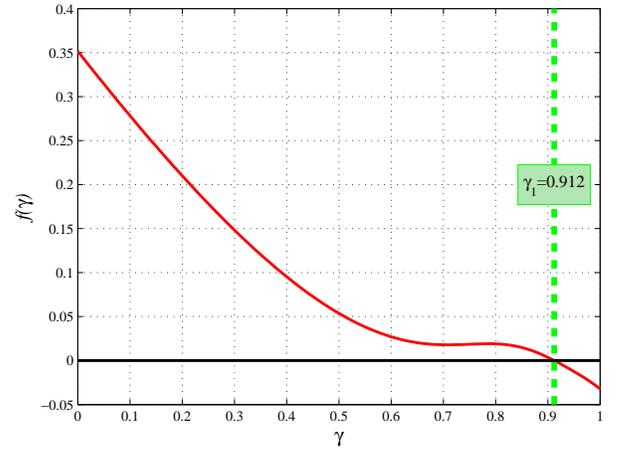
which has the following **unique** solution as shown in Fig. 4(a):

$$\gamma^H = \gamma^R = \gamma^C = \gamma_1 = 0.912.$$

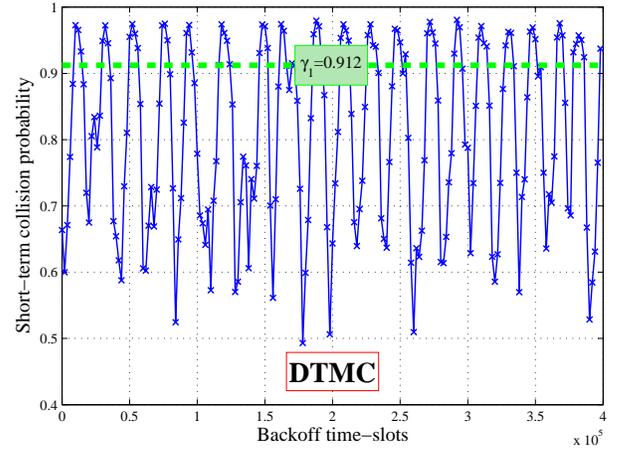
Since there is only one solution, one might be much inclined to hazard the conjecture by Bianchi *et al.* [2], [12] that the collision probability is approximately  $\gamma_1$ . However, there is a stable limit cycle around this equilibrium. In other words, the oscillation is stable, *i.e.*, not transient but lasting forever. The event-average collision probability obtained through simulations is 0.869 which is less than  $\gamma^H$  or  $\gamma^C$ .

We can see from Fig. 4(b) that, unlike the previous example, the trajectory of instantaneous collision probability forms almost periodic oscillation and does not tend to concentrate around the unique equilibrium  $\gamma_1$ . Though the oscillation is not deterministic but stochastic, it clearly persists as time goes to infinity. The period of the oscillation empirically can be computed from Fig. 4(b) as between 19000 and 20000 time-slots. The oscillation and its period are exactly predicted from the trajectories of the ODE model (solid lines) as shown in Fig. 4(c). The instability of  $\gamma_1$  can be decided by the eigenvalues of the corresponding Jacobian matrix.

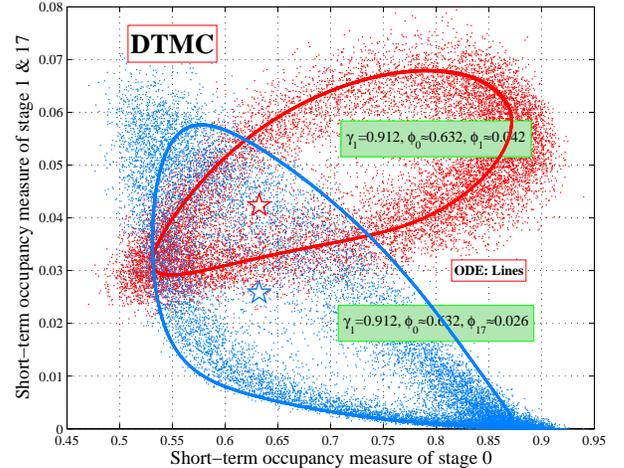
The decoupling assumption does not hold in the asymptotic sense; in contrast, nodes are coupled by the oscillations of the occupancy measure, an emerging property of the system dynamics.



(a)  $f(\gamma)$  versus  $\gamma$



(b) Short-term average collision probability vs. backoff time-slots



(c) Short-term average occupancy measure in backoff stages,  $(\phi_0(t), \phi_1(t))$  and  $(\phi_0(t), \phi_{17}(t))$ . *Solid lines and stars* ( $\star$ ): mean field limits; *dots*: DTMC simulation.

Fig. 4. Oscillation Example: There is a unique solution ( $\gamma_1$ ) to the fixed point equation but the decoupling assumption does *not* hold in the asymptotic sense. Short-term average statistics measured for each 2000 backoff time-slots suggest stable oscillation around the unique equilibrium.

## VI. CONCLUDING REMARKS WITH A CONJECTURE

Since it is axiomatic that the fixed point equation (FPE), called Bianchi's formula, must have a unique solution in order

to provide an approximation, there has been a speculation that the uniqueness of the solution might assure the validity of the FPE, which has been the main subject of previous approaches by Kumar *et al.* [12], and Ramaiyan *et al.* [18]. One counterexample in our paper has shown that this speculation is not always true, putting another emphasis on asymptotic validation of the decoupling assumption which underlies the formula.

Thanks to recent advances in mean field theory [1], [7] and also [22], we have analyzed the validity of the FPE by determining the stability of an ordinary differential equation (ODE). In the course of establishing stability, we obtained an illuminating insight that not only monotonicity but also mildness of scaled attempt rate guarantees the uniqueness of the equilibrium, which made the logical relations between them clear. Paradoxically, the mathematical formalism of mean field theory presented us a *succinct* stability condition (MINT), whose main implication is as follows: to achieve perfect decoupling between nodes as population  $N$  grows, in addition to reducing the attempt probability at  $k$ th backoff to  $q_k/N$ , we need to *further* diminish the node activity such that the scaled attempt rate satisfies  $q_k \leq 1$ . The existence of such an upper bound appears to be in best agreement with our usual intuition. We also have shown that there are infinitely many constructions of  $q_k$  which maximize the aggregate throughput as well as satisfy (MINT) and (MONO), hence the condition (MINT) is practical as well.

Though an EDCA prioritization mechanism causes a new type of coupling between the evolutions of per-class remaining idle times and backoff stages, which has been an intricate complication [22], another penetration, also formalized by mean field argument, has led us to an extended form of an ODE model spinning off a generalized FPE as well.

Lastly, we conjecture that (MINT) implies the global stability of (27) and (28) as well, as observed in our exhaustive simulations. We believe that it is provable with a Lyapunov function though the form of which is unknown yet. Although theoretical support to this conjecture is not available, we hope the discussion can introduce the challenging side of the open stability problem.

## REFERENCES

- [1] M. Benaïm and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Perf. Eval.*, 65(11-12):823–838, Nov. 2008.
- [2] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE J. Select. Areas Commun.*, 18(3):535–547, Mar. 2000.
- [3] G. Bianchi, I. Tinnirello, and L. Scalia. Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations. *IEEE Network*, 19(4):28–34, July 2005.
- [4] P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 2nd ed., 1999.
- [5] C. Bordenave, D. McDonald, and A. Proutiere. Random multi-access algorithm: A mean field analysis. In *Proc. 43th Annual Allerton Conf. Comm., Control, and Comput.*, Monticello, IL, Sept. 2005.
- [6] C. Bordenave, D. McDonald, and A. Proutiere. Performance of random medium access control: An asymptotic approach. In *Proc. ACM Sigmetrics*, June 2008.
- [7] C. Bordenave, D. McDonald, and A. Proutiere. A particle system in interaction with a rapidly varying environment: Mean field limits and applications. *Networks and Heterogeneous Media*, 5(1):31–62, Mar. 2010.

- [8] M. Carvalho and J. Garcia-Luna-Aceves. A scalable model for channel access protocols in multihop ad hoc networks. In *Proc. ACM MobiCom*, Sept. 2004.
- [9] K. Duffy. Mean field markov models of wireless local area networks. *Markov Processes and Related Fields*, 16(2):295–328, 2010.
- [10] S. Ethier and T. Kurtz. *Characterization and convergence*. Wiley, 1986.
- [11] M. Garetto, T. Salonidis, and E. Knightly. Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks. *ACM/IEEE Trans. Networking*, 16(4):864–877, Aug. 2008.
- [12] A. Kumar, E. Altman, D. Miorandi, and M. Goyal. New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs. *ACM/IEEE Trans. Networking*, 15(3):588–601, June 2007.
- [13] P. Kumar. An interview with Dr. P. R. Kumar. *Science Watch Newsletter*, available at <http://esi-topics.com/wireless/interviews/PRKumar.html>, June 2006.
- [14] T. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6:223–240, 1978.
- [15] B.-J. Kwak, N.-O. Song, and L. Miller. Performance analysis of exponential backoff. *ACM/IEEE Trans. Networking*, 13(2):343–355, Apr. 2005.
- [16] K. Medepalli and F. Tobagi. Towards performance modeling of IEEE 802.11 based wireless networks: A unified framework and its applications. In *Proc. IEEE Infocom*, Apr. 2006.
- [17] A. Proutiere. Pushing mean field asymptotics to the limits: Stability and performance of random medium access control. In *Proc. Stochastic Networks Conference*, ENS, Paris, June 2008.
- [18] V. Ramaiyan, A. Kumar, and E. Altman. Fixed point analysis of single cell IEEE 802.11e WLANs: Uniqueness, multistability. *ACM/IEEE Trans. Networking*, 16(5):1080–1093, Oct. 2008.
- [19] J. Robinson and T. Randhawa. Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function. *IEEE J. Select. Areas Commun.*, 22(5):917–928, June 2004.
- [20] T. Sakurai and H. Vu. MAC access delay of IEEE 802.11 DCF. *IEEE Wireless Commun.*, 6(5):1702–1710, May 2007.
- [21] G. Sharma, A. Ganesh, and P. Key. Performance analysis of contention based medium access control protocols. In *Proc. IEEE Infocom*, Barcelona, Spain, Apr. 2006.
- [22] G. Sharma, A. Ganesh, and P. Key. Performance analysis of contention based medium access control protocols. *IEEE Trans. Inform. Theory*, 55(4):1665–1681, Apr. 2009.
- [23] A.-S. Sznitman. Topics in propagation of chaos. In P.-L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XIX - 1989*, volume 1464 of *Lecture Notes in Math.*, pp. 165–251. Springer, Berlin, 1991.

## APPENDIX

### A. Alternative Proof of Lemma 1

We show the existence and uniqueness of the equilibrium point. Differentiating the right-hand side of (11) with respect to  $\bar{q}$ , we can see that the following equation determines the sign of the derivative.

$$\begin{aligned} \delta_K &:= \sum_{k=0}^K k\gamma^{k-1} \left( \sum_{j=0}^K \frac{\gamma^j}{q_j} \right) - \sum_{j=0}^K \gamma^j \left( \sum_{k=0}^K \frac{k\gamma^{k-1}}{q_k} \right) \\ &= \sum_{k=0}^K \sum_{j=0}^K \gamma^{k+j-1} \left( \frac{k}{q_j} - \frac{j}{q_k} \right). \end{aligned}$$

Consider a proper subsum,  $\delta_\kappa$ , which can be obtained by replacing  $K$  with  $\kappa \in \{1, \dots, K-1\}$ . Recall that  $q_0 \geq q_1$  by the assumption; then it is easy to see that  $\delta_1 \leq 0$  is true. Now suppose  $\delta_\kappa$  is zero or negative. We show  $\delta_{\kappa+1} \leq 0$  if  $\delta_\kappa \leq 0$ . Rearranging terms of  $\delta_{\kappa+1}$ , it is not difficult to obtain:

$$\delta_{\kappa+1} = \delta_\kappa + \left[ \sum_{i=0}^{\kappa} \gamma^{\kappa+i} (\kappa+1-i) \left( \frac{1}{q_i} - \frac{1}{q_{\kappa+1}} \right) \right]$$

where the second term on the right-hand side is zero or negative as  $q_i$  is nonincreasing for  $i \in \{0, \dots, K\}$ . As  $\delta(K)$  is zero or negative, we can conclude that the right-hand side of (11) is a nonincreasing function which is positive and converges to  $(K+1)/\sum_{k=0}^K q_k^{-1}$  at  $\bar{p} = \infty$ . This conclusion

taken together with the fact that the left-hand side of (11) is an identical function from  $[0, \infty)$  to  $[0, \infty)$  proves that there exists a unique equilibrium point  $\bar{q}$ .

### B. Proof of Lemma 3

First, we note from (33) that the right-hand sides of (29) and (30) are nonincreasing in  $\gamma^H$  and  $\gamma^C$ , respectively. The proof of this fact is almost identical to that of Lemma 1.

Assume that there are two solutions  $(\bar{q}^H, \bar{q}^L)$  and  $(\hat{q}^H, \hat{q}^L)$  of the fixed point equation (29)-(32) and  $\hat{q}^H \geq \bar{q}^H$ , without loss of generality. If we assume that  $\hat{q}^L \geq \bar{q}^L$ , it follows from (32) that  $\hat{\gamma}^C \geq \gamma^C$ . Because the right-hand side of (30) is nonincreasing in  $\gamma^C$ , we must have  $\hat{q}^L \leq \bar{q}^L$  and hence  $\hat{\gamma}^C \geq \gamma^C$ . Now we have shown by contradiction that  $\hat{q}^H \geq \bar{q}^H$  implies  $\hat{q}^L \leq \bar{q}^L$  and  $\hat{\gamma}^C \geq \gamma^C$ .

Moreover, we can rewrite (31) in the following form:

$$\begin{aligned} \gamma^H &= \frac{(1 - \gamma^R)^\Delta + \gamma^R \sum_{i=0}^{\Delta-1} (1 - \gamma^R)^i}{\left\{ \sum_{i=0}^{\Delta-1} (1 - \gamma^R)^i \right\} + \frac{(1 - \gamma^R)^\Delta}{\gamma^C}} \\ &= \frac{1}{\left\{ \sum_{i=0}^{\Delta-1} (1 - \gamma^R)^i \right\} + \frac{(1 - \gamma^R)^\Delta}{\gamma^C}} \end{aligned} \quad (38)$$

where the second equality can be easily verified. As  $\gamma^R = 1 - e^{-\bar{q}^H}$  is increasing in  $\bar{q}^H$ ,  $\bar{q}^H \geq \hat{q}^H$  implies  $\hat{\gamma}^R \geq \gamma^R$ . Combining these with the fact that (38) is increasing in  $\gamma^R$  and  $\gamma^C$ , we can establish that  $\hat{q}^H \geq \bar{q}^H$  implies  $\hat{\gamma}^H \geq \gamma^H$ . On the other hand, since the right-hand side of (29) is nonincreasing in  $\gamma^H$ , the inequality  $\hat{\gamma}^H \geq \gamma^H$  must imply  $\hat{q}^H \leq \bar{q}^H$ .

In conclusion, if we assume  $\hat{q}^H \geq \bar{q}^H$ , we have  $\hat{q}^H \leq \bar{q}^H$ , which implies that  $\hat{q}^H = \bar{q}^H$ . Then it automatically follows that  $\hat{q}^L = \bar{q}^L$ ,  $\hat{\gamma}^H = \gamma^H$ , and  $\hat{\gamma}^L = \gamma^L$ .

We have yet to establish the existence of the solution. We first note that the left-hand sides of (29) and (30) are identical functions of  $\bar{q}^H$  and  $\bar{q}^L$ , respectively, from  $[0, \infty)$  to  $[0, \infty)$ . Because (32) is increasing in  $\bar{q}^L$ , for each fixed  $\bar{q}^H$ , the right-hand side of (30) is a positive nonincreasing function of  $\bar{q}^L$  by the proof of Lemma 1. Likewise, as (31) is increasing in  $\bar{q}^H$  for each fixed  $\bar{q}^L$ , the right-hand side of (29) is a positive nonincreasing function of  $\bar{q}^H$  by the proof of Lemma 1. This completes the proof.

### C. Proof of Lemma 4

Multiplying both sides of (29) and (30) respectively by  $(1 - \gamma^H)$  and  $(1 - \gamma^C)$  yields the following equations:

$$\bar{q}^H (1 - \gamma^H) = \sigma^H \frac{\sum_{k=0}^{K^H} (\gamma^H)^k}{\sum_{k=0}^{K^H} \frac{(\gamma^H)^k}{q_k^H}} \cdot (1 - \gamma^H), \quad (39)$$

$$\bar{q}^L (1 - \gamma^C) = \sigma^L \frac{\sum_{k=0}^{K^L} (\gamma^C)^k}{\sum_{k=0}^{K^L} \frac{(\gamma^C)^k}{q_k^L}} \cdot (1 - \gamma^C). \quad (40)$$

The proof is similar to that of Lemma 3 except that:

- (i) We use the fixed point equation (39), (40), (31) and (32).
- (ii) We note from Lemma 2 that the right-hand sides of (39) and (40) are decreasing respectively in  $\bar{q}^H$  and  $\bar{q}^L$ , and less than

or equal to the left-hand sides of (39) and (40) respectively at  $\bar{q}^H = 1$  and  $\bar{q}^L = 1$ .

To complete the proof, it is sufficient to show that the left-hand sides of (39) and (40) are increasing respectively in  $\bar{q}^H$  and  $\bar{q}^L$ . It follows from the proof of Lemma 2 that (35) implies  $\bar{q}^H \leq 1$  and  $\bar{q}^L \leq 1$ . It is also obvious from the form of  $\bar{q}^L (1 - \gamma^C) = \bar{q}^L e^{-\bar{q}^H - \bar{q}^L}$  that the left-hand side of (40) is increasing in  $\bar{q}^L \in [0, 1]$ .

To sum up again, it is now enough to show that the left-hand side of (39) is increasing in  $\bar{q}^H \in [0, 1]$ . To establish this, we rewrite (31) in a compact form

$$\gamma^H = 1 - \left\{ \frac{e^{-\bar{q}^H} h(\bar{q}^H, \bar{q}^L)}{h(\bar{q}^H, \bar{q}^L) + 1} + \frac{e^{-\bar{q}^H - \bar{q}^L}}{h(\bar{q}^H, \bar{q}^L) + 1} \right\}$$

where  $h(\bar{q}^H, \bar{q}^L) := \left( e^{\bar{q}^H \Delta} - 1 \right) \cdot \frac{1 - e^{-\bar{q}^H - \bar{q}^L}}{1 - e^{-\bar{q}^H}}$ . Differentiating  $\bar{q}^H (1 - \gamma^H)$  with respect to  $\bar{q}^H$  yields

$$(1 - \bar{q}^H) \frac{e^{-\bar{q}^H} h + e^{-\bar{q}^H - \bar{q}^L}}{h + 1} + \bar{q}^H \frac{e^{-\bar{q}^H} - e^{-\bar{q}^H - \bar{q}^L}}{(h + 1)^2} \cdot \frac{dh}{d\bar{q}^H}$$

where  $h$  is a shorthand notation for  $h(\bar{q}^H, \bar{q}^L)$ . The first term of the above equation is positive for  $\bar{q}^H \in (0, 1)$ . The sign of the second term is determined by  $\frac{dh}{d\bar{q}^H}$  which is nonnegative because  $h$  can be rearranged as

$$h(\bar{q}^H, \bar{q}^L) = \sum_{i=1}^{\Delta} e^{\bar{q}^H i} \cdot \left( 1 - e^{-\bar{q}^H - \bar{q}^L} \right)$$

which is nondecreasing in  $\bar{q}^H$ . This completes the proof.

**Jeong-woo Cho** received his B.S., M.S., and Ph.D. degrees in Electrical Engineering and Computer Science from KAIST, Daejeon, South Korea, in 2000, 2002, and 2005, respectively. From September 2005 to July 2007, he was with the Telecommunication R&D Center, Samsung Electronics, South Korea, as a Senior Engineer. From August 2007 to August 2010, he held postdoc positions in the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, and at the Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. He is now an assistant professor in the School of Information and Communication Technology at KTH Royal Institute of Technology, Stockholm, Sweden. His current research interests include performance evaluation in various networks such as peer-to-peer network, wireless local area network, and delay-tolerant network.

**Jean-Yves Le Boudec** is full professor at EPFL and fellow of the IEEE. He graduated from École Normale Supérieure de Saint-Cloud, Paris, where he obtained the Agrégation in Mathematics in 1980 (rank 4) and received his doctorate in 1984 from the University of Rennes, France. From 1984 to 1987 he was with INSA/IRISA, Rennes. In 1987 he joined Bell Northern Research, Ottawa, Canada, as a member of scientific staff in the Network and Product Traffic Design Department. In 1988, he joined the IBM Zurich Research Laboratory where he was manager of the Customer Premises Network Department. In 1994 he joined EPFL as associate professor.

His interests are in the performance and architecture of communication systems. In 1984, he developed analytical models of multiprocessor, multiple bus computers. In 1990 he invented the concept called “MAC emulation” which later became the ATM forum LAN emulation project, and developed the first ATM control point based on OSPF. He also launched public domain software for the interworking of ATM and TCP/IP under Linux. He proposed in 1998 the first solution to the failure propagation that arises from common infrastructures in the Internet. He contributed to network calculus, a recent set of developments that forms a foundation to many traffic control concepts in the Internet.

He earned the Infocom 2005 Best Paper award, with Milan Vojnovic, for elucidating the perfect simulation and stationarity of mobility models, the 2008 IEEE Communications Society William R. Bennett Prize in the Field of Communications Networking, with Bozidar Radunovic, for the analysis of max-min fairness and the 2009 ACM Sigmetrics Best Paper Award, with Augustin Chaintreau and Nikodin Ristanovic, for the mean field analysis of the age of information in gossiping protocols.

He is or has been on the program committee or editorial board of many conferences and journals, including Sigcomm, Sigmetrics, Infocom, Performance Evaluation and ACM/IEEE TRANSACTIONS ON NETWORKING. He co-authored the book *Network Calculus* (2001) with Patrick Thiran and is the author of the book *Performance Evaluation of Computer and Communication Systems* (2010).

**Yuming Jiang** received his BSc from Peking University, China, in 1988, MEng from Beijing Institute of Technology, China, in 1991, and PhD from National University of Singapore, Singapore, in 2001. He worked with Motorola from 1996 to 1997. From 2001 to 2003, he was a Member of Technical Staff and Research Scientist with the Institute for Infocomm Research, Singapore. From 2003 to 2004, he was an Adjunct Assistant Professor with the Electrical and Computer Engineering Department, National University of Singapore. From 2004 to 2005, he was with the Centre for Quantifiable Quality of Service in Communication Systems (Q2S), Norwegian University of Science and Technology (NTNU), Norway, supported in part by the Fellowship Programme of European Research Consortium for Informatics and Mathematics (ERCIM). Since 2005, he has been with the Department of Telematics, NTNU, as a Professor. He visited Northwestern University, USA from 2009 to 2010.

He was Co-Chair of IEEE Globecom2005 - General Conference Symposium, TPC Co-Chair of 67th IEEE Vehicular Technology Conference (VTC) 2008, and General/TPC Co-Chair of International Symposium on Wireless Communication Systems (ISWCS) 2007-2010. He is first author of the book “*Stochastic Network Calculus*”. His research interests are the provision, analysis and management of quality of service guarantees in communication networks. In the area of network calculus, his focus has been on developing models and investigating their basic properties for stochastic network calculus (snetcal), and recently also on applying snetcal to performance analysis of wireless networks.