On the Validity of the Decoupling Assumption for Analyzing the 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. A necessary condition for the validity of this approach is the existence and uniqueness of a solution to a fixed point equation. However, it was also recently pointed out that this condition is not sufficient; in contrast, a necessary and sufficient condition is a global stability property of the associated ordinary differential equation. Such a property was established only for a specific case, namely for a homogeneous system (all nodes have the same parameters) and when the number of backoff stages is either 1 or infinite and with other restrictive conditions. In this paper, we give a simple condition that establishes the validity of the decoupling assumption for the homogeneous case. We also discuss the heterogeneous and the differentiated service cases and show that the uniqueness condition is not sufficient; we exhibit one case where the fixed point equation has a unique solution but the decoupling assumption is not valid.**

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

I. Introduction

Living up to the growing and insatiable hunger for higher wireless throughput of users, IEEE 802.11n was ratified and released recently in October 2009, to which a lot of enterprises are reported to migrate. 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, *i.e.*, $N \to \infty$.

Most existing work on performance evaluation of the 802.11 MAC protocol [2], [9], [10], [15] relies on the "decoupling assumption" which was first adopted in the seminal work of 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, which in turn amounts to the assumption that the backoff processes are *independent*[1]. 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 [10] even for small number of backoff stages, *i.e.*, 3 and 4, unless the population is very small. A similar point was stressed by P. R. Kumar in an interview with Science Watch Newsletter [11]:

"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) [10], also called Bianchi's formula. Kumar *et al.* [10] revisited the FPE and axiomized 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 [10], already adopted in the field [15], [12], is that the full interference model, also called the single-cell model [10] and is the main focus of our work, leads to the *backoff synchrony property* [14] which implies the backoff process can be completely separated and analyzed solely through the FPE technique.

It is however pointed out by the work of Benaïm and Le Boudec [1, Section 8.2], using a mean field convergence method, that the uniqueness of a solution to the FPE does not necessarily lead to the validity of the decoupling assumption in the stationary regime. That is, the decoupling assumption along with its FPE has not been theoretically validated yet, though regarded as a plausible assumption especially when the number of nodes is large ($N \gg 1$). The main purpose of this paper is to provide an answer to the following challenging question with appropriate mathematical formalism.

Q:*"Under which conditions is the FPE valid?"*

That is, we refound the FPE technique on mean field theories [6], [1] which render down the validity of the FPE into the stability of a nonlinear *ordinary differential equation* (ODE), to which a scaled version of the original Markov chain is shown to converge when $N \to \infty$. A comprehensive summary of the literature and the outstanding questions raised therein has been recently made by Duffy [8].

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 and Y. Jiang are with the Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway (email: {jeongwoo,jiang}@q2s.ntnu.no).

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

[1]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'.

In line with the preceding argument, Bordenave *et al.* in [6, Theorem 5.4] studied the **homogeneous** case (all nodes have same per-stage backoff probabilities) for the case where the number of backoff stages is infinite. They found the following sufficient condition for global stability of the ODE, hence for validity of the decoupling assumption:

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

where q_k is the *re-scaled*[2] attempt probability (to be defined in Section II-B) 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.* [17] obtained a result for $K = 1$ and mentioned the difficulty to go beyond.

Before offering our answer, we discover an intriguing fact that not only (i) the MONOtonicity ((**MONO**) in Sec. II) but also (ii) the Mild INTensity of per-stage attempt probability ((**MINT**) in Sec. II) implies the uniqueness of a solution to the FPE, which is naturally a necessary condition for stability. Moreover, the latter (**MINT**) is proven to guarantee the global stability of the ODE. The finding greatly simplifies the story because simply that the attempt probability is *upper-bounded* by the reciprocal of the population suffices for each node being completely decoupled from the rest, namely $q_k \leq 1$ for all k. Moreover, it is shown that, 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 two mechanisms to differentiate the per-class settings of (i) the contention window (CW) and (ii) the idle time after each transmission. Since the former, CW differentiation, necessarily implies there are two or more classes, we call the corresponding system **heterogeneous**. The latter, 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 [10, Section VI].

[2]This word, used as an antonym of 'scale' in this work, means 'to increase the size of something'.

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 constitutes a separatrix which divides the initial condition space into two regions. We also consider a heterogenous 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.

This paper discovers a fundamental condition for perfect decoupling of wireless nodes under the asymptotic regime which is essential for the design of 802.11 in a predictable manner and founds an extended fixed point equation of greater generality on mean field theory, which constitutes the contributions of the paper. We also stress that the stability condition established in this work for the first time has been tantalizing other researchers as well, *e.g.*, [17, 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 proves a global stability condition of the ODE, which is in turn shown to be capable of optimizing the throughput. In Section III, we elaborate on another complexities arising from EDCA and derive its corresponding ODE model. Some counterexamples in Section IV illustrate the utility of the ODE models. Concluding remarks and a challenging outstanding problem are given in Section V.

II. Mean Field Technique Revisited

The backoff process in 802.11 is governed by a few rules if the duration of per-stage backoff is taken to be exponential: (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. Each node following the randomized access procedure of 802.11 distributed coordination function (DCF) generates a *backoff value* after receiving the Short Inter-Frame Space (SIFS) if it has a packet to send. This backoff value is uniformly distributed over $\{0, 1, \cdots, 2b_0-1\}$ (or $\{1, 2, \cdots, 2b_0\}$) where $2b_0$ is the initial contention window.

Whenever the medium is idle for the duration of a Distributed Inter-Frame Space (DIFS), a node unfreezes (starts) its countdown procedure of the backoff and decrements the backoff by one per every time-slot. It freezes the countdown procedure as soon as the medium becomes busy. There exist $K+1$ backoff stages whose indices belong to the set $\mathbb{K} := \{0, 1, \cdots, K\}$ where we assume $K > 0$. If two or more wireless nodes finish their countdowns at the same time-slot, there occurs a collision between RTS (ready to send) packets if the CSMA/CA (carrier sense multiple access with collision avoidance) is implemented, otherwise two 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. 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 whose contention window is set to $2b_0$. In the IEEE 802.11b standard, $m = 2$, $K = 6$ (7 attempts per packet), and $2b_0 = 32$ are used.

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 [10], [15], [6], [1].

B. The Bianchi's Formula

In performance analysis of 802.11, Bianchi's formula and its many variants are probably the most known [2], [7], [9], [10], [12], [13], [15], [16]. Assuming that there are N nodes, the Bianchi's formula can be written compactly in a more general *fixed point equation* (FPE) form:

$$\bar{p} = \frac{\sum_{k=0}^{K} \gamma^k}{\sum_{k=0}^{K} \frac{\gamma^k}{p_k}}, \tag{1}$$

$$\gamma = 1 - \mathrm{e}^{-N\bar{p}} \tag{2}$$

where $\bar{p}$ and γ respectively designate the average attempt rate 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 γ are not affected by whether attempts in backoff stage K are successful or not.

The main 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 first-order approximation of $\bar{p}$ and γ. It is, perhaps, surprising that whether the Bianchi's formula is valid or not has never been completely agreed upon, despite the fact that the FPE has been a *de facto* principal tool for the analysis.

Exactly under which condition the FPE holds is recently being rediscovered with rigorous mathematical arguments [6], [1], called *mean field independence*. Although the particle interaction model proposed in [1] overcomes some limitations and broadens applicability of the mean field model proposed by Bordenave *et al.* [6], both of them have found that, as the number of particles goes to infinity, *i.e.*, $N \to \infty$, 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.

C. Validation of Decoupling Assumption: ODE model

Let us dive into the details of the ODE model. Both [1] and [6] rely upon the concept of intensity scaling for mathematical tractability. The intensity scaling means that the intensity, defined as the number of state (backoff stage) transitions per node per time-slot, is vanishing [6], *i.e.*, converges to 0 as $N \to \infty$ [1]. This vanishing intensity precludes aggregate probabilities of nodes from being *saturated* (becoming 0 or 1) and makes formulation of the mean field differential equations possible. Specifically, in this work, 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 N, such that each node in backoff stage k attempts transmission with probability q_k/N.
- **Re-scaling** is to *accelerate* the evolution of time-slots by N, such that a variable at t before this operation is translated into another variable at t/N. For example, see $\boldsymbol{\Phi}(t)$ and $\boldsymbol{\phi}(t/N)$ in the below.

Then the scaled version of the Markov chain converges to an ODE system as $N \to \infty$. 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 [6]. The limit variables which we obtain by applying the re-scaling and the limit operation $N \to \infty$ are dubbed ***mean field limits (MFL)*** in this paper.

Denoting by $X_n(t)$ the backoff stage of node n at time-slot t, the occupancy measure (or empirical measure) of backoff stage k at time-slot t is defined as

$$\Phi_k(t) := \frac{1}{N} \sum_{n=1}^{N} 1_{\{X_n(t)=k\}} \tag{3}$$

where $1_{\{\cdot\}}$ is the indicator function. It is shown in [6] that $\boldsymbol{\Phi}(Nt)$ converges in probability to $\boldsymbol{\phi}(t)$ which is the solution of the ODE:

$$\frac{\mathrm{d}\phi_0}{\mathrm{d}t}(t) = \bar{q}(t)\,(1-\gamma(t)) - q_0\phi_0(t) + \underbrace{q_K\phi_K(t)\gamma(t)}_{\text{inflow from } K} \tag{4}$$

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

$$\frac{\mathrm{d}\phi_k}{\mathrm{d}t}(t) = q_{k-1}\phi_{k-1}(t)\gamma(t) - q_k\phi_k(t), \tag{ODE}$$

which is the drift equation for $k \in \{1, \cdots, K\}$. Here $\bar{q}(t) := \sum_{k=0}^{K} q_k \phi_k(t)$ is the MFL of the average attempt rate and $\gamma(t)$ is that of the collision probability to be defined very soon. It is important to note that the above system is *degenerate*[3], because we also have a *manifold* relation $\phi_0(t) \equiv 1 - \sum_{k=1}^{K} \phi_k(t)$, which can be plugged into (ODE) to eliminate (4), whereupon we only need to consider the K-dimensional system (ODE). We will use the reduced version (ODE) throughout this work to simplify the exposition. This system (ODE) will be called **homogeneous** because all nodes adopt the same parameter set q_k and K.

The drift equation (ODE) can be intuitively understood. For example, the first term and second term on the right-hand side in (ODE) are respectively the inflow caused by collisions in the $(k-1)$th backoff stage and the outflow caused by any attempts in the kth backoff stage. Note that the underbraced term in (4) was not considered in [6], and exists only in networks with finite backoff stages.

We recapitulate here the derivation of $\gamma(t)$ in [6]. Recall that the attempt probability of every node in backoff stage k is taken to be q_k/N. The collision probability of a tagged node n in backoff stage $X_n(t)$ necessarily takes the form

$$\Gamma(t,n) := 1 - \left(1 - \frac{q_{X_n(t)}}{N}\right)^{-1} \prod_{k=0}^{K} \left(1 - \frac{q_k}{N}\right)^{N \cdot \Phi_k(t)} \tag{5}$$

which is the probability that at least one other node than n attempts transmission at time-slot t. Here we can see that the term $q_{X_n(t)}/N$ vanishes as $N \to \infty$. Just as $\mathbf{\Phi}(Nt)$ converges to $\boldsymbol{\phi}(t)$ as $N \to \infty$, the re-scaled version $\Gamma(Nt, n)$ converges to its MFL as $N \to \infty$. Thus we similarly define the MFL of collision probability as follows:

$$\gamma(t) := \lim_{N \to \infty} \Gamma(Nt, n).$$

Then it follows from the definition of exponential function $\lim_{N\to\infty}(1 - x/N)^N = \exp(-x)$ that we have

$$\gamma(t) = 1 - \mathrm{e}^{-\sum_{k=0}^{K} q_k \phi_k(t)} = 1 - \mathrm{e}^{-\bar{q}(t)} \tag{6}$$

which is the final form we use in (ODE). Remark that $\Gamma(t, n)$ depends on the backoff stage $X_n(t)$ where node n is, whereas its MFL $\gamma(t)$ is common to all nodes.

Equating the right-hand sides of (4) and (ODE) 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:

$$\bar{q} = \frac{\sum_{k=0}^{K} \gamma^k}{\sum_{k=0}^{K} \frac{\gamma^k}{q_k}}, \tag{FPE1}$$

$$\gamma = 1 - \mathrm{e}^{-\bar{q}}. \tag{FPE2}$$

The theoretical limit of mean field analysis represented by (ODE) needs to be clearly understood. The nonlinear ODE model only implies that *any* node will be in backoff stage k with the common probability $\phi_k(t/N)$ under the asymptotic regime, $N \to \infty$. The component ratio $\boldsymbol{\phi}(t)$, in general a time-varying solution of (ODE), is not guaranteed to be constant. Bordenave *et al.* [6, Theorem 5.4] studies its global stability of the asymptotic case when $K = \infty$, 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 [17, Lemma 3].

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:

$$\boxed{q_k \text{ is nonincreasing in } k.} \tag{\textbf{MONO}}$$

$$\boxed{\text{(FPE1)-(FPE2) has a unique solution.}} \tag{\textbf{UNIQ}}$$

The following lemma holds as long as the right-hand side of (FPE2) is increasing in $\bar{q}$. That is, the lemma does not fully exploits the exponential form of (FPE2). It is remarkable that Lemma 1 was originally established by Kumar *et al.* [10, 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:

$$\bar{q}(t) \leq 1, \quad \forall t \geq 0. \tag{7}$$

As we are interested in *global* stability, we need to show that the solutions of (ODE) 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 (7) holds for any initial condition $\boldsymbol{\phi}(0)$ *if and only if*

$$\boxed{q_k \leq 1, \ \forall k.} \tag{\textbf{MINT}}$$

Interestingly, the above upper bound on the per-stage attempt probabilities 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 $q_{\max} := \max_{k \in \mathbb{K}} q_k$, it is clear that (**MINT**) is equivalent to $q_{\max} \leq 1$. First, we have

$$\bar{q} = \frac{\sum_{k\in\mathbb{K}} \gamma^k}{\sum_{k\in\mathbb{K}} \frac{\gamma^k}{q_k}} \leq \frac{\sum_{k\in\mathbb{K}} \gamma^k}{\sum_{k\in\mathbb{K}} \frac{\gamma^k}{q_{\max}}} = q_{\max} \leq 1. \tag{8}$$

[3] A degenerate system has a *singular* Jacobian matrix which means that its linearization cannot determine the local stability of the system.

Multiplying the both sides of (FPE1) by $\mathrm{e}^{-\bar{q}}$ yields:

$$\bar{q}\mathrm{e}^{-\bar{q}} = \frac{\sum_{k\in\mathbb{K}} \gamma^k}{\sum_{k\in\mathbb{K}} \frac{\gamma^k}{q_k}} \cdot \mathrm{e}^{-\bar{q}} = \frac{1}{\sum_{k\in\mathbb{K}} \frac{\gamma^k}{q_k}} \cdot \frac{\sum_{k\in\mathbb{K}} \gamma^k}{\sum_{k=0}^{\infty} \gamma^k}. \tag{9}$$

The second factor of (9) can be rearranged as

$$\sum_{k\in\mathbb{K}} \gamma^k(1-\gamma) = (1+\cdots+\gamma^K) - (\gamma+\cdots+\gamma^{K+1}) = 1-\gamma^{K+1} = 1-(1-\mathrm{e}^{-\bar{q}})^{K+1}$$

which is a decreasing function of $\bar{q}$. As the first factor of (9) is also a decreasing function of $\bar{q}$, (9) is decreasing in $\bar{q}$. On the other hand, $\bar{q}\mathrm{e}^{-\bar{q}}$ is increasing in $\bar{q} \in [0,1]$ and the range of $\bar{q}\mathrm{e}^{-\bar{q}}$ is $[0, \mathrm{e}^{-1}]$. Since (9) decreases from q_0 at $\bar{q}=0$ to

$$\frac{\sum_{k\in\mathbb{K}} (1-\mathrm{e}^{-1})^k}{\sum_{k\in\mathbb{K}} \frac{(1-\mathrm{e}^{-1})^k}{q_k}} \cdot \mathrm{e}^{-1}$$

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, (FPE1) and (FPE2) have a unique solution. ■

Unlike Lemma 1, the forms of both (FPE1) and (FPE2) 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.

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

Theorem 1 (Stability Condition)
(**MINT**) implies the global stability of (ODE).

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

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

Since $\boldsymbol{\phi}(t)$ is a probability measure on a finite sample space $\mathbb{K}$, $\boldsymbol{\phi}(t)$ is tight [4]. Appealing to this, we can pick a convergent *sub*sequence $\{t_i\}$ such that $\lim_{t_i\to\infty} \phi_k(t_i) = \phi_k(\infty)$ 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 $\lim_{t\to\infty} \nu(t) = \bar{q}^l$. Consider the *degenerate* version of (ODE) which has one additional equation with respect to $\frac{\mathrm{d}\phi_0}{\mathrm{d}t}(t)$. By plugging the substitution $\bar{q}(t) \Rightarrow \nu(t)$ into this, we get the following modified ODE:

$$\frac{\mathrm{d}\varphi_0}{\mathrm{d}t}(t) = \nu(t)\mathrm{e}^{-\nu(t)} - q_0\varphi_0(t) + q_K\varphi_K(t)(1-\mathrm{e}^{-\nu(t)}),$$
$$\frac{\mathrm{d}\varphi_k}{\mathrm{d}t}(t) = q_{k-1}\varphi_{k-1}(t)(1-\mathrm{e}^{-\nu(t)}) - q_k\varphi_k(t).$$

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

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

where we used $\gamma^l := \lim_{t\to\infty}(1-\mathrm{e}^{-\nu(t)})$ 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 $t\to\infty$. Thus $\varphi(\infty)$ should satisfy

$$q_0\varphi_0(\infty) = \bar{q}^l\mathrm{e}^{-\bar{q}^l} + q_K\varphi_K(\infty)\left(1-\mathrm{e}^{-\bar{q}^l}\right), \tag{10}$$
$$q_k\varphi_k(\infty) = q_{k-1}\varphi_{k-1}(\infty)\left(1-\mathrm{e}^{-\bar{q}^l}\right), \tag{11}$$

for $k \in \mathbb{K}\backslash\{0\}$ because $\lim_{t\to\infty}\nu(t) = \bar{q}^l$. Plugging (11) into (10) yields

$$q_k\varphi_k(\infty) = \bar{q}^l(1-\mathrm{e}^{-\bar{q}^l})^k \Big/ \sum_{j\in\mathbb{K}} (1-\mathrm{e}^{-\bar{q}^l})^j \ . \tag{12}$$

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

$$\varphi_0(t) = \mathrm{e}^{-q_0 t}\phi_0(0) + \int_0^t \mathrm{e}^{q_0(s-t)}\left\{\nu(s)\mathrm{e}^{-\nu(s)} + q_K\varphi_K(s)\left(1-\mathrm{e}^{-\nu(s)}\right)\right\}\mathrm{d}s, \tag{13}$$

$$\varphi_k(t) = \mathrm{e}^{-q_k t}\phi_k(0) + \int_0^t \mathrm{e}^{q_k(s-t)} q_{k-1}\varphi_{k-1}(s)\left(1-\mathrm{e}^{-\nu(s)}\right)\mathrm{d}s, \tag{14}$$

where $k \in \mathbb{K}\backslash\{0\}$. First we have $\nu(t) \leq 1$ from the assumption (**MINT**). Since $1-\mathrm{e}^{-x}$ and $x\mathrm{e}^{-x}$ 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 (14) into (13) K times that $\varphi_0(t) \leq \phi_0(t)$ and hence $\varphi_k(t) \leq \phi_k(t)$, $\forall t \geq 0$ and $\forall k \in \mathbb{K}$. That is, $\phi_k(t)$ is *lower-bounded* by $\varphi_k(t)$.

From (12) and the definition of the subsequence $\{t_i\}$, we have the following relation:

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

where we recall $\phi_k(\infty)$ was defined as $\lim_{t_i\to\infty}\phi_k(t_i) = \phi_k(\infty)$. This result taken together with $\varphi_k(t) \leq \phi_k(t)$ proves $\varphi_k(\infty) = \phi_k(\infty)$, $\forall k \in \mathbb{K}$, and therefore, $\sum_{k\in\mathbb{K}}\varphi_k(\infty) = 1$. Then it necessarily follows that $\bar{q}^l$ should satisfy (FPE1) and (FPE2) 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 $\limsup_{t\to\infty}\bar{q}(t) = \bar{q}^u$, $\lim_{t_i\to\infty}\bar{q}(t_i) = \bar{q}^u$ and $\lim_{t_i\to\infty}\phi_k(t_i) = \phi_k(\infty)$. This will show $\lim_{t\to\infty}\bar{q}(t) = \bar{q}$. That is, there is only one limit point for $\bar{q}(t)$.

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

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

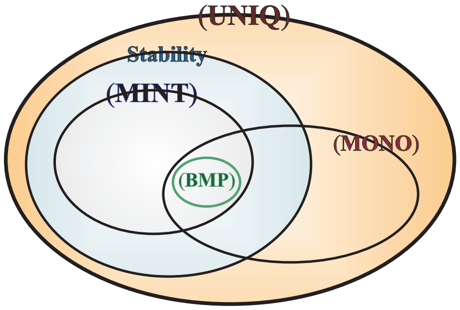


Fig. 1. Logical relations between conditions.

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

Remark 1 **R1.1** [Answer to **Q1**] This result **justifies** the FPE approach used in [2], [7], [9], [10], [12], [13], [15], [16] through rigorous arguments under the weakest assumption. That is, the decoupling assumption is validated, as long as the per-stage attempt probability is mild, *i.e.*, (**MINT**).

R1.2 Another result of [6, Theorem 5.4] for the case $K = \infty$ requires an additional *monotonicity* condition along with a condition on attempt probability at $k = 0$ as shown in the conditions (**BMP**) in Section I to prevent wireless node to escape to infinite backoff stage. These conditions, designated also by (**BMP**) in Fig. 1, correspond to a proper subset of the intersection of (**MINT**) and (**MONO**). As compared with [6, Theorem 5.4], Theorem 1 is a much stronger yet more practical argument due to finite K.

R1.3 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 validity of the decoupling assumption. It is still *open* whether (**MONO**) implies the global stability or not.

R1.4 Informally, the theorem is demonstrated by the fact that the solution $\boldsymbol{\phi}(t)$ 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 $t \to \infty$ at the same time as its solution $\varphi(t)$ lower-bounds $\boldsymbol{\phi}(t)$ such that $\boldsymbol{\phi}(t)$ is *squeezed* into $\boldsymbol{\phi}$ as $t \to \infty$.

From another viewpoint, the theorem can be restated as follows: the ODE is globally stable if the collision probability $\gamma(t) \le 1 - \mathrm{e}^{-1} = 0.632$ for any initial condition.

It is naturally apparent that *weakening* the activity of nodes is *indispensable* to decoupling the interaction between nodes, hence a precondition to decoupling. Viewing the intensity scaling as a weakening method, we note Theorem 1 requires even stronger weakening as follows:

> To accomplish decoupling, it is enough to **further** weaken the node activity such that $q_k \le 1$, along with the intensity scaling.

D. Achievable Throughput

Let L and L_c denote the durations of a successful packet transmission and a collision, expressed in terms of time-slot. Also the fixed overhead for each successful transmission is denoted by L_o. Assuming that $\bar{q}(t) \to \bar{q}$ as N tends to infinity, we can define the achievable throughput or alternatively the MFL of the aggregate throughput, as in [5, Section 5]:

$$\Omega(\bar{q}) := \frac{\mathsf{P}_1(\bar{q}) \cdot L}{\mathsf{P}_1(\bar{q}) \cdot (L + L_o) + \mathsf{P}_0(\bar{q}) + \mathsf{P}_c(\bar{q}) \cdot L_c} \tag{15}$$

where $\mathsf{P}_1(\bar{q}) := \bar{q}\mathrm{e}^{-\bar{q}}$, $\mathsf{P}_0(\bar{q}) := \mathrm{e}^{-\bar{q}}$, and $\mathsf{P}_c(\bar{q}) := 1 - \mathsf{P}_1(\bar{q}) - \mathsf{P}_0(\bar{q})$ 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 (5) and thus omitted.

Since (15) holds on the condition that (ODE) is *globally stable* such that $\lim_{t\to\infty} \bar{q}(t) = \bar{q}$, we can use (15) so long as (**MINT**) holds. Then the result of Theorem 1 poses another question:

"Does (**MINT**) *suffice for maximizing* (15)*?"*

Dividing the denominator of (15) by its nominator, we can see that maximizing (15) is equivalent to minimizing

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

Differentiating this expression shows that the global maximum of (15) is at the solution of the following equation:

$$\frac{1}{L_c} - 1 = (\bar{q} - 1)\mathrm{e}^{\bar{q}} \tag{16}$$

whose right-hand side is monotonically increasing in $\bar{q}$ over the domain $(0, \infty)$. 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 (16), which is from now denoted by $\bar{q} = \bar{q}^*$. If $L_c = 1$, the solution is $\bar{q}^* = 1$. We can easily show the following corollary.

Corollary 1 Suppose $L_c \ge 1$. $\bar{q}^* \le 1$ maximizes (15).

In the meantime, for $q_k = q_0/m^k$ and $\bar{q} = \bar{q}^*$, plugging (FPE2) into (FPE1) yields:

$$\frac{\bar{q}^*}{q_0} = \frac{\sum_{k=0}^{K} \left(1 - \mathrm{e}^{-\bar{q}^*}\right)^k}{\sum_{k=0}^{K} \left(1 - \mathrm{e}^{-\bar{q}^*}\right)^k m^k}. \tag{17}$$

The right-hand side of (17) is decreasing in $m \in (0, \infty)$. For given optimal solution $\bar{q}^*$, one can use (17) to find q_0 and m which satisfy (**MINT**) and maximize (15) at the same time. For instance, in order to obtain nonincreasing q_k, one can simply set $q_0 = 1$ and compute m from (17) where $m \ge 1$ is warranted because the left-hand side of (17) is no greater than 1 and the right-hand side of (17) decreases from 1 at $m = 1$ to 0 at $m = \infty$. Denoting by $m = m^*$ the solution to (17), we have another corollary.

Corollary 2 Suppose $L_c \ge 1$. For any $q_0 \in [\bar{q}^*, 1]$, let $q_k = q_0/(m^*)^k$. q_k is nonincreasing and satisfies (**MINT**).

Note that $L_c \ge 1$ in all versions of IEEE 802.11 MAC, regardless of the usage of the RTS/CTS mechanism. In summary, (**MINT**) is **enough** to maximize (15).

III. 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 freezed (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 $k \in \{0, \cdots, K\}$. We extend this feature by allowing per-class setting of K and q_k for *any* $k \in \{0, \cdots, K\}$ 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* preemptive prioritization to a certain class by holding back other classes from attempting transmissions for a few time-slots. This preemption 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}, $\Phi_k^{\mathrm{H}}(t)$ and $\Phi_k^{\mathrm{L}}(t)$.

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-B where the stage evolution of nodes depends only on their own stage density, *i.e.*, the occupancy measure $\mathbf{\Phi}(\cdot)$, 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 $\mathbf{\Phi}^{\mathrm{H}}(t)$ and $\mathbf{\Phi}^{\mathrm{L}}(t)$. 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 evolutions 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.* [17, Section IV] in a similar context also reckoned this difficulty though they have not solved it.

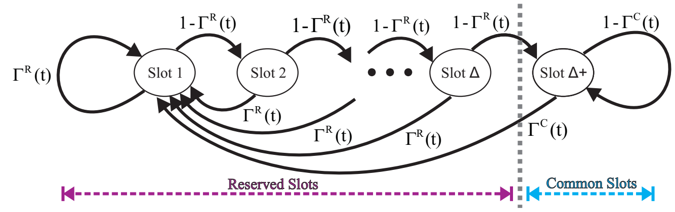


Fig. 2. Evolution of slot type follows a nonhomogeneous Markov chain.

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.*, $\Phi_k^{\mathrm{H}}(t)$ and $\Phi_k^{\mathrm{L}}(t)$ 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

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

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

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

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

$$\Gamma^{\mathrm{R}}(t) := 1 - \prod_{k=0}^{K^{\mathrm{H}}} \left(1 - \frac{q_k^{\mathrm{H}}}{N}\right)^{N \cdot \Phi_k^{\mathrm{H}}(t)},$$

$$\Gamma^{\mathrm{C}}(t) := 1 - \prod_{k=0}^{K^{\mathrm{H}}} \left(1 - \frac{q_k^{\mathrm{H}}}{N}\right)^{N \cdot \Phi_k^{\mathrm{H}}(t)} \prod_{k=0}^{K^{\mathrm{L}}} \left(1 - \frac{q_k^{\mathrm{L}}}{N}\right)^{N \cdot \Phi_k^{\mathrm{L}}(t)}$$

where we intentionally use the letter Γ which is the same as the collision probability in (5) because in the mean field limit the additional term $q_{X_n(t)}/N$ in (5) vanishes.

From the viewpoint of an 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 per-class AIFS value, the node may attempt transmission with its per-stage probability $q^{\mathrm{H,L}}_{X_n(t)}/N$.

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 $\Gamma^{\mathrm{R,C}}(t)$ and the idle probabilities $1-\Gamma^{\mathrm{R,C}}(t)$ as well. Here in Fig. 2 reserved time-slots and common time-slots are respectively denoted by the notations 'Slot 1'–'Slot Δ' and 'Slot $\Delta+$'. Note that $\Delta+$ means that, after any Δ 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 $\mathbf{\Phi}^{\mathrm{H,L}}(t)$ are also changing over time-slots, hence $\Gamma^{\mathrm{R,C}}(t)$ is.

The simplification of the analysis bases upon the following intuitive observation:

O: *"As population grows, the stage distribution (density) varies much slower than the type of time-slots."*

which is essentially due to the intensity scaling. Formally speaking, the occupancy measures, $\Phi^{\mathrm{H}}_k(t)$ and $\Phi^{\mathrm{L}}_k(t)$, evolve at a rate of $\Theta(1/N)$ which ultimately vanishes as $N \to \infty$, whereas the probability that the slot-type changes for each time-slot does not vanish and is strictly positive. Therefore, we can analyze the evolution of slot type **as if** the occupancy measures were constant. Solving the balance equations as if the Markov chain were homogeneous yields the following stationary distributions for each slot type:

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

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

which satisfy $\pi^{\mathrm{R}}(t)+\pi^{\mathrm{C}}(t) \equiv 1$. Note that however, 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 $\Phi^{\mathrm{H}}_k(t)$ and $\Phi^{\mathrm{L}}_k(t)$ are denoted by $\phi^{\mathrm{H}}_k(t)$ and $\phi^{\mathrm{L}}_k(t)$ as in Section II-C. After manipulation akin to (6), we can show that the MFLs of collision probability for the different types of time-slots, R and C, take the forms

$$\gamma^{\mathrm{R}}(t) = 1-\mathrm{e}^{-\bar{q}^{\mathrm{H}}(t)} \quad \text{and} \quad \gamma^{\mathrm{C}}(t) = 1-\mathrm{e}^{-\bar{q}^{\mathrm{H}}(t)-\bar{q}^{\mathrm{L}}(t)}$$

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

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

Let $\boldsymbol{g}^{\mathrm{R}}(t)$ and $\boldsymbol{g}^{\mathrm{C}}(t)$ denote the drifts of $(\boldsymbol{\phi}^{\mathrm{H}}(t), \boldsymbol{\phi}^{\mathrm{L}}(t))$ under the conditions that all time-slots are of slot type R or C, respectively. By means of the informal observation in Section III-B, Benaïm and Le Boudec have formally established a few results [1], by appealing to which, we now derive the final extended ODE. To this aim, as in Section II-C, assume that we eliminate manifolds by using the substitutions $\phi^{\mathrm{H}}_0(t) \equiv \sigma^{\mathrm{H}} - \sum_{k=1}^{K^{\mathrm{H}}}\phi^{\mathrm{H}}_k(t)$ and $\phi^{\mathrm{L}}_0(t) \equiv \sigma^{\mathrm{L}} - \sum_{k=1}^{K^{\mathrm{L}}}\phi^{\mathrm{L}}_k(t)$ and hence we consider $(K^{\mathrm{H}}+K^{\mathrm{L}})$-dimensional ODE. From the fact that at a time-slot of slot-type R, only Class H users are allowed to attempt transmission, whereas at a time-slot of slot-type C, all users are allowed to do so, we have

$$\boldsymbol{g}^{\mathrm{R}}(t) = \left(\begin{array}{c} q^{\mathrm{H}}_{K^{\mathrm{H}}-1}\phi^{\mathrm{H}}_{K^{\mathrm{H}}-1}(t)\gamma^{\mathrm{R}}(t) - q^{\mathrm{H}}_{K^{\mathrm{H}}}\phi^{\mathrm{H}}_{K^{\mathrm{H}}}(t) \\ \vdots \\ q^{\mathrm{H}}_0\phi^{\mathrm{H}}_0(t)\gamma^{\mathrm{R}}(t) - q^{\mathrm{H}}_1\phi^{\mathrm{H}}_1(t) \\ \hline 0 \\ \vdots \\ 0 \end{array}\right),$$

$$\boldsymbol{g}^{\mathrm{C}}(t) = \left(\begin{array}{c} q^{\mathrm{H}}_{K^{\mathrm{H}}-1}\phi^{\mathrm{H}}_{K^{\mathrm{H}}-1}(t)\gamma^{\mathrm{C}}(t) - q^{\mathrm{H}}_{K^{\mathrm{H}}}\phi^{\mathrm{H}}_{K^{\mathrm{H}}}(t) \\ \vdots \\ q^{\mathrm{H}}_0\phi^{\mathrm{H}}_0(t)\gamma^{\mathrm{C}}(t) - q^{\mathrm{H}}_1\phi^{\mathrm{H}}_1(t) \\ \hline q^{\mathrm{L}}_{K^{\mathrm{L}}-1}\phi^{\mathrm{L}}_{K^{\mathrm{L}}-1}(t)\gamma^{\mathrm{C}}(t) - q^{\mathrm{L}}_{K^{\mathrm{L}}}\phi^{\mathrm{L}}_{K^{\mathrm{L}}}(t) \\ \vdots \\ q^{\mathrm{L}}_0\phi^{\mathrm{L}}_0(t)\gamma^{\mathrm{C}}(t) - q^{\mathrm{L}}_1\phi^{\mathrm{L}}_1(t) \end{array}\right).$$

Interestingly, we can apply [1, Theorems 1 & 2][4] to show that the resultant drift of $(\boldsymbol{\phi}^{\mathrm{H}}(t), \boldsymbol{\phi}^{\mathrm{L}}(t))$ becomes:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\begin{array}{c}\boldsymbol{\phi}^{\mathrm{H}}(t)\\ \boldsymbol{\phi}^{\mathrm{L}}(t)\end{array}\right] = \boldsymbol{g}^{\mathrm{R}}(t)\cdot\pi^{\mathrm{R}}(t) + \boldsymbol{g}^{\mathrm{C}}(t)\cdot\pi^{\mathrm{C}}(t) \tag{18}$$

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

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

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

This result has the implication that we can derive (18) as a **linear combination** of two kinds of MFLs $\boldsymbol{g}^{\mathrm{R,C}}(t)$ and $\pi^{\mathrm{R,C}}(t)$ where the latter $\pi^{\mathrm{R,C}}(t)$ can be defined as

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

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

Finally, after some manipulation of (18), we have the following enhanced ordinary differential equation:

$$\frac{\mathrm{d}\phi^{\mathrm{H}}_k}{\mathrm{d}t}(t) = q^{\mathrm{H}}_{k-1}\phi^{\mathrm{H}}_{k-1}(t)\gamma^{\mathrm{H}}(t) - q^{\mathrm{H}}_k\phi^{\mathrm{H}}_k(t), \tag{eODE1}$$

$$\frac{\mathrm{d}\phi^{\mathrm{L}}_k}{\mathrm{d}t}(t) = \pi^{\mathrm{C}}(t)\left\{q^{\mathrm{L}}_{k-1}\phi^{\mathrm{L}}_{k-1}(t)\gamma^{\mathrm{C}}(t) - q^{\mathrm{L}}_k\phi^{\mathrm{L}}_k(t)\right\}, \tag{eODE2}$$

[4] The corresponding assumptions can be easily checked.

where (eODE1) and (eODE2) respectively hold for $k \in \{1, \cdots, K^{\mathrm{H}}\}$ and $k \in \{1, \cdots, K^{\mathrm{L}}\}$. Here we use the following shorthand notation:

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

whose form is obvious from (18).

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

$$\bar{q}^{\mathrm{H}} = \sigma^{\mathrm{H}} \frac{\sum_{k=0}^{K^{\mathrm{H}}} \left(\gamma^{\mathrm{H}}\right)^k}{\sum_{k=0}^{K^{\mathrm{H}}} \frac{(\gamma^{\mathrm{H}})^k}{q_k^{\mathrm{H}}}}, \tag{eFPE1}$$

$$\bar{q}^{\mathrm{L}} = \sigma^{\mathrm{L}} \frac{\sum_{k=0}^{K^{\mathrm{L}}} \left(\gamma^{\mathrm{C}}\right)^k}{\sum_{k=0}^{K^{\mathrm{L}}} \frac{(\gamma^{\mathrm{C}})^k}{q_k^{\mathrm{L}}}}, \tag{eFPE2}$$

$$\gamma^{\mathrm{H}} = \pi^{\mathrm{R}} \left(1 - \mathrm{e}^{-\bar{q}^{\mathrm{H}}}\right) + \pi^{\mathrm{C}} \left(1 - \mathrm{e}^{-\bar{q}^{\mathrm{H}} - \bar{q}^{\mathrm{L}}}\right), \tag{eFPE3}$$

$$\gamma^{\mathrm{C}} = 1 - \mathrm{e}^{-\bar{q}^{\mathrm{H}} - \bar{q}^{\mathrm{L}}}. \tag{eFPE4}$$

Remark 2 **R2.1** It is remarkable that the extended ODE model laid out in (eODE1) and (eODE2) *encompasses* the homogeneous system in Section II, and the heterogeneous system in Section III as well, which has the two prioritization functionalities.

R2.2 For instance, if $\Delta = \infty$, we have $\pi^{\mathrm{R}}(t) = 1$ and the ODE model reduces to the homogeneous system (ODE). On the other hand, if $\Delta = 0$, we have $\pi^{\mathrm{C}}(t) = 1$ and the ODE model reduces to a purely heterogeneous system, implying that the AIFS differentiation is disabled.

R2.3 What is the most surprising is that the FPE (eFPE1)-(eFPE4) **coincides** with that proposed in [10, Section VI], which was derived rather intuitively.

In the following, we give the meanings of three conditions akin to those in Section II-C, by adopting which we present two lemmas.

$$\boxed{q_k^{\mathrm{H}} \text{ and } q_k^{\mathrm{L}} \text{ are nonincreasing in } k,} \tag{eMONO}$$

$$\boxed{\text{(eFPE1)-(eFPE4) has a unique solution,}} \tag{eUNIQ}$$

$$\boxed{q_k^{\mathrm{H}} \le 1 \text{ and } q_k^{\mathrm{L}} \le 1,\ \forall k.} \tag{eMINT}$$

Lemma 3 (Monotonicity Implies Uniqueness) (**eMONO**) implies (**eUNIQ**).

Lemma 4 (Mild Intensity Implies Uniqueness) (**eMINT**) implies (**eUNIQ**).

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 $\Delta \ge 0$ 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 content with having stated the problem precisely with its inherent technical difficulty.

IV. Selected Counterexamples

Before proceeding to selected examples, we must bridge the gap between ODE models and the backoff processes in 802.11. This gap emerged right on applying the *intensity scaling* in Section II-C, which requires per-stage backoff probability be q_k/N. Putting $p_k = q_k/N$, we note that substituting q_k by Np_k should yield a reasonable approximation if p_k is small. After removing the re-scaling from (ODE), we have

$$\frac{\mathrm{d}\phi_k}{\mathrm{d}t}(t) = p_{k-1}\phi_{k-1}(t)\gamma(t) - p_k\phi_k(t) \tag{ODE'}$$

where $\gamma(t) := 1 - \mathrm{e}^{-N\bar{p}(t)}$ and $\bar{p}(t) := \sum_{k=0}^{K} p_k\phi_k(t)$. It is no wonder that the equilibrium of (ODE') satisfies the Bianchi's formula, (1) and (2).

A. Example 1: Multistability

Consider the **homogeneous** system (ODE'). Plugging (1) into (2) yields:

$$f(\gamma) := = 1 - \exp\left(-N\frac{\sum_{k=0}^{K}\gamma^k}{\sum_{k=0}^{K}\frac{\gamma^k}{p_k}}\right) - \gamma = 0 \tag{19}$$

which is a function of only γ. 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

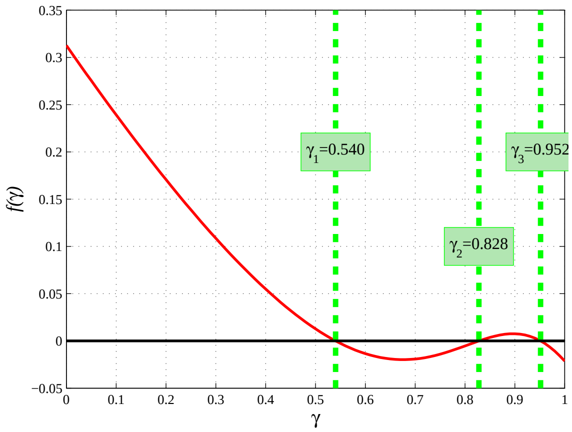
$$(p_0, p_1, \cdots, p_{12}) = \left(\frac{1}{3200}, \frac{1}{160}, \frac{m}{160}, \cdots, \frac{m^{11}}{160}\right)$$

where $m = 6/5 = 1.2$. The roots of (19) can be computed from Fig. 3(a) as

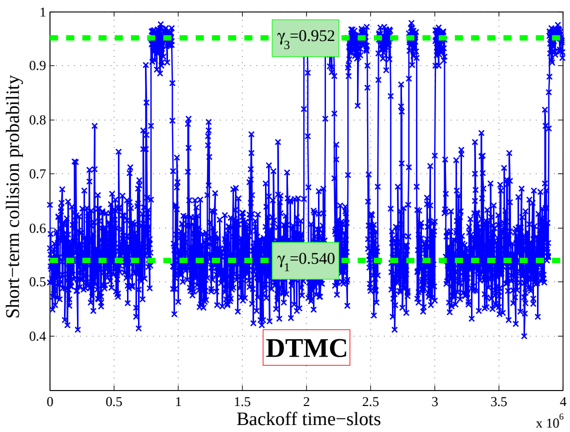
$$(\gamma_1, \gamma_2, \gamma_3) = (0.540, 0.828, 0.952).$$

We have simulated the corresponding Discrete Time Markov Chain (DTMC) to obtain short-term averages of collision probability $\gamma(t)$ and occupancy measure $\phi_k(t)$. The instantaneous collision probability for each 2000 backoff time-slots is shown in Fig. 3(b), which tends to concentrate 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 γ_1 nor γ_3.

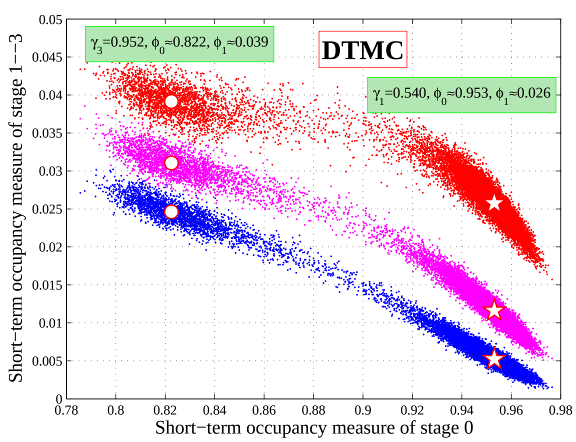
Recall that $\phi_k(t)$ denotes the fraction of nodes in backoff stage k. Fig. 3(c) shows the short-term average of the fraction of nodes in backoff stage $k \in \{1, \cdots, 3\}$ versus that in backoff stage 0. From the top to the bottom, the short-term occupancy measures of stage $1-3$ are shown in order, where the two kinds of markers, *i.e.*, circle ($\circ$) and star ($\star$), stand for the occupancy measures at two equilibriums, γ_3 and γ_1, which are computed from (ODE'). The *bistability* of this system is precisely predicted from either two modes of behavior of (ODE') or the eigenvalues of Jacobian matrices at the three equilibrium points.



(a) $f(\gamma)$ versus γ

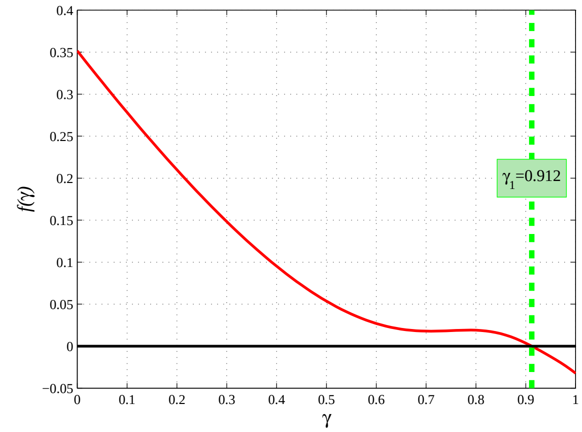


(b) Short-term average collision probability vs. backoff time-slots

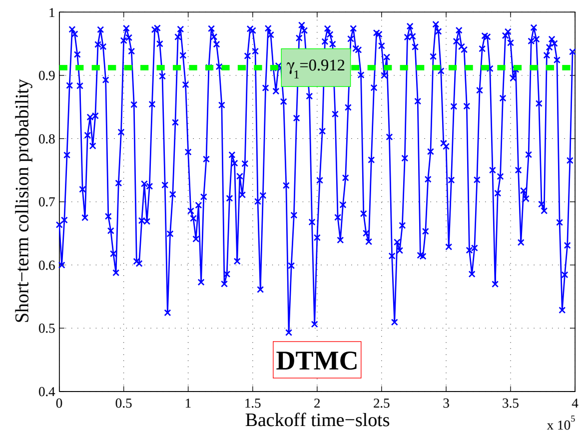


(c) Short-term average occupancy measure in backoff stages, $(\phi_0(t), \phi_1(t))$, $(\phi_0(t), \phi_2(t))$ and $(\phi_0(t), \phi_3(t))$. *Stars* ($\star$) *and circles* ($\circ$): mean field limits; *dots*: DTMC simulation.

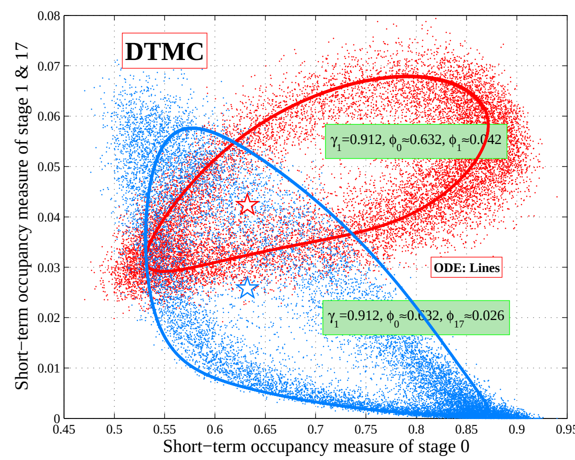
Fig. 3. Bistabilty Example: There are three solutions to the fixed point equation, two of which (γ_1 and γ_3) are stable and the other one (γ_2) is unstable. Short-term average statistics measured for each 2000 backoff time-slots suggest bistability.



(a) $f(\gamma)$ versus γ



(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 (γ_1) to the fixed point equation but the decoupling assumption does ***not*** hold. Short-term average statistics measured for each 2000 backoff time-slots suggest stable oscillation around the unique equilibrium.

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 $\pi^{\mathrm{R}} = 0$ and $\pi^{\mathrm{C}} = 1$. Suppose there are two classes H and L such that population of each class is $N^{\mathrm{H}} = N^{\mathrm{L}} = 640$. The numbers of backoff stages are assumed to be equal, *i.e.*, $K^{\mathrm{H}} + 1 = K^{\mathrm{L}} + 1 = 21$. The attempt probability at each backoff stage is:

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

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

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

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

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

$$\gamma^{\mathrm{H}} = \gamma^{\mathrm{R}} = \gamma^{\mathrm{C}} = \gamma_1 = 0.912.$$

Since there is only one solution, one might be much inclined to hazard the conjecture by Bianchi *et al.* [2], [10] that the collision probability is approximately γ_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 γ^{H} or γ^{C}.

Similarly, we have run DTMC simulations to obtain the short-term average statistics. 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 γ_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 (sold lines) as shown in Fig. 4(c). The unstability of γ_1 can be decided by the eigenvalues of the corresponding Jacobian matrix.

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

V. 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.* [10], and Ramaiyan *et al.* [15]. One counterexample in our paper has shown that this speculation is not always true, putting another emphasis on validation of the decoupling assumption which underlies the formula.

Thanks to recent advances in mean field theory [6], [1] and also [17], 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 per-stage attempt probability 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 kth backoff to q_k/N, we need to *further* weaken the node activity such that the rescaled attempt probability 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 established that this condition is sufficient for optimizing the aggregate throughput, hence it 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 [17], 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 (eODE1) and (eODE2) 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. Performance of random medium access control: An asymptotic approach. In *Proc. ACM Sigmetrics*, June 2008.

[6] 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.

[7] 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.

[8] K. Duffy. Mean field markov models of wireless local area networks. *Markov Processes and Related Fields*, accepted for publication.

[9] 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.

[10] 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.

[11] P. Kumar. An interview with Dr. P. R. Kumar. *Science Watch Newsletter, available at http://esi-topics.com/wireless/interviews/PRKumar.html*, June 2006.

[12] B.-J. Kwak, N.-O. Song, and L. Miller. Performance analysis of exponenetial backoff. *ACM/IEEE Trans. Networking*, 13(2):343–355, Apr. 2005.

[13] 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.

[14] 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.

[15] 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.

[16] T. Sakurai and H. Vu. MAC access delay of IEEE 802.11 DCF. *IEEE Trans. Wireless Commun.*, 6(5):1702–1710, May 2007.

[17] 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.

Appendix

For notational simplicity, the sets of backoff stages $\{0,\cdots,K^{\mathrm{H}}\}$, $\{0,\cdots,K^{\mathrm{L}}\}$ are denoted by $\mathbb{K}^{\mathrm{H}}$ and $\mathbb{K}^{\mathrm{L}}$, respectively.

A. Alternative Proof of Lemma 1

We show the existence and uniqueness of the equilibrium point. Recall the notation $\mathbb{K}=\{0,\cdots,K\}$. Differentiating the right-hand side of (FPE1) with respect to $\bar{q}$, we can see that the following equation determines the sign of the derivative.

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

Consider a proper subsum, δ_κ, which can obtained by replacing K with $\kappa\in\{1,\cdots,K-1\}$. 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 δ_κ 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:

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

where the second term on the right-hand side is zero or negative as q_i is nonincreasing for $i\in\mathbb{K}$. As $\delta(K)$ is zero or negative, we can conclude that the right-hand side of (FPE1) is a nonincreasing function which is positive and converges to $(K+1)/\sum_{k\in\mathbb{K}}q_k^{-1}$ at $\bar{p}=\infty$. This conclusion taken together with the fact that the left-hand side of (FPE1) 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 (**eMONO**) that the right-hand sides of (eFPE1) and (eFPE2) are nonincreasing in γ^{H} and γ^{C}, respectively. The proof of this fact is almost identical to that of Lemma 1.

Assume that there are two solutions $(\bar{q}^{\mathrm{H}},\bar{q}^{\mathrm{L}})$ and $(\acute{\bar{q}}^{\mathrm{H}},\acute{\bar{q}}^{\mathrm{L}})$ of the fixed point equation (eFPE1)-(eFPE4) and $\acute{\bar{q}}^{\mathrm{H}}\geq\bar{q}^{\mathrm{H}}$, without loss of generality. If we assume that $\acute{\bar{q}}^{\mathrm{L}}\geq\bar{q}^{\mathrm{L}}$, it follows from (eFPE4) that $\acute{\gamma}^{\mathrm{C}}\geq\gamma^{\mathrm{C}}$. Because the right-hand side of (eFPE2) is nonincreasing in γ^{C}, we must have $\acute{\bar{q}}^{\mathrm{L}}\leq\bar{q}^{\mathrm{L}}$ and hence $\acute{\gamma}^{\mathrm{C}}\geq\gamma^{\mathrm{C}}$. Now we have shown by contradiction that $\acute{\bar{q}}^{\mathrm{H}}\geq\bar{q}^{\mathrm{H}}$ implies $\acute{\bar{q}}^{\mathrm{L}}\leq\bar{q}^{\mathrm{L}}$ and $\acute{\gamma}^{\mathrm{C}}\geq\gamma^{\mathrm{C}}$.

Moreover, we can rewrite (eFPE3) in the following form:

$$\begin{aligned}\gamma^{\mathrm{H}} &= \frac{\left(1-\gamma^{\mathrm{R}}\right)^{\Delta}+\gamma^{\mathrm{R}}\sum_{i=0}^{\Delta-1}\left(1-\gamma^{\mathrm{R}}\right)^{i}}{\left\{\sum_{i=0}^{\Delta-1}\left(1-\gamma^{\mathrm{R}}\right)^{i}\right\}+\frac{(1-\gamma^{\mathrm{R}})^{\Delta}}{\gamma^{\mathrm{C}}}}\\ &= \frac{1}{\left\{\sum_{i=0}^{\Delta-1}\left(1-\gamma^{\mathrm{R}}\right)^{i}\right\}+\frac{(1-\gamma^{\mathrm{R}})^{\Delta}}{\gamma^{\mathrm{C}}}}\end{aligned}\tag{20}$$

where the second equality can be easily verified. As $\gamma^{\mathrm{R}}=1-\mathrm{e}^{-\bar{q}^{\mathrm{H}}}$ is increasing in $\bar{q}^{\mathrm{H}}$, $\acute{\bar{q}}^{\mathrm{H}}\geq\bar{q}^{\mathrm{H}}$ implies $\acute{\gamma}^{\mathrm{C}}\geq\gamma^{\mathrm{C}}$ and $\acute{\gamma}^{\mathrm{R}}\geq\gamma^{\mathrm{R}}$. Combining these with the fact that (20) is increasing in γ^{R} and γ^{C}, we can establish that $\acute{\bar{q}}^{\mathrm{H}}\geq\bar{q}^{\mathrm{H}}$ implies $\acute{\gamma}^{\mathrm{H}}\geq\gamma^{\mathrm{H}}$. On the other hand, since the right-hand side of (eFPE1) is nonincreasing in γ^{H}, the inequality $\acute{\gamma}^{\mathrm{H}}\geq\gamma^{\mathrm{H}}$ must imply $\acute{\bar{q}}^{\mathrm{H}}\leq\bar{q}^{\mathrm{H}}$.

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

We still have to establish the existence of the solution. We first note that the left-hand sides of (eFPE1) and (eFPE2) are identical functions of $\bar{q}^{\mathrm{H}}$ and $\bar{q}^{\mathrm{L}}$, respectively, from $[0,\infty)$ to $[0,\infty)$. Because (eFPE4) is increasing in $\bar{q}^{\mathrm{L}}$, for each fixed $\bar{q}^{\mathrm{H}}$, the right-hand side of (eFPE2) is a positive nonincreasing function of $\bar{q}^{\mathrm{L}}$ by the proof of Lemma 1. Likewise, as (eFPE3) is increasing in $\bar{q}^{\mathrm{H}}$ for each fixed $\bar{q}^{\mathrm{L}}$, the right-hand side of (eFPE1) is a positive nonincreasing function of $\bar{q}^{\mathrm{H}}$ by the proof of Lemma 1. This completes the proof.

C. Proof of Lemma 4

Multiplying both sides of (eFPE1) and (eFPE2) respectively by $(1-\gamma^{\mathrm{H}})$ and $(1-\gamma^{\mathrm{C}})$ yields the following equations:

$$\bar{q}^{\mathrm{H}}(1-\gamma^{\mathrm{H}})=\sigma^{\mathrm{H}}\frac{\sum_{k\in\mathbb{K}^{\mathrm{H}}}\left(\gamma^{\mathrm{H}}\right)^{k}}{\sum_{k\in\mathbb{K}^{\mathrm{H}}}\frac{(\gamma^{\mathrm{H}})^{k}}{q_k^{\mathrm{H}}}}\cdot(1-\gamma^{\mathrm{H}}),\tag{eFPE1'}$$

$$\bar{q}^{\mathrm{L}}(1-\gamma^{\mathrm{C}})=\sigma^{\mathrm{L}}\frac{\sum_{k\in\mathbb{K}^{\mathrm{L}}}\left(\gamma^{\mathrm{C}}\right)^{k}}{\sum_{k\in\mathbb{K}^{\mathrm{L}}}\frac{(\gamma^{\mathrm{C}})^{k}}{q_k^{\mathrm{L}}}}\cdot(1-\gamma^{\mathrm{C}}).\tag{eFPE2'}$$

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

E.1 We adopt the fixed point equation (eFPE1'), (eFPE2'), (eFPE3) and (eFPE4).

E.2 We note from Lemma 2 that the right-hand sides of (eFPE1') and (eFPE2') are decreasing respectively in $\bar{q}^{\mathrm{H}}$ and $\bar{q}^{\mathrm{L}}$, and less than or equal to the left-hand sides of (eFPE1') and (eFPE2') respectively at $\bar{q}^{\mathrm{H}}=1$ and $\bar{q}^{\mathrm{L}}=1$.

To complete the proof, it is sufficient to show that the left-hand sides of (eFPE1') and (eFPE2') are increasing respectively in $\bar{q}^{\mathrm{H}}$ and $\bar{q}^{\mathrm{L}}$. It follows from the proof of Lemma 2 that (**eMINT**) implies $\bar{q}^{\mathrm{H}}\leq 1$ and $\bar{q}^{\mathrm{L}}\leq 1$. It is also obvious from the form of $\bar{q}^{\mathrm{L}}(1-\gamma^{\mathrm{C}})=\bar{q}^{\mathrm{L}}\mathrm{e}^{-\bar{q}^{\mathrm{H}}-\bar{q}^{\mathrm{L}}}$ that the left-hand side of (eFPE2') in increasing in $\bar{q}^{\mathrm{L}}\in[0,1]$.

To sum up again, it is now enough to show that the left-hand side of (eFPE1') is increasing in $\bar{q}^{\mathrm{H}}\in[0,1]$. To establish this, we rewrite (eFPE3) in a compact form

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

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

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

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

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

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