Identification and Insight into a Long Transitory Phase in Random-Access Protocols

Cristina Cano
Universitat Oberta de Catalunya
Av. Carl Friedrich Gauss 5, Building B3
08860, Castelldefels (Barcelona, Spain)
Email: ccanobs@uoc.edu

David Malone
Hamilton Institute
Maynooth University, Maynooth, Ireland
Email: david.malone@nuim.ie

***Abstract*—In this work we show that random-access protocols, which are used in a range of networks (e.g. WiFi, power line communications and Internet of Things), may experience a high-throughput, extremely long (of the order of hours) transitory phase. This behaviour is not highlighted by common analysis techniques and experimental evaluations, which can lead to incorrect prediction of network performance. We identify factors that led to this transitory behaviour being overlooked in previous work. Via numerical analysis and experimental evaluation, we establish under which conditions this transitory phase occurs. Additionally, we give insight into the duration of this transitory period and its statistical properties.**

I. Introduction

A noteworthy feature of random-access protocols is that they provide a service rate dependent on the actual number of stations with a packet pending for transmission, which may not be monotonic in the number of backlogged stations. In this work we show that this state-dependent service rate may, in combination with Poisson arrivals, cause an extremely long transitory period (of the order of hours) under certain conditions. In particular, when the buffer size of the stations is large enough to be considered infinite and we operate with the sum of arrival rates slightly higher than the service rate that the system could serve in *saturation* (when all stations have a packet buffered for transmission). Consider, for example, a network formed by 50 nodes using Homeplug 1.0 Medium Access Control (MAC) [1]. Fig. 1 shows the evolution of throughput (measured every second). Observe that the network remains in a high-throughput phase for several hours and suddenly changes to a lower-throughput phase.

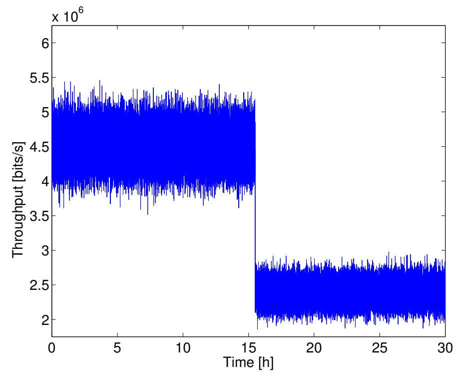


Fig. 1. Evolution of throughput in Homeplug MAC from [1].

To illustrate why this effect takes place, consider a set of nodes with no previous packets buffered for transmission. Suppose packets are generated at a rate slightly higher than the maximum rate the network could serve in saturation. Initially, the probability that a large percentage of the nodes contend for the channel at the same time is small. Thus, compared to the case in which many stations are backlogged, the time to transmit a packet is smaller, as the conditional collision probability is smaller. Consequently, the probability that a large number of packets accumulate for transmission is also low, and higher throughput than that achieved when many stations are backlogged can be obtained. However, with infinite buffer sizes, this situation cannot be maintained in the long run. Eventually, the number of backlogged stations will become large and so will the time to transmit a packet, leading to an increased number of the queued packets. As the service rate is lower than the arrival rate, the system of coupled queues is unstable [2]. Thus the long-term behaviour of these nodes corresponds to saturation throughput. Note that this effect does not correspond to bistability as, under these conditions, the queue lengths diverge and so a return to the initial high-throughput phase is eventually impossible.

We show in this work that the time to observe this long-term behaviour may be extremely long (e.g. Fig. 1), potentially leading to incorrect results from experimental assessments. Additionally, common analytical models used for prediction of network performance cannot identify the long-term operation of the network and so provide two solutions (one corresponding to the high-throughput phase and the other to saturation). Unawareness of the two-solution effect and the long duration of the transitory phase have contributed to incorrect validation of analytical and experimental results in previous work.

In this work we determine under which conditions this transitory phase occurs and give insight into its duration. The specific contributions are as follows:

1) We identify the factors contributing to this transitory phase being overlooked in previous work.
2) We determine the range of packet arrival rates for which a long transitory period may occur.

3) Via numerical analysis and experimental evaluation, we give insight into the duration of this transitory phase and its statistical properties.
4) We contribute further understanding of random-access protocols to avoid incorrect prediction of long-term results. We show this incorrect prediction has a significant potential impact on performance evaluation, parametrisation and optimisation.

This article is organised as follows. In Section II, we point to previous results that incorrectly predict long-term performance and provide an overview of related work. In Section III, we describe factors that contribute to the transitory behaviour being overlooked in common analysis and experimental evaluation and provide recommendations to avoid misprediction of network performance. In Section IV, we give more insight into the duration and statistical properties of this long transitory phase. We conclude the paper with some final remarks.

II. Related Work

Borst et al. [2] show that, when considering an infinite buffer size, the system of coupled queues that models the network behaviour of random-access protocols is unstable when the total arrival rate is higher than the saturated service rate (when all stations are backlogged). We show in this work that, for a range of arrival rates just above the saturated service rate, the system of coupled queues potentially observes a high-throughput, long-duration transitory phase. The possibility of this long transitory phase was postulated in [3] without experimental findings or formal proof.

The bistability of random-access protocols is well known, e.g. [4] and [5]. Particularly well studied is the situation of an infinite number of stations with a single packet buffer (e.g. [6], [7]), where the number of backlogged stations tends to infinity over time. For the case of a finite population of users and a single-packet buffer, a long transitory period has been identified [8], noting that the system may operate outside the stationary regime for long periods. Our aim, in contrast, is to consider the system with a finite number of stations but with larger buffer sizes, as is common in Ethernet, WiFi and power line communication systems in regimes where a long transitory phase may lead to incorrect long-term results.

Note that our contributions are fundamentally different to previous literature on bistability of random-access protocols. In particular, we complement previous contributions by considering the dynamics of the coupled system of queues when the long-term operation of the network corresponds to saturation, as defined in [2]. Thus, the interest in this work is on unstable queues (with the number of packets buffered for transmission increasing without bound) that *appear* stable for a long duration before the system saturates. To give insight into the transitory regime in this case is challenging, as common queuing theory analysis is not applicable to unstable systems. We make use of similar techniques to [8] in this work (e.g. the first hitting/exit time of a Markov model simplifying the full system of coupled queues) to estimate the transitory period between attracting equilibria, and will make use of drift analysis similar to that applied to Aloha [4], [5].

We first identified the long transitory behaviour in the Homeplug MAC in [1] (see Fig. 1) and showed that previous model results are incomplete [9], as simulation and analytic results do not correspond to the long-term network performance (inducing a relative error of 200% for some configurations). This effect is not Homeplug-specific but common to many random-access protocols. An example of such incorrect prediction in WiFi networks can be found in [10], where an analytical model with infinite buffer sizes is proposed. Higher throughput than saturation is predicted when 50 and 100 nodes are contending for the channel (reaching relative errors of 9% and 50%, respectively). In that work, disagreement is observed between analytical and experimental results (see Fig. 6 in [10]): the throughput found in simulation for certain packet arrival rates corresponds to neither the saturated throughput nor a higher throughput phase. We believe this is caused by averaging results from the transitory and long-term operations.

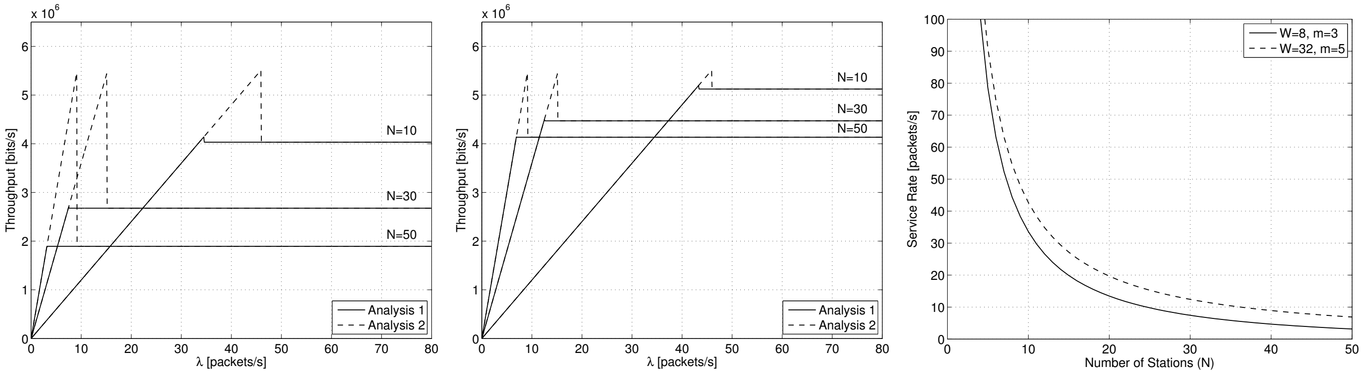
III. Factors Contributing to Performance Misprediction

Given that the system of coupled queues is unstable when the arrival rate is higher than the saturated service rate [2], how is it then possible that other previous work mispredicts network performance, showing higher throughput than that obtained in saturation? Both analysis and experimental assessments may provide the performance of the high-throughput transitory phase, thus contributing to incorrect validation of analytic models with experimental data. In this section we identify the factors, both from analysis and experimental assessments, that contribute to misprediction of long-term behaviour.

A. Factors Related to Analytical Modeling

When analysing the performance of a network of nodes using a random-access protocol, a decoupling approximation is commonly used in order to make the analysis tractable. Under this approximation each queue is modelled as independent of other queues in the network. These models involve solving a fixed point equation, but, in contrast to saturated models, the solution may not be unique in models that consider unsaturated conditions [11]. In particular, we have detected the long transitory phase as an additional solution under two conditions: *i)* for a range of packet arrival rates slightly higher than the load the system could serve in saturation and *ii)* under infinite (or large enough to be considered infinite) buffer size.

The fundamental limitation of these analytical models is that they do not consider the number of instantaneous contending/backlogged stations, i.e., the number of queues that have at least a packet buffered at the same time. Our intuition from Section I regarding the long transitory phase suggests it arises because of the difference between two extreme cases: the queues being mostly empty or saturated conditions. Under the decoupling approximation assumption, both solutions are equally feasible. However, as previously mentioned, when the arrival rate is above the saturated service rate and stations



(a) Throughput ($W = 8$, $m = 3$) (b) Throughput ($W = 32$, $m = 5$) (c) Service Rate in Saturated Conditions

Fig. 2. The two solutions in throughput and the service rate in saturated conditions obtained from the DCF renewal reward analysis.

have infinite buffer size, the only long-term solution is that of saturation.

DCF Example: To illustrate the limitation of analyses based on the decoupling approximation to obtain the long-term network performance of the system, let us consider a common analytical model of the IEEE 802.11 DCF MAC considering variable numbers of stations, traffic arrival rates and infinite buffer space. All stations are homogeneous and belong to the same collision domain (details in Appendix).

Fig. 2 shows the aggregated throughput as well as the service rate in saturated conditions for different numbers of nodes (N) and packet arrival rates (λ). We have considered setting the minimum contention window (W) to 8 and 32 and the number of backoff stages (m) to 3 and 5. Thus, the maximum contention windows are 64 and 1024 respectively (refer to [12] for more details). Parameters shown in Table I are used with data payload (L) set to 1500 bytes. This model involves solving a fixed-point equation. The solution labelled as *Analysis 1* is obtained by iterating an initial point representing saturation while the solution labelled as *Analysis 2* considers iterating an initial lightly-loaded point. Both represent solutions of the model.

Figs. 2(a) and 2(b) show that while these solutions are often the same, two distinct solutions are possible. Fig. 2(c) shows the saturation service rate and that two solutions emerge when the arrival rate is above this service rate. It seems that the decoupling assumption may cause misprediction of long-term network performance if the iterative solver converges to the *Analysis 2* solution.

1) Challenges to Model the Coupled Behaviour: The vast majority of analytical models of random-access protocols are based on the decoupling approximation. However, some authors have already pointed out the inaccuracy of this assumption and have made efforts to model the coupled dynamics of the system of queues, as proposed in [13], [14].

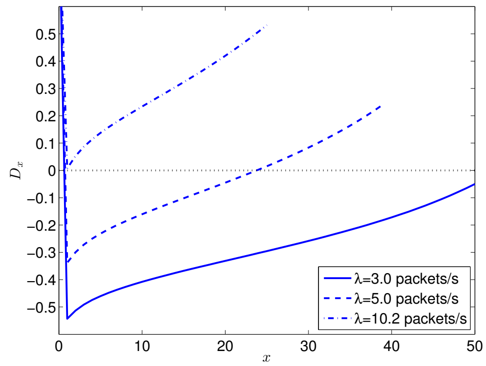
The main problem in considering the number of packets at each buffer and modelling the coupled behaviour is the large resulting state space. Keeping track of the number of packets buffered for transmission at each queue results in a state space of Q^N, where Q denotes the maximum length of the queues and N the number of stations. Consequently, the complexity of the system is considerably higher than modelling the network assuming that the decoupling approximation holds. Furthermore, as $Q \to \infty$, the analysis becomes intractable.

TABLE I
SYSTEM PARAMETERS OF THE IEEE 802.11B SPECIFICATION.

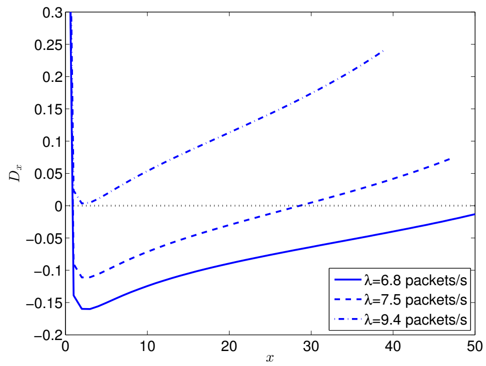
Parameter	Value in IEEE 802.11b
R_{data}	11 Mbps
$R_{\text{basic}}/R_{\text{PHY}}$	1 Mbps
L_{MACH}	272 bits
L_{PLCPPre}	144 bits
L_{PLCPH}	48 bits
L_{ack}	112 bits
σ	20 μs
DIFS	50 μs
SIFS	10 μs

2) Method to Identify the Potential for Two Solutions: Figs. 2(a) and 2(b) show that as λ increases the analysis provides only the long-term throughput solution, regardless of the initial conditions for iteration. Observe that despite our efforts to choose the initial values for the solver to produce two solutions, only one solution is obtained when the two solutions are very close (as seen in Fig. 2(a) for $N = 10, \lambda \approx 34$). Thus, a method to identify the range of packet arrival rates for which two solutions exist can be useful to avoid misprediction.

In order to gather insight into the system of coupled queues with reduced complexity, we model the system as a Markov chain with states in $\{0, 1, \ldots N\}$ in which the state represents the number of backlogged stations ($X = n_x$). Although the queue occupancy probability is not memoryless, in this section we model it as independent of the previous states as follows. The probability that a station remains still backlogged after transmitting is the standard queue busy probability: $\rho_x = \min(\lambda/\mu(x), 1)$, where x denotes the current state, i.e., the number of backlogged stations and $\mu(x)$ is determined from a model for state-dependent services (e.g. the Appendix with x saturated stations). Using this approximation, we model the queue coupling but ignore the queue occupancy at each node and only consider whether the queue has a packet pending for



(a) DCF ($W = 8$, $m = 3$)



(b) DCF ($W = 32$, $m = 5$)

Fig. 3. Drift analysis for the number of backlogged stations in DCF ($N = 50$).

transmission.

We set the number of states of the Markov chain to $1 + N'$ (from having no backlogged station up to the case in which N' stations are backlogged), where N' is the smallest value of N for which the condition $\lambda \geq \mu(n), \forall n \geq N'$. Observe that $\rho_x = 1$ when N' stations have a packet pending for transmission. Thus, in order to partially capture the conclusions in [2] without actually tracking the queue dynamics, we consider the last state to be absorbing, i.e., $P(N' \mapsto N') = 1$. The transition probabilities for $x < N'$ are shown in Eq. 1.

$$
\begin{aligned}
P(x \mapsto x+1 \leq N') &= \frac{(N-x)\lambda}{(N-x)\lambda + x\mu(x)}, \\
P(x \mapsto x-1 \geq 0) &= \frac{x\mu(x)(1-\rho_x)}{(N-x)\lambda + x\mu(x)}, \\
P(x \mapsto x) &= \frac{x\mu(x)\rho_x}{(N-x)\lambda + x\mu(x)}. \qquad (1)
\end{aligned}
$$

Now, similar to [4], we define the drift, D_x, in a given state x as the expected change in the number of backlogged stations, which according to Eq. 1 is:

$$
D_x = \frac{(N-x)\lambda - x\mu(x)(1-\rho_x)}{(N-x)\lambda + x\mu(x)}. \qquad (2)
$$

DCF Example: Fig. 3 shows the drift for each state considering $N = 50$ nodes for DCF ($W = 8, m = 3$ and $W = 32, m = 5$). We have obtained the state-dependent service rates using the analysis in the Appendix for saturated conditions. Three selected packet arrival rates, λ, are shown for each case. The smallest λ corresponds to arrival rates below saturated operation in Fig. 2, while the two higher rates correspond to operation above this stability limit. For the middle arrival rates ($\lambda = 5$ and $\lambda = 7.5$ for $W = 8, m = 3$ and $W = 32, m = 5$ respectively) the decoupled model provided two solutions in Fig. 2, while there was a single solution in the other cases.

Note in Fig. 3 that for the lowest packet arrival rates considered (unsaturated conditions), the drift is negative above a certain x. Thus, the system tends to move to states with a reduced number of backlogged stations and we observe an attractive equilibrium point close to 0.5 and 1 backlogged stations respectively. The contrary happens for the highest packet arrival rates: the drift is always positive, meaning that the system tends to move to states with a high number of backlogged stations.

Interestingly, the behaviour is more complex for the middle packet arrival rates where two solutions are obtained in Fig. 2. In these cases, we observe three equilibria: an attractive one at states with a small number of backlogged stations (approximately $x = 0.5$ and 1, respectively); a repelling one with a moderate number of backlogged stations ($x = 24$ and 29); and an attractive one at the maximum number of backlogged stations. Once the number of backlogged stations moves past the repelling equilibrium, the positive drift means that the system tends towards all stations being backlogged. Therefore, when these three equilibria are observed, we expect the system to operate in the transitory phase for a certain duration (i.e., the equilibrium point at states with small number of backlogged stations) but then, after exceeding a threshold, it tends to move to states with a large number of backlogged stations. Ultimately, results in [2] tell us that the full system remains in this state.

Note that the range of values for which we see multiple equilibria in the drift analysis closely matches the range of packet arrival rates for which multiple solutions exist in Fig. 2. An equilibrium corresponds to network conditions where the number of backlogged stations is approximately constant, these conditions also correspond to a situation with an approximately fixed collision probability for this level of transmissions. Thus, we do expect an approximate correspondence between the equilibria in this section and solutions of the fixed-point equations in the previous section. In fact, we observe the presence of a third solution to the fixed-point equation which is not discovered by the simple iterative method, corresponding to the repelling equilibrium.

Consequently, we believe this drift analysis provides a simple method for identification of packet arrival rates for which a long transitory phase is potentially present. Note, we also see in Fig. 3, that as we increase the arrival rate, the drift curve moves upwards, and so the repelling equilibrium moves to a smaller number of stations. This means we expect the threshold value to reduce as λ increases. A reduced threshold is more likely to be crossed, and so we expect that the duration of the transitory period will decrease.

B. *Factors Related to Experimental Assessment*

In simulation, we do observe that as λ increases, the system rapidly converges to the long-term operation of the network even if there is potentially a transitory phase. However, performing an experimental evaluation with arrival rate just above the saturated service rate of the system of coupled queues can provide wrong results as the length of the transitory phase can be extremely long (see Fig. 1 in Section I). Specifically, when the experiments are started with the queues empty, it can take a long time to observe the long-term behaviour since the system has to reach a point at which a large number of nodes are simultaneously contending for the channel. This can easily defeat the aim of a burn-in period for a simulation.

One way to obtain the long-term performance is to start with the queues empty, run the experiments for a long time until the system moves to long-term operation and then start taking statistics of the performance metrics of interest. The drawback of this method is that, due to the extremely long duration of the transient period, the statistics from the transitory phase must be discarded in order to reduce the bias in performance results.

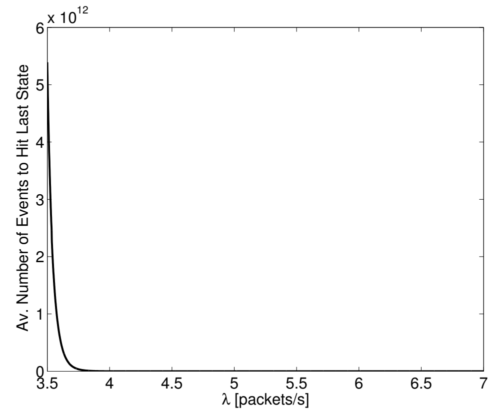
We suggest a more practical way to force the system to quickly reach long-term operation: to start the experiments with a number of packets preloaded in the queues. If the queues are unstable, the initial conditions are closer to the long-term behaviour, and the long-term throughput is obtained faster. This technique is based on the recommendation to set the initial conditions to those in steady-state proposed in [15]. This method is shown to be more effective in estimating the steady-state mean if compared to discarding the transients, as previously described. The disadvantage of this technique is that, if the system of coupled queues is stable, there is a transitory phase during which those extra packets will be released. However, the previous method for identifying the range of packet arrival rates for which a potential transitory exists based on drift analysis can be used to determine the cases where queues should be preloaded. Alternatively, simulations beginning with preloaded queues could be compared with simulations beginning with empty queues, and significant discrepancies could be regarded as reason for caution.

IV. Insights into the Duration of the Long Transitory

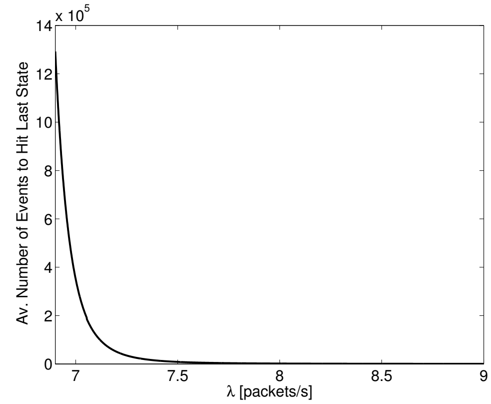
In this section we provide insights into the duration of the long transitory phase and its statistical properties using numerical analysis and experimental evaluations.

A. *A Lower Bound on the Duration of the Transitory Phase*

Here we provide a lower bound on the duration of the transitory phase via numerical analysis. With this purpose, we reuse the Markov model from Section III-A2, which tracked the number of backlogged stations, to perform a hitting time analysis to the state in which arrival rate will be greater than the service rate, so the drift leads us to all stations being backlogged.



(a) DCF ($W = 8$, $m = 3$)



(b) DCF ($W = 32$, $m = 5$)

Fig. 4. Results of the hitting time analysis on the system modelling the number of backlogged stations).

1) Hitting Time to Limiting States: Starting with the queues empty, we consider the instant, on average, at which a limiting state N', where the arrival rate exceeds the service rate is first hit. This metric allows us to track the number of events (arrivals/departures) elapsed since the network start-up until N' stations are backlogged. We can thus analytically obtain the average number of events to hit state $x = N'$ starting from $x = 0$ (denoted as $h(0)$) by solving the system of linear equations formed by Eq. 3 along with $h(N') = 0$.

$$h(x < N') = 1 + \frac{(N-x)\lambda}{(N-x)\lambda + x\mu(x)} h(x+1) + \frac{x\mu(x)(1-\rho_x)}{(N-x)\lambda + x\mu(x)} h(x-1 > 0) + \frac{x\mu(x)\rho_x}{(N-x)\lambda + x\mu(x)} h(x). \tag{3}$$

This analysis is similar to the FET analysis in [8] and is extremely computationally efficient. Thus it can be used to perform an extensive numerical evaluation. However, results are affected by the approximation of not tracking the queue occupancies of the nodes. Also observe that while reaching $x = N'$ is necessary to end the transitory phase, having N' backlogged stations is not sufficient for the system to move to long-term operation. If the number of packets in the queues is small when $x = N'$, there is some non-zero probability that stations transmit those packets without a consequent increase in the number of packets accumulated for transmission. Thus, the average number of events to hit state N' is a lower bound for the events necessary to escape from the transitory period. We illustrate this later in this section.

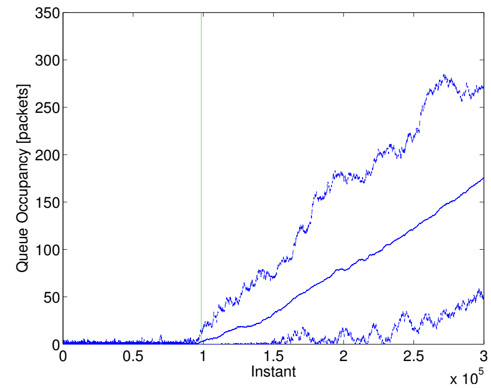
DCF Example: We have solved the linear system in Eq. 3 for $N = 50$ nodes for the DCF ($W = 8$, $m = 3$ and $W = 32$, $m = 5$), obtaining the state-dependent service rates ($\mu(x)$) from the analytical model presented in the Appendix, considering saturated conditions. Fig. 4 shows the average number of events to hit state N' for different arrival rates. Observe that, in all cases considered, as the packet arrival rate increases, the average number of events to hit state N' tends to zero. On the contrary, for reduced packet arrival rates, it can be substantial. One might expect that the mean hitting time would have a relatively simple leading term. We found that a simple curve for the form $A(\lambda - \lambda_0)^\alpha$ provides a relatively good fit, where λ_0 is close to the saturation throughput and $\alpha < 0$ is also dependent on the network configuration. Considering the hitting time analysis provides a lower bound for the duration of the transitory phase, these results demonstrate that for arrival rates above than the saturated throughput ($\mu(N)$), the duration of the transitory phase can be extremely long.

2) Modelling the System of Coupled Queues: To illustrate the fact that the previous hitting time analysis is a lower bound on the actual duration of the transitory phase, we present here experimental results on the queue occupancy evolution of the system modeling the coupled behaviour of the queues. We model the system of N parallel queues as a discrete Markov chain with states in $\{0, 1, \ldots, Q\}^N$, the state represents the number of packets waiting for transmission at each queue: $X = (X_1, ..., X_N)$. We assume the system of parallel queues to be homogeneous, i.e., same maximum queue length, packet arrival and conditional service rate distribution at all queues. The number of backlogged stations in a given state x is denoted by n_x and represents the number of queues with at least one packet pending for transmission. We take into account Poisson arrivals at rate λ packets/s. The service rate ($\mu(n_x)$) is also considered to be exponentially distributed as well as state-dependent. Including dependence on the number of backlogged stations (n_x) changes quantities in the model, such as the conditional collision probability and/or the average backoff duration. The transition probabilities among the different states of this process are:

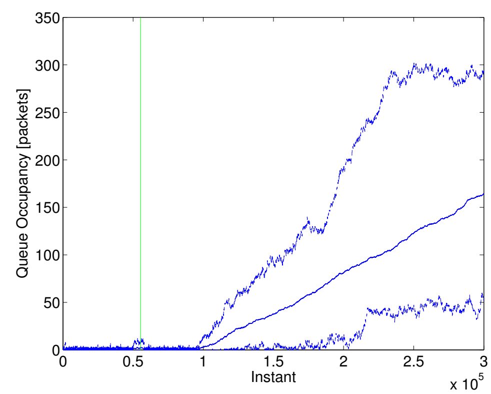
$$\begin{aligned} P(x \mapsto x + e_i \leq q) =& \frac{\lambda}{N\lambda + n_x \mu(n_x)}, \\ P(x \mapsto x - e_i \geq 0) =& \frac{\mu(n_x)}{N\lambda + n_x \mu(n_x)}, \end{aligned} \tag{4}$$

with the relational operators being element-wise and q and e_i being the all Q and i-th unit vectors in $\mathbb{Z}_+^N$, respectively. The number of backlogged stations in state x is computed as:

$$n_x = \sum_{i=1}^{N} I(x_i), \tag{5}$$



(a) Simulation Run 1



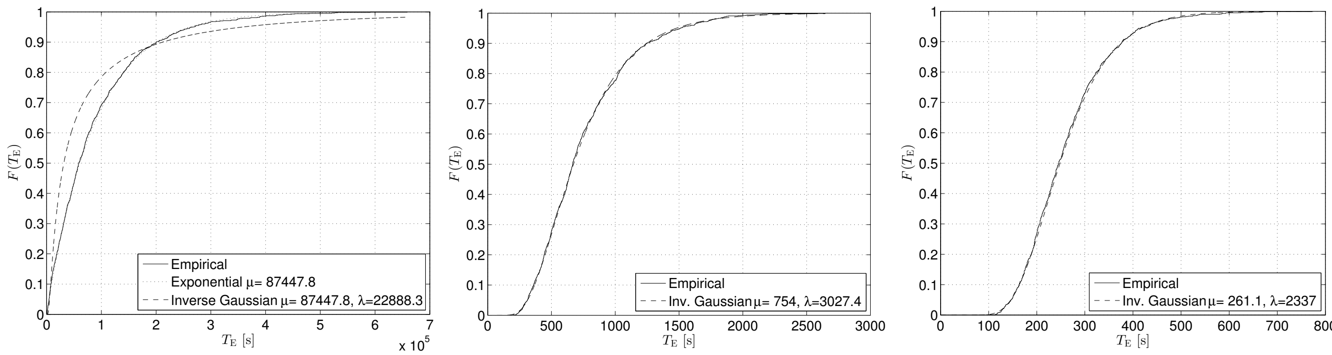
(b) Simulation Run 2

Fig. 5. Evolution of the queue occupancy (maximum, average, minimum) of the N nodes for two simulation runs in DCF ($\lambda = 7.5$ packets/s). Hitting instant of first state such that $\lambda > \mu(n_x)$ also displayed (vertical line).

where $I(x_i)$ is the indication function of having at least one packet pending for transmission in queue i (i.e., $I(x_i) = 1$ if $x_i > 0$ and 0 otherwise).

This system provides a close description of the behaviour of the network with the main assumption being Poisson service. As we mentioned in Section III, the complexity of solving it explicitly is prohibitive. Its state space is of the order of Q^N and we are interested in the case when $Q \to \infty$. In fact, the system is computationally intractable even for small N, even taking into account that the transition matrix is sparse. However, assessment via Monte Carlo simulation is practical.

DCF Example: We track the queue evolution at every instant (packet arrival/departure) of Monte Carlo simulations using the state-dependent service rates from the model of the DCF described in the Appendix. The minimum, average and maximum queue length of $N = 50$ nodes using the DCF protocol with $\lambda = 7.5$ packets/s for two different simulation runs are depicted in Fig. 5. Observe that the queue occupancies remain low for a long interval, until they start to increase to the maximum queue length (effectively leaving the transitory phase and entering into saturated conditions). We plot the instant at which the limiting state (first x such that $\lambda > \mu(n_x)$) is first reached (vertical line). Note that, in Fig. 5(a), this instant coincides with the moment at which the queues start



(a) $\lambda = 7.5$ packets/s (b) $\lambda = 7.75$ packets/s (c) $\lambda = 8$ packets/s

Fig. 6. Empirical CDF obtained via network simulations of the last instant $N-1$ stations backlogged using in DCF with $W = 32$, $m = 5$. Inverse Gaussian (and exponential in Fig. 6(a)) distribution with parameters selected to best fit the empirical distribution also depicted.

to be filled with packets. However, in Fig. 5(b), the system is able to recover from this situation and remain in the transitory phase for a longer interval, showing that the hitting time analysis in the previous subsection actually corresponds to a lower bound on the duration of the transitory period.

B. Transient Duration and Statistical Properties

For empirical assessment of the transitory phase duration and evaluation of its statistical properties, we use a network simulator based on the SENSE framework [16]. Packet interarrival times are modelled as exponentially distributed while the service rate strictly follows the DCF random-access procedure.

We estimate the long transitory duration as T_E, that we define as the last time instant at which $N-1$ stations were backlogged once the average queue occupancy reached 75% of the maximum queue size (with $Q = 1000$ packets). The empirical Cumulative Distribution Functions (CDF) of T_E, obtained from 1000 simulation runs for different packet arrival rates with $N = 50$ and $W = 32$, $m = 5$ is shown in Fig. 6.

First, observe that for shorter durations of the transitory period (Figs. 6(b-c)), the empirical distribution resembles that of an inverse Gaussian (best goodness of fit). However, for longer durations of the transitory period (Fig. 6(a)), the distribution obtained can be better described as an exponential. These numerical results suggest that the actual distribution could be described as a combination of two distributions, with the exponential one having more influence for longer durations (smaller packet arrival rates) and the inverse Gaussian being more relevant for shorter durations (higher packet arrival rates). Second, note that the actual length of the transitory period in simulations when the packet arrival rate is 7.5 packets/s is in the order of several hours (Fig. 6(a)).

V. Final Remarks

In this work we identify that there is a potential to observe a long-duration transitory phase in random-access protocols when we operate right above the saturated service rate of the system of coupled queues and under certain circumstances, such as infinite buffer size and exponentially distributed interarrival of packets.

We first identified that previous work has overlooked this transitory regime, leading to significant misprediction of long-term performance. We then determined the relation between incorrect prediction of long-term performance with *i)* the use of iterative solvers of analytical models based on the decoupling approximation and *ii)* the presence of the extremely long transitory phase in experimental evaluations.

Then, we determined the range of packet arrival rates for which there is a potential for a long transitory phase to occur by evaluating the equilibrium points via drift analysis of the system modeling the number of backlogged stations. We also provided a lower bound on the long transitory duration via a hitting times numerical evaluation and gave more insight into the actual duration of the transitory phase and its statistical properties through network simulations.

Appendix: Renewal Reward Analysis of the WiFi DCF Medium Access Control Protocol

We take a common renewal reward approach [17], [18], [19] to model the DCF network performance that makes use of the decoupling approximation. We also consider: *i)* large enough to be considered infinite buffer size and retry limit, *ii)* exponentially distributed interarrival of packets, *iii)* ideal channel conditions, and *iv)* that all nodes can overhear each other's transmissions. The mean queue occupancy (ρ) of a node is derived considering the service rate (μ) and the packet arrival rate from the network layer (λ) as $\rho = \min(\lambda/\mu, 1)$.

The service rate depends on the following: *i)* the total time on average spent in transmitting packets that result in a collision and *ii)* the time spent successfully transmitting the packet, both including the total average backoff duration until the successful frame transmission ($E[w]\alpha$):

$$\mu = 1/((n_\mathrm{t} - 1)(E[w]\alpha + T_\mathrm{c}) + E[w]\alpha + T_\mathrm{s}), \qquad (6)$$

where n_t is the average number of attempts to successfully transmit a packet. A successful transmission (T_s) and a collision (T_c) are computed as: $T_\mathrm{s} = T_\mathrm{c} = \text{DIFS} + T_\mathrm{fra} + \text{SIFS} + T_\mathrm{ack}$, where T_fra and T_ack denote the times to transmit the frame and the acknowledgement, respectively. We compute T_{fra} as shown in Eq. 7 and T_{ack} as in Eq. 8.

$$T_{\text{fra}} = \frac{L_{\text{PLCPPre}} + L_{\text{PLCPH}}}{R_{\text{PHY}}} + \frac{L_{\text{MACH}}}{R_{\text{basic}}} + \frac{L}{R_{\text{data}}}, \quad (7)$$

$$T_{\text{ack}} = \frac{L_{\text{PLCPPre}} + L_{\text{PLCPH}}}{R_{\text{PHY}}} + \frac{L_{\text{ack}}}{R_{\text{basic}}}, \quad (8)$$

with L_{PLCPPre}, L_{PLCPH}, L_{MACH}, L_{ack} and L being the length of the PLCP preamble, PLCP header, MAC header, acknowledgement and data payload, respectively, while R_{PHY}, R_{basic} and R_{data} denote the physical, basic and data rates [12].

Under the decoupling assumption with an infinite number of retries, the average number of attempts to transmit a frame (n_{t}) is computed as $n_{\text{t}} = /(1-p)$, where the conditional collision probability (p) is obtained as the complementary of having at least one of the other $N-1$ nodes transmitting a frame in the same slot (Eq. 9), with τ denoting the attempt rate of a node.

$$p = 1 - (1-\tau)^{N-1}. \quad (9)$$

We view the attempt rate as a regenerative process, where the renewal events are when the MAC begins processing a new frame. Thus, we apply the renewal reward theorem (Eq. 10).

$$\tau = \frac{n_{\text{t}}}{n_{\text{t}}\,(E[w]+1) + I}. \quad (10)$$

The term I in Eq. 10 accounts for the number of slots in idle state (when there is no packet waiting in the queue for transmission) and is computed as the probability of having an empty queue over the probability of a packet arrival in a slot. Considering an M/M/1 FIFO queue, we then compute I as in Eq. 11.

$$I = \frac{1-\rho}{1-e^{-\lambda\alpha}}, \quad (11)$$

where α is the average slot duration while the node is in backoff, which is derived depending on the type of slot that is overheard. A slot can be empty if no other node transmits (that occurs with p_{e} probability) and, in such a case, its duration is σ (defined in [12]). Otherwise, it can be occupied due to a successful transmission (that happens with probability p_{s}) or a collision (that occurs with p_{c} probability), with durations T_{s} and T_{c}, respectively. Thus: $\alpha = p_{\text{s}}T_{\text{s}} + p_{\text{c}}T_{\text{c}} + p_{\text{e}}\sigma$, with:

$$\begin{aligned} p_{\text{s}} &= (n-1)\tau(1-\tau)^{N-2}, \\ p_{\text{e}} &= (1-\tau)^{N-1}, \\ p_{\text{c}} &= 1 - p_{\text{s}} - p_{\text{e}}. \end{aligned} \quad (12)$$

The average number of backoff slots can be computed as shown in Eq. 13 derived in [17].

$$E[w] = \frac{1-p-p(2p)^m}{1-2p}\frac{W}{2} - \frac{1}{2}. \quad (13)$$

Finally, we obtain the throughput as: $S = \rho\mu L$.

Acknowledgment

This publication has emanated from research supported in part by the grant from the Spanish Ministry of Science TEC2015-71303-R (MINECO/FEDER) and from the Science Foundation Ireland (SFI), co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

References

[1] C. Cano and D. Malone, "On efficiency and validity of previous Homeplug MAC performance analysis," *Computer Networks. Available at: http://dx.doi.org/10.1016/j.comnet.2015.03.005*, 2015.

[2] S. Borst, M. Jonckheere, and L. Leskelä, "Stability of parallel queueing systems with coupled service rates," *Discrete Event Dynamic Systems*, vol. 18, no. 4, pp. 447–472, 2008.

[3] K. H. Suleiman, T. Javidi, M. Liu, and S. Kittipiyakul, "The impact of MAC buffer size on the throughput performance of IEEE 802.11," *Technical Report*, 2008.

[4] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data networks*. Prentice-Hall International New Jersey, 1992, vol. 2.

[5] L. Dai, "Stability and delay analysis of buffered aloha networks," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 8, pp. 2707–2719, 2012.

[6] D. J. Aldous, "Ultimate instability of exponential back-off protocol for acknowledgment-based transmission control of random access communication channels," *Information Theory, IEEE Transactions on*, vol. 33, no. 2, pp. 219–223, 1987.

[7] F. P. Kelly, "Stochastic models of computer communication systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 379–395, 1985.

[8] L. Kleinrock and S. S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *Communications, IEEE Transactions on*, vol. 23, no. 4, pp. 410–423, 1975.

[9] M. Y. Chung, M.-H. Jung, T.-J. Lee, and Y. Lee, "Performance analysis of HomePlug 1.0 MAC with CSMA/CA," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1411–1420, 2006.

[10] J. M. Pitts and O. M. Shepherd, "Analysing the transition between unsaturated and saturated operating conditions in 802.11 network scenarios," in *Proceedings of the Military Communications Conference (MILCOM).*, 2008, pp. 1–7.

[11] K. R. Duffy, "Mean field Markov models of wireless local area networks," *Markov Processes and Related Fields*, vol. 16, no. 2, pp. 295–328, 2010.

[12] IEEE Std 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *ANSI/IEEE Std 802.11*, 1999 Edition (Revised 2007).

[13] M. Garetto and C.-F. Chiasserini, "Performance analysis of 802.11 WLANs under sporadic traffic," in *Proceedings of the Springer Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, 2005, pp. 1343–1347.

[14] M. K. Panda and A. Kumar, "State dependent attempt rate modeling of single cell IEEE 802.11 WLANs with homogeneous nodes and Poisson arrivals," in *Proceedings of the First IEEE International Communication Systems and Networks and Workshops (COMSNETS)*, 2009, pp. 1–10.

[15] J. R. Wilson and A. A. B. Pritsker, "Evaluation of startup policies in simulation experiments," *Simulation*, vol. 31, no. 3, pp. 79–89, 1978.

[16] G. Chen, J. Branch, M. Pflug, L. Zhu, and B. Szymanski, "SENSE: A Sensor Network Simulator," *Advances in Pervasive Computing and Networking*, pp. 249–267, 2004.

[17] G. Bianchi and I. Tinnirello, "Remarks on IEEE 802.11 DCF performance analysis," *IEEE Communication Letters*, vol. 9, no. 8, pp. 765–767, 2005.

[18] K. Medepalli and F. A. Tobagi, "System centric and user centric queueing models for IEEE 802.11 based wireless LANs," in *Proceedings of the 2nd International Conference on Broadband Networks (BroadNets)*, 2005, pp. 612–621.

[19] B. Bellalta, M. Oliver, M. Meo, and M. Guerrero, "A simple model of the IEEE 802.11 MAC protocol with heterogeneous traffic flows," in *Proceedings of the IEEE Eurocon*, November 2005.