On Stability and Maximum Throughput of Exponential Backoff Mechanisms with Two Users

Flaminio Borgonovo
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Piazza L. da Vinci 32, 20133
Milano, Italy
borgonov@elet.polimi.it

Matteo Cesana
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Piazza L. da Vinci 32, 20133
Milano, Italy
borgonov@elet.polimi.it

ABSTRACT

The Binary Exponential Backoff (BEB) mechanism plays a key role in determining stability and efficiency of popular MAC protocols, such as IEEE 802.3 and IEEE 802.11. Unfortunately, the maximum throughput allowed by this protocol under different system parameters and user's multiplicity is largely unknown. In this paper we evaluate the capacity of a two-station system under Bernoulli arrivals, which is proved in the literature to be lower bounded by 0.3 under balanced arrivals. We refer to the general exponential backoff law b^{-i} and show that, when close to capacity, the system behaves in a simple manner, which allows us to assess capacity without resorting to the general Markov Chain approach. Our approach shows that the capacity of BEB under balanced load is equal to 0.6096. We also provide the stability region when arrivals are unbalanced in the two queues. Simulation results support our finding.

Categories and Subject Descriptors

C.4 [**Computer Systems Organization**]: Performance of Systems

General Terms

Performance, Theory

Keywords

Binary Exponential Backoff, Markov Chain, Throughput

1. INTRODUCTION

The Binary Exponential Backoff (BEB) mechanism plays a key role in determining stability and efficiency of popular MAC protocols, such as IEEE 802.3 and IEEE 802.11, and consequently, its analysis has attracted much attention in the last years. Several works have appeared trying to assess throughput, stability, and delay performance, but, unfortunately, even the limited goal of determining the capacity

of such protocols appears to be a hard task because of the complexity of the analysis.

In [2], Aldous considers a slotted channel with Markovian BEB, where stations transmit the packet at the head of their queue with probability b^{-i}, $b > 1$, where i is the number of consecutive collisions suffered by the packet. The author shows that BEB is unstable under the infinite population model and any positive packet arrival rate λ, so that its capacity is zero.

When a finite number M of station is assumed, however, BEB capacity evaluation is still an open issue. The common analytical approach here resorts to a Markov Chain whose state space is described by vector
$(q_1, q_2, \ldots, q_M, b_1, b_2, \ldots, b_M)$, where q_i and b_i represent the queue length and the backoff state index at station i, respectively. Unfortunately, even in the simplest case $M = 2$ an exact analysis is unfeasible due to the complexity of the state space. Nevertheless, bounds on capacity have been given. In [5], Goodman *et al.* prove an arrival frequency $\lambda^*(M) > 0$ does exists such that the system is stable if $\lambda(M) < \lambda^*(M)$, where $\lambda^*(M) \geq 1/M^{\alpha \log M}$ for some constant α. In [1], Al-Ammal *et al.* improve the bound in [5] proving that BEB is stable for arrival rates smaller than $1/\alpha M^{1-\eta}$, where $\eta < 0.25$. Finally, in [4], Håstad *et al.* show, using the same analytical model as in [5], that BEB is unstable whenever $\lambda_i > \lambda/M$ for $1 \leq i \leq M$, and $\lambda > 0.567 + 1/(4M - 2)$, where λ is the system arrival rate and λ_i is the arrival rate at node i.

An approximated analysis of non-Markovian BEB adopting uniform transmission windows with non-zero initial value is introduced in [3]: here the model considers saturation conditions, i.e., queues always full, and adopts a Markov Chain approach based on the simplifying assumption that collision probability is constant in time and equal to its average value. This method has gained popularity due to its ability to provide results in good accordance with simulations for the cases with $M > 2$ (see for example [6]). Unfortunately, this method cannot provide definite results on capacity, especially when the BEB range is unlimited and the initial window size is zero, i.e., new packets are immediately transmitted.

In this paper, we derive the capacity of a two-station system under Bernoulli arrivals and Markovian BEB, the same as in [5], where capacity has been shown to be lower bounded by 0.3 under balanced arrivals. Our approach exploits the specificity of the $M = 2$ case, and does not resort to the general Markov Chain method. We find that the capacity

of BEB under balanced load is equal to 0.6096, and we also provide the stability region when traffic is unbalanced. The paper is organized as follows: Section 2 provides the analytical evaluation of BEB capacity and presents simulation results to support it; in Section 3, the stability region of BEB under unbalanced traffic is derived. Concluding remarks are drawn in Section 4.

2. ANALYSIS

Our analysis is inspired to the "saturation condition" approach of [3]. The latter is based on the fact that capacity coincides with the throughput of the system when queues are always full, a concept derived from queueing theory. In our case, things are somehow more complicated since queues interact through the backoff protocol. As a matter of fact, we show below that were the two queues always full the system would not be stable, as one of the stations would capture the channel and transmit forever. This does not happen, however, if queues, though arbitrary large, are finite in length. Therefore, we evaluate capacity as the throughput with very large queues.

We consider a two-station system with Bernoulli arrivals at each queue with intensity $\lambda/2$, later extending the achieved results to the unbalanced case. Stations transmit the packets at the head of their queue according to the exponential backoff law b^{-i}, $b > 1$, where i is the backoff state index, i.e., the number of consecutive collisions suffered by the packet. If both stations transmit in the same slot a collision occurs and both backoff indexes are incremented.

Our argument is based on the specific behavior of system under heavy load, when queues are large. In Figure 1(a), we have reported the state diagram of backoff indexes in saturation conditions (queues always full) where we have denoted $\beta = 1/b$. Due to the perfect symmetry of states (i, j) and (j, i), the diagram in Figure 1(a) can be reduced to the diagram in Figure 1(b). Here, states $(i, i), \forall i$, are easily recognized as transient states, since they can never be re-entered. The other states constitute an irreducible set of states which we show to represent a non-positive-recurrent Markov Chain $\mathcal{C}$, i.e., where the asymptotic probability of any state (i, j) is zero. To see this, we change $\mathcal{C}$ into another Markov chain $\mathcal{D}$, whose state space is shown in Figure 2(a). This is obtained by changing the probability flux $(1, j) \rightarrow (2, j+1)$, in $\mathcal{C}$, into the probability flux $(1, j) \rightarrow (1, j+1)$. The changes introduced assure that if chain $\mathcal{D}$ is non-positive-recurrent, then also chain $\mathcal{C}$ is non-positive-recurrent. In fact, to see that $\mathcal{C}$ is non-positive-recurrent, let consider chain $\mathcal{E}$ in Figure 2(b), which presents the same probability fluxes as chain $\mathcal{D}$ for the states that are common to both. Therefore, the solution of the equilibrium equations for chain $\mathcal{E}$ is the same as in chain $\mathcal{D}$ for the same states. This solution is easily derived as

$$\pi_i = \pi_1 \frac{\beta^{i-1}}{\prod_{k=2}^{i}(\beta + \beta^k + \beta^{k+1})}. \tag{1}$$

where index i denotes the right index of states in chain $\mathcal{E}$. The solution for the remaining states in chain $\mathcal{D}$ (states $(0, 2), (0, 3), \ldots, (0, i)$, is easily obtained as

$$\pi_i' = \pi_i \frac{1 - \beta^i}{\beta^{i-1}} = \pi_1 \frac{1 - \beta^i}{\prod_{k=2}^{i}(\beta + \beta^k + \beta^{k+1})}. \tag{2}$$

From (1) and (2) it can be seen that for any non zero value of π_1 we have $\sum(\pi_i + \pi_i') = \infty$, and, therefore, chain $\mathcal{D}$ is non-positive-recurrent.

Being chain $\mathcal{C}$ non-positive-recurrent means that either state indexes goes to infinity. However, when the right index is large only one path to infinity becomes possible. In fact, we have the following transition probabilities out of state (i, j):

$$p_{(i,j)->(i+1,j+1)} = b^{-(i+j)},$$
$$p_{(i,j)->(0,j)} = b^{-i}(1 - b^{-j}). \tag{3}$$

When j increases without limit the former becomes negligible (in probability) with respect to the latter, so that only the path $(i, j)- > (0, j)$ becomes possible. In the same way, path $(1, j+1)- > (2, j+2)$ becomes negligible (in probability) with respect to path $(1, j+1)- > (0, j+1)$ and we can conclude that for high values of j the only path to infinity is represented by the pattern $(0, j)- > (1, j+1)- > (0, j+1)$. In these condition the channel is captured by the station with the lowest index that keeps on transmitting, rarely interrupted by a collision with the other station (transition $(0, j)- > (1, j+1)$), immediately followed by a correct transmission from the former that again captures the channel. In this phase collisions occur at an average distance equal to b^j, where j is the backoff state of the backlogged station.

Now, if the system is stable, but starting with very large queues, at the beginning the behavior of the backoff chain is the same as in the infinite case and the path cited above becomes again the only one possible, because in the first phase the system behaves as in saturation conditions. Therefore, as in the previous case, one station captures the channel and keeps on transmitting until its queue empties.

To find BEB capacity, we introduce the concept of *cycle*. We call *cycle* the time period needed to re-enter backoff state $(0, 0)$. Cycles are composed of three periods, **0**, **a** and **b**, where **a** and **b** can be absent, whose length we denote by t_0, t_a and t_b respectively.

Period 0 is an initial phase in which both stations contend for the channel, occasionally transmitting with success, and where queues build up. Period **0** ends by entering period **a**, during which one station, say station 1, captures the channel, i.e., it keeps on transmitting, with rare collisions until its queue empties. At this point station 1 enters period **b**, in which empty slots are interrupted either by transmissions of newly arrived packets at station 1 or by rare collisions with station 1. This period ends when the station 2 succeeds in transmitting, at which slot both backoff indexes are re-set to zero and the cycle ends. As long as the input traffic λ is close to capacity, cycles alternate on the channel, queues increase and empty, showing a stationary behavior.

If the traffic increases beyond capacity, the system behavior becomes unstable. Starting with empty queues, cycles still alternate on the channel, but, since arriving packets are more than successful transmissions, queues increase their average content without limit, and so does the capture period, until asymptotically one of the stations can no longer transmit, never resets its backoff index and its throughput drops to zero.

In the stable region, the system throughput can be evaluated as the average throughput in a cycle and can be expressed as
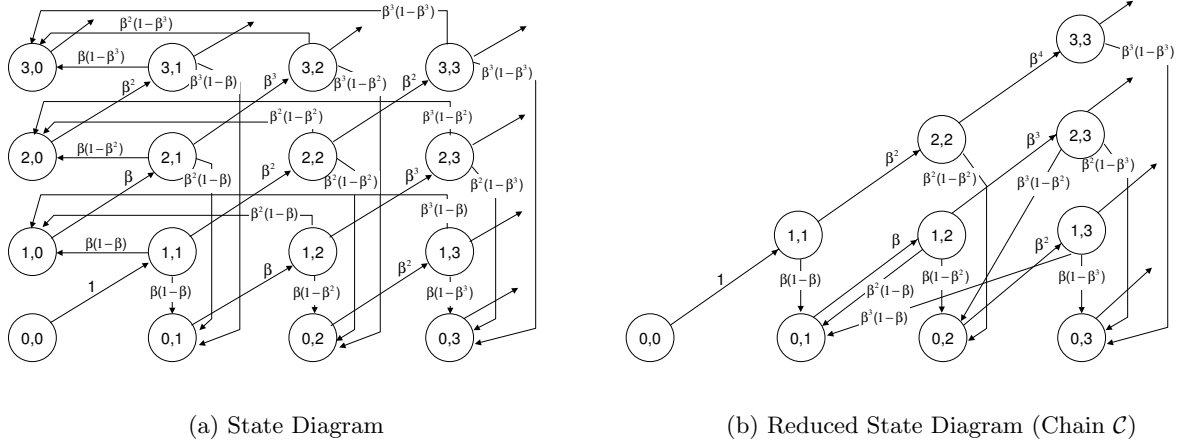
(a) State Diagram

(b) Reduced State Diagram (Chain $\mathcal{C}$)

Figure 1: *State diagrams of backoff indexes with queues always full.*

(a) Modified State Diagram (Chain $\mathcal{D}$)

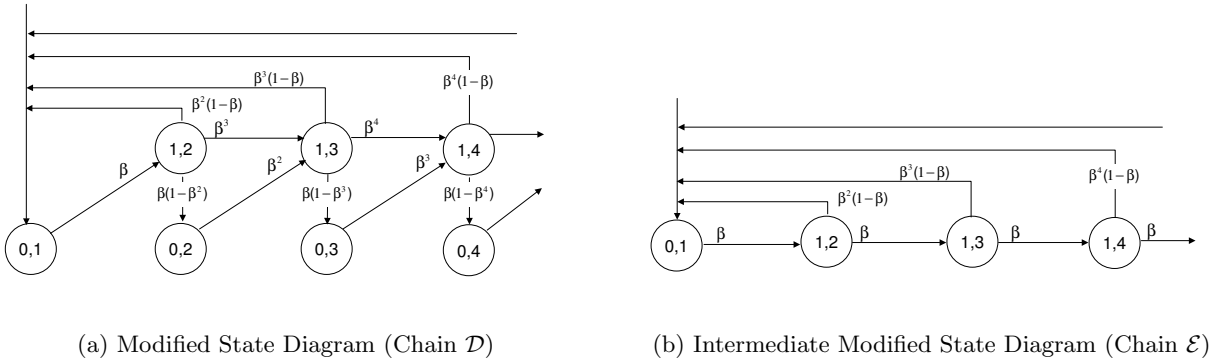(b) Intermediate Modified State Diagram (Chain $\mathcal{E}$)

Figure 2: *Modified State diagrams.*

$$\lambda = \frac{E[t_0]\lambda_0 + E[t_a]\lambda_a + E[t_b]\lambda_b}{E[t_0] + E[t_a] + E[t_b]}, \qquad (4)$$

where λ_0, λ_a and λ_b denote the throughput during periods **0**, **a** and **b**, respectively.

When the traffic approaches capacity, the queues increase with no limit together with $E[t_a]$ and $E[t_b]$, so that $E[t_0]$ becomes negligible with respect to the cycle length and can be dropped from (4). In the same circumstances, according to what has been explained above, λ_a approaches one while λ_b approaches the input traffic λ_1, actually $\lambda_1 = \lambda/2$. In fact, station 2 transmits very rarely, and station 1 immediately and successfully transmits all the packets entering its queue at rate λ_1. Hence, we can write:

$$\lambda = \frac{E[t_a] + E[t_b]\lambda/2}{E[t_a] + E[t_b]} = \frac{E[t_a]/E[t_b] + \lambda/2}{E[t_a]/E[t_b] + 1}. \qquad (5)$$

Equation (5) may be regarded as an equation in λ whose solution yields capacity, provided we are able to evaluate $E[t_a]$ and $E[t_b]$. To this purpose, we note that a very small increase in λ toward capacity causes a dramatic increase
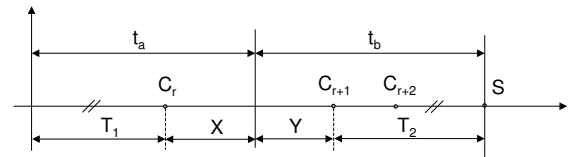

Figure 3: *Relations among the different period definitions.*

in $E[t_a]$ and $E[t_b]$, being their ratio almost constant; this means that the absolute values of $E[t_a]$ and $E[t_b]$, together with the queue length at the beginning of the cycle, have no influence on the solution of (5) provided they are large enough.

To proceed further, referring to Figure 3, let denote by C_r

the last collision before queue 1 empties, being r is the back-off index of station 2, and by S the first successful transmission of station 2; then, periods t_a and t_b can be decomposed as shown in the same figure:

$$E[t_a + t_b] = E[T_1] + E[X] + E[Y] + E[T_2]$$

Period T_1 is the sum of the backoff periods of station 2 up to beginning of period r. These periods have an average value b^i, being i the state index of backoff 2, so that we have:

$$E[T_1] = \sum_{i=0}^{r-1} l_i = \sum_{i=0}^{r-1} b^i = \frac{b^r - 1}{b - 1} \simeq \frac{b^r}{b - 1} \qquad (6)$$

Random variable Y, owing to the memoryless backoff mechanism, is geometrically distributed as an entire backoff period, i.e., with average $l_r = b^r$. Therefore, the entire period $E[Y + T_2]$ is the sum of backoff periods of station 2 from state r until success. Each subsequent collision occurs with probability $\lambda/2$, the probability that at station 1, that has an empty queue, arrives a new packet, and those events are statistically independent since arrivals are statistically independent. Therefore the number of backoff periods in t_b is a r.v. geometrically distributed with parameter $\lambda/2$, and we have

$$E[Y + T_2] = \sum_{k=0}^{\infty} (1 - \lambda/2)(\lambda/2)^k \sum_{i=r}^{k+r} b^i = \frac{b^r}{1 - b\lambda/2} \qquad (7)$$

where the right result holds for $\lambda/2 < 1/b$, otherwise $E[Y + T_2]$ is infinite and the system is not stable. The same condition for stability is stated in [5] for the binary case $b = 2$.

Substituting (6) and (7) into (5) yields:

$$\lambda = \frac{b^r(\frac{1}{b-1} + \frac{\lambda/2}{1 - b\lambda/2}) + E[X]}{b^r(\frac{1}{b-1} + \frac{1}{1 - b\lambda/2}) + E[X]}. \qquad (8)$$

Note that setting $E[X] = 0$ provides a lower bound to the right-hand side and therefore the lower bound to capacity:

$$\lambda_{cl} = \frac{1}{b} \qquad (9)$$

To say more on the efficiency and capacity we need to evaluate $E[X]$. If events C_j determined by the backoff process of station 2 were stationary and reversible, then X would have the same distribution of Y. However, events C_j are not stationary. In the Appendix 4, we show that, given any very large instant n, the distance to the first event previous to n, $X(n)$, presents the same distribution of the distance to the first event next to n, $Y(n)$. Furthermore, there are a variety of arguments that suggest that random instants t_a behave as the deterministic instant n in determining X, although we can not provide a rigorous proof. Therefore we explicitly assume $E[X] = E[Y] = b^r$. Exploiting this equality (8) becomes

$$\lambda = \frac{\frac{b}{b-1} + \frac{\lambda/2}{1 - b\lambda/2}}{\frac{b}{b-1} + \frac{1}{1 - b\lambda/2}}. \qquad (10)$$

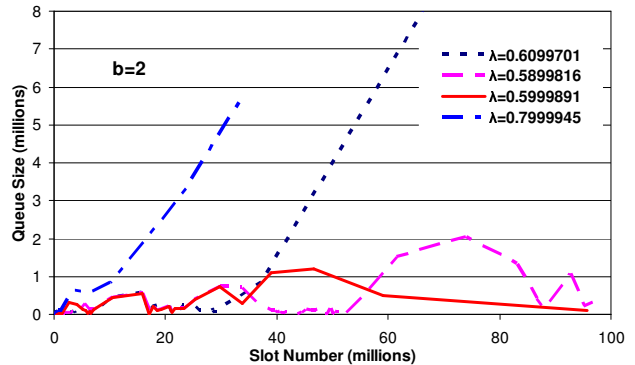| $1/b$	0.5	0.6	0.7	0.8	0.9	1
capacity	0.6096	0.6830	0.7545	0.8283	0.9083	1

Figure 4: *Queue size at the end of each cycle versus simulation time for different total arrival rates in the* $b = 2$ *case.*

that, solved in λ finally provides the capacity as

$$\lambda_c = \frac{b^2 + 3b - 1 - \sqrt{b^4 - 2b^3 + 7b^2 - 6b + 1}}{2b^2} \qquad (11)$$

Table 1 reports the capacity values for different values of the backoff parameter b.

Figure 2 shows some simulation results obtained simulating the 2-stations system with a time-slotted channel for transmission. We have reported sample curves for the BEB case ($b = 2$) representing the size of the queue at the end of the cycle, having selected only cycles that presents channel capture with backoff index at least equal to 6. Even if this simulation analysis cannot be quantitative, it clearly shows that, in the case of $\lambda = 0.60$, the queue dimension stays limited suggesting a stable behavior, whereas when arrival rate grows beyond $\lambda = 0.61$, the queue dimension clearly diverges, indicating an unstable behavior.

3.   GENERALIZATIONS

In the general case in which λ_1 differ from λ_2, we have two versions of (5), according to the identity i of the station that captures the channel:

$$\eta_i = \frac{E[t_{ia}] + E[t_{ib}]\lambda_i}{E[t_{ia}] + E[t_{ib}]}. \qquad (12)$$

However, (12) no longer represents throughput, but rather the efficiency η_i at which packets are successfully transmitted during the cycles captured by station i. As packets are successfully transmitted in proportions λ_i/λ, the throughput of station i can be evaluated as:
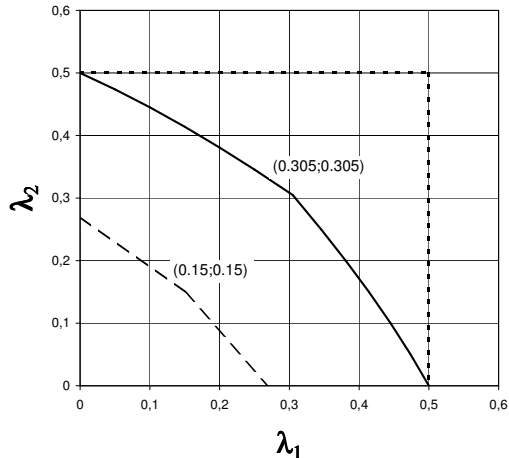
Figure 5: *Capacity curve and upper bound to the stability region (below) and lower bound to the instability region (above) provided in [5] in the BEB case.*

$$\lambda_i = \frac{E[t_{1a}] + E[t_{1b}]\lambda_i}{E[t_{1a}] + E[t_{1b}]} \; \frac{\lambda_i}{\lambda_1 + \lambda_2}, \qquad i = 1, 2 \qquad (13)$$

In Figure 5 we have reported the capacity curve obtained by (13) in the BEB case, together with the upper bound to the stability region and the lower bound to the instability region provided in [5].

As a further generalization, observe that (7) holds true even with Poisson arrivals. In fact, during period t_b multiple Poisson arrivals in a slot at queue 1 are queued so that the output consists of bursts of packets of finite average length that succeed in t_b with average $\lambda/2$. Owing to the great extent of t_b, collision probability when station 2 transmit is still $\lambda/2$.

4. CONCLUDING REMARKS

We have studied stability and capacity of the Binary Exponential Backoff (BEB) in the case of two contending stations, and we have proven that, in this scenario, BEB is stable for total arrival rates up to $\lambda = 0.6096$, which represents also the capacity of the system.

5. REFERENCES

[1] H. Al-Ammal, L. A. Goldberg, and P. MacKenzie. An improved stability bound for binary exponential backoff. *Theory Comput. Syst.*, 30(3):229–244, Dec. 2001.

[2] D. J. Aldous. Ultimate instability of exponential back-off protocol for acknowledgment-based transmission control of random access communication channels. *IEEE Trans. Inform. Theory*, 15(2):219–223, Mar. 1987.

[3] G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *Journ. on Select. Areas Commun.*, 18(3):535–547, March 2000.

[4] J. Hastad, T. Leighton, and B. Rogoff. Analysis of backoff protocols for multiple access channels. *SIAM Jour. on Comput.*, 25(4):740–744, 1996.

[5] J.Goodman, A. G. Greenberg, N. Madras, and P. March. Stability of binary exponential backoff. *Jour. of ACM*, 35(3):579–602, July 1988.

[6] B.-J. Kwak, N.-O. Song, and L. Miller. Performance analysis of exponential backoff. *IEEE/ACM Trans. on Net.*, 13(2):343–355, April 2005.

APPENDIX

Referring to the Markov Chain $J(n)$ that represents the backoff index J at time n, we adopt the following notation

$$\pi_j(n) = P(J(n) = j)$$

and

$$P(j, k, n, n+1) = P(J(n+1) = k/J(n) = j).$$

By the Bayes formula, we have:

$$P(j, j, n+1, n) = P(j, j, n, n+1)\frac{\pi_j(n)}{\pi_j(n+1)}. \qquad (14)$$

Referring to the large-queue conditions we have discussed in the previous section, when n is large and state j reaches very high values collisions, state changes are very rare so that we have

$$\frac{\pi_j(n)}{\pi_j(n+1)} \simeq 1. \qquad (15)$$

and from (14) we have:

$$P(j, j, n+1, n) = P(j, j, n, n+1). \qquad (16)$$

We may thus conclude that the probability of leaving state j is the same in the two directions of the time axis and, therefore, random variable X presents the same probability distribution as Y (geometric).