Analysis of random-access MAC schemes

M. Veeraraghavan and Tao Li

April 9, 2004

(title change; dropped “stable” in (27) and added sentences in red)


First-order analysis: if we assume there are infinite number of nodes, the number of new arrivals in a slot is a Poisson r.v. with parameter $\lambda$. If retransmissions arrive in a highly random fashion, then the total number of transmissions plus retransmissions in a slot, $X$, is a Poisson r.v. with parameter $G$, where $G > \lambda$. The probability of a successful transmission $S$ is given by

$$S = Ge^{-G}$$

(1)

because a successful transmission occurs if only one transmission or retransmission is attempted, i.e. $X = 1$. Slot time is 1. Hence (1). While this analysis shows that the maximum throughput (departure rate) occurs at $G = 1$ because

$$\frac{dS}{dG} = e^{-G} - Ge^{-G} = 0,$$

(2)

and the value of $S$ at this point is $e^{-1} = 0.368$. There are two points at which arrival rate $\lambda$ will equal departure rate $S$. Are both of these stable points? This analysis does not give us any insights into the dynamics of the system with a finite number of hosts. Below is such an analysis.

[From [4], page 279: Let $n$ be the number of hosts with data to send (called “backlogged nodes”). $q_r$ is the probability that such a backlogged node will transmit a packet during the next slot. If $m$ is the total number of nodes, this means one or more of the remaining $m - n$ nodes

![Diagram](attachment:image.png)

Figure 1: DTMC model for slotted ALOHA: state is number of backlogged nodes)
transmit on the next slot if there is a new packet arrival (from higher layer). Assume independence of packet generation among all the nodes. We also assume there is no buffering. If a packet is waiting at a backlogged node, new arrivals are discarded and will need to be “generated” as new arrivals in a later slot. Let \( \lambda \) be the total packet arrival rate. Therefore, per node, it is \( \lambda / m \).

The probability that there are no new arrivals is \( e^{-\lambda/m} \), assuming each slot time is 1. Thus \( q_a = 1 - e^{-\lambda/m} \) is the probability that an unbacklogged node will transmit a packet in a slot.

Probability that \( i \) unbacklogged nodes transmit packets in a given slot is:

\[
Q_a(i, n) = \binom{m-n}{i} (1-q_a)^{m-n-i} q_a^i
\]  

(3)

Probability that \( i \) backlogged nodes transmit packets in a given slot is:

\[
Q_r(i, n) = \binom{n}{i} (1-q_r)^{n-i} q_r^i
\]  

(4)

Reasoning through when a packet is successfully transmitted, we can derive the transition probabilities for the DTMC shown in

\[
P_{n, n+i} = \begin{cases} 
Q_a(i, n) & 2 \leq i \leq (m-n) \\
Q_a(1, n)[1 - Q_r(0, n)] & i = 1 \\
Q_a(1, n)Q_r(0, n) + Q_a(0, n)[1 - Q_r(1, n)] & i = 0 \\
Q_a(0, n)Q_r(1, n) & i = -1 
\end{cases}
\]  

(5)

If there are no new arrivals and only one retransmission attempt, it is successful and hence the state decreases by 1 (last term). If there are \( i \) new arrivals with \( i \geq 2 \), then there will not be any successful transmissions because of collisions. Therefore the state increases by \( i \) (first term). If there is 1 arrival and at least one retransmission, then the number of backlogged nodes increases by 1 and there is no successful transmission (second term). If there are no retransmission attempts and 1 new arrival, then that arrival gets through and so the number of backlogged nodes stays unchanged (first part of the third term). Similarly if there are no new arrivals and only either no retransmissions or at least two retransmissions, then number of backlogged nodes stays unchanged and there is no successful transmission (second part of the third term).

To understand the impact of \( q_r \) and the number of nodes \( m \), consider the following:
Probability of success: \( P_{\text{succ}} = Q_a(1, n)Q_r(0, n) + Q_a(0, n)Q_r(1, n) \) \hspace{1cm} (6)

Drift (expected change in backlog over one slot time) \( D_n = (m - n)q_a - P_{\text{succ}} \) \hspace{1cm} (7)

Define attempt rate \( G(n) \) as the expected number of attempted transmissions in a slot when the system is in state \( n \):

\[
G(n) = (m - n)q_a + nq_r
\]

\[
P_{\text{succ}} = Q_a(1, n)Q_r(0, n) + Q_a(0, n)Q_r(1, n)
\]

\[
P_{\text{succ}} = \left( \binom{m-n}{1}(1-q_a)^{m-n-1}q_a \right) \times \binom{n}{0}((1-q_r)^n0q_r) + \\
\left( \binom{m-n}{0}(1-q_a)^{m-n}0q_a \right) \times \binom{n}{1}((1-q_r)^n-1q_r) + \\
(m-n)(1-q_a)^{m-n-1}q_a(1-q_r)^n + n(1-q_a)^{m-n}(1-q_r)^n-1q_r
\]

(11)

If \( q_a \) and \( q_r \) are small, we can use the approximation:

\[
(1-x)^y = e^{-xy}, \text{ when } x \text{ is small (shouldn’t we require } y \text{ to be large?)}
\]

\[
P_{\text{succ}} = (m - n)q_a e^{-q_a(m-n-1)} e^{-q_n} + nq_r e^{-q_a(m-n)} e^{-q_r(n-1)}
\]

\[
P_{\text{succ}} = (m-n)q_a e^{-[q_a(m-n) + q,n]} e^{q_a} + nq_r e^{-[q_a(m-n) + q,n]} e^{q_r}
\]

(14)

If \( q_a \) and \( q_r \) are small, \( e^{q_a} \) and \( e^{q_r} \) are approximated to 1, and

\[
P_{\text{succ}} \approx G(n) e^{-G(n)} \text{ (following from (8)).}
\]

See Fig. 2. x-axis: state \( n \); also the attempt rate \( G(n) \), for some fixed value of \( m \), \( \lambda \) and \( q_r \). y-axis is \( P_{\text{succ}} \) for the curved plot. The linear plot is \( (m - n)q_a \), the arrival rate. If arrival rate in a slot is equal to \( P_{\text{succ}} \), then drift is 0. There will be no change in the backlog. If \( n \) is to the left of the unstable point, it means \( P_{\text{succ}} > (m - n)q_a \), which means backlog will not build up. Beyond unstable point, the opposite is true, i.e., \( P_{\text{succ}} < (m - n)q_a \), which means backlog will build up. To the right of the unstable point, since \( D \) is positive, it will drift toward the undesired stable point after which \( P_{\text{succ}} > (m - n)q_a \), but at this operating point, \( P_{\text{succ}} \) is low, which means throughput is low. Reason that two of the crosspoints are listed as “stable” is because drift is the
expected change in state from one slot to another and hence the system tends to move in the direction of the drift. In the two points marked stable, drift is decreasing, while at the point marked “unstable” drift increases. The “desired/undesired” characteristic is obvious.

What value of $q_r$ and arrival rate $\lambda$ should be used for stable behavior for a given $m$?

To achieve stability (what is stability - under no buffering assumption, if a packet arrives when one is already in backlog queue at a host, it is simply thrown away, there is a steady-state behavior for all arrival rates (all?)), one approach is to run the system at $G(n) = 1$ because at this value, $P_{succ}$ is maximized. Change $q_r$ dynamically as $n$ changes. If there is a collision, decrease $q_r$, if there is an idle slot, increase $q_r$. This is just like rate adjustment in TCP. As stated on page 281 in [4], as $q_r$ is increased, the delay in retransmitting a collided packet decreases. Manipulating $q_r$ is nothing but changing the sending rate!

Measuring $n$ and changing $q_r$ is challenging (much like figuring out ideal sending rate in TCP), and hence many ALOHA-stabilizing algorithms have been proposed. The binary exponential backoff algorithm in Ethernet is a stabilization technique. Splitting algorithms constitute another technique, where if a collision occurs in the $k^{th}$ slot, among all nodes involved in the collision, only some attempt retransmission on the $(k + 1)^{th}$ slot (approx. half - determined by each
node by flipping a coin) and ones not involved in the collision do not enter the fray. If the attempt
was successful, the second subset gets to try on the \((k + 2)^{th}\) slot. If was not successful, then the
first subset splits again.

2. **Analysis of CSMA Slotted Aloha [4]**

Let \(\tau\) be the max. propagation delay and detection delay (time required to determine if some
other node is currently transmitting) - i.e., time by which all nodes detect an idle channel after a
transmission ends. If \(C\) is the bit rate and \(L\) is the average packet size, then

\[
\beta = \frac{C\tau}{L},
\]

where \(\beta\) is equivalent to \(\tau\) expressed in packet transmission units. Idle slots in CSMA have a
duration \(\beta\). Packets arriving during an idle slot wait for the start of the next slot and are transmit-
ted immediately. A packet arriving when a transmission is ongoing is marked as a “backlogged”
and is transmitted with probability \(q_r\) after each subsequent idle slot. In mac.ppt, we referred to
such a scheme as non-persistent CSMA, stating that when a node senses the medium (which it
does when it has something to send), and finds the medium busy, it waits for some random time
before sensing the medium again.

Use Markov chain to analyze CSMA Slotted Aloha. It is a DTMC with the discrete time unit
set to \(\beta\). Assume all packets are of unit (fixed) length. Can extend to case when this is not true. \(\beta\)
will be less than 1, which is the time to send a data packet. Note that every busy slot (success or
collision) has to be followed by an idle slot for the nodes to detect that the medium is free. The
system changes state only at the end of idle slots.
Consider a sample function as shown in Fig. 3. Consider the time from the end of one idle slot to the next. This time can be $\beta$ or it can be $1 + \beta$ because after the busy period, we necessarily have an idle slot. Call this interval between the ends of two idle slots $X$.

$$P[X = \beta] = (1 - q_r)^n e^{-\lambda \beta}$$

$$P[X = 1 + \beta] = 1 - (1 - q_r)^n e^{-\lambda \beta}$$

This is because the probability of no transmissions in a $\beta$ idle slot is $(1 - q_r)^n e^{-\lambda \beta}$, where $(1 - q_r)^n$ is the probability that none of the backlogged nodes transmit and $e^{-\lambda \beta}$ is that there were no new arrivals in the previous slot (because if there were new arrivals, they are simply sent in non-persistent CSMA). Set $p = (1 - q_r)^n e^{-\lambda \beta}$.

Therefore, the average value of $X$, the time between the end of 1 idle slot to the end of the next idle slot, is $\beta + 1 - (1 - q_r)^n e^{-\lambda \beta}$:

$$E[X] = \beta p + (1 + \beta)(1 - p) = \beta + 1 - p.$$  \hspace{1cm} (19)

The expected number of arrivals within time $X$ is:

$$E[\text{arrivals}] = \lambda[\beta + 1 - (1 - q_r)^n e^{-\lambda \beta}]$$

The expected number of departures within the same interval is the probability of a successful transmission:

Using the logic of (6): $P_{\text{succ}} = Q_a (1, n) Q_r (0, n) + Q_a (0, n) Q_r (1, n)$

$$P_{\text{succ}} = \lambda \beta e^{-\lambda \beta} (1 - q_r)^n + e^{-\lambda \beta} \binom{n}{1} q_r (1 - q_r)^{n-1} = e^{-\lambda \beta} (1 - q_r)^n \left( \lambda \beta + \frac{nq_r}{1 - q_r} \right)$$

\hspace{1cm} (22)
Drift in state \( n \) is the expected number of arrivals within this interval \( X \) minus the expected number of departures within the same interval:

\[
D_n = \lambda [\beta + 1 - (1 - q_r)^n e^{-\lambda \beta}] - e^{-\lambda \beta} (1 - q_r)^n \left( \lambda \beta + \frac{n q_r}{1 - q_r} \right)
\]  
(23)

For small \( q_r \), we approximate \( (1 - q_r)^n \approx 1 - q_r^n \) (see Appendix).

\[
D_n \approx \lambda [\beta + 1 - e^{-q_r^n} e^{-\lambda \beta}] - e^{-\lambda \beta} e^{-q_r^n} (\lambda \beta + \frac{n q_r}{1 - q_r})
\]  
(24)

\[
D_n \approx \lambda [\beta + 1 - e^{-g(n)}] - e^{-g(n)} (\lambda \beta + n q_r) = \lambda [\beta + 1 - e^{-g(n)}] - g(n) e^{-g(n)}, \text{ where}
\]  
(25)

\[g(n) = \lambda \beta + q_r n : \text{mean number of attempted transmissions after a transition into state } n\]  
(26)

If the call arrival rate \( \lambda < \frac{g(n) e^{-g(n)}}{[\beta + 1 - e^{-g(n)}]} \), drift is negative

\[
D_n = \lambda \beta + 1 - e^{-g(n)} (\lambda \beta + n q_r) = \lambda \beta + 1 - e^{-g(n)} - g(n) e^{-g(n)}, \text{ where}
\]  
(27)

The ratio in (27) is the mean number of departures divided by mean duration of a state holding time. Hence it is the departure rate in state \( n \). Plot this ratio as a function of \( g \). Max. point occurs when \( g = \sqrt{2 \beta} \), at which point the departure rate can be approximated to \( 1 / (1 + \sqrt{2 \beta}) \) for small \( \beta \). For small \( \beta \), little time is wasted in idle slots. Most of the waste comes from collisions. The rhs of (27) becomes:

\[
\frac{\sqrt{2 \beta} e^{-\sqrt{2 \beta}}}{\beta + 1 - e^{-\sqrt{2 \beta}}} = \frac{\sqrt{2 \beta}}{\beta + 1 - e^{-\sqrt{2 \beta}} - 1}
\]  
(28)

Approximate \( e^{-\sqrt{2 \beta}} \approx 1 - \sqrt{2 \beta} + (\sqrt{2 \beta})^2 / 2 \):

\[
\frac{\sqrt{2 \beta}}{(\beta + 1)e^{-\sqrt{2 \beta}} - 1} = \frac{\sqrt{2 \beta}}{(\beta + 1)(1 + \sqrt{2 \beta} + \beta) - 1} = \frac{\sqrt{2 \beta}}{(\beta + 1) + \beta \sqrt{2 \beta} + \sqrt{2 \beta} + \beta^2 + \beta - 1}
\]  
(29)

\[
= \frac{\sqrt{2 \beta}}{2\beta + \beta \sqrt{2 \beta} + \beta^2 + \sqrt{2 \beta}} = \frac{1}{1 + \sqrt{2 \beta}} \text{ (ignoring the } 3/2 \text{ and square power terms of } \beta \text{).}
\]  
(30)

\( \beta / q_r \) is the expected idle time that a backlogged node must wait before attempting retransmission (because the number of idle slots waited is a geometric r.v. with parameter \( q_r \) and mean of a geometric r.v with parameter \( p \) is \( 1/p \)). If \( \beta \) is small and \( \lambda \) is modest, \( q_r \) can be small and yet
main a small total delay. This means unstability will not set in too easily in a CSMA system unlike in slotted Aloha. For large $q_r$, one can expect unstability because backlog will keep increasing.

The arrival rate is independent of $n$ here because unlike in the slotted Aloha analysis, here we assume an infinite number of nodes, and $\lambda$ as the aggregate arrival rate.

3. Analysis of Ethernet (CSMA/CD)

Additional: stop transmission if a collision is detected. Think of time being divided into slots and minislots, where a minislot is of duration $\beta$, the time (expressed in packet transmission units) for propagation and detection. If only one node transmits within a minislot, then its whole packet gets through. If two transmit, then they both detect the collision by the end of the minislot and stop transmission. Thus, there is contention in the minislots, but once a node gets through the minislot, it is home free for the rest of the transmission of its packet. So [4] says it is as though a host is “effectively reserving the channel for the completion of the packet,” and thus the minislot can be viewed as a “signaling” phase for the resource reservation!

Analyze CSMA/CD with a Markov chain [9], pg. 444. Time taken for a collision detection is $2\tau$ (see mac.ppt).

![Figure 4:Total transmission time to send a frame in CSMA/CD](image)

$$t_v = \frac{L}{C} + \tau + 2\tau J = \frac{L}{C}[1 + \beta(1 + 2J)], \quad (31)$$

where $J$ is the average number of retransmissions.

The mean length of the collision interval can be determined because the number of collisions is a geometric random variable with parameter $\nu$. Each collision takes $2\tau$ sec. Mean of a geometric r.v. is $1/p$ if $p$ is the parameter of the r.v. Hence:

$$J = \sum_{k=1}^{\infty} k\nu(1-\nu)^{k-1} = \frac{1}{\nu} \quad (32)$$
\[ v = np(1-p)^{n-1}, \]  
where \( p \) is the probability that exactly one station transmits and is hence successful; \( n \) is the number of stations. The value of \( p \) at which \( v \) is maximized:

\[
\frac{dv}{dp} = n(1-p)^{n-1} - np(n-1)(1-p)^{n-2} = 0 \text{ or } p = \frac{1}{n}
\]  

or (34)

\[
v_{\text{max}} = \left(1 - \frac{1}{n}\right)^{n-1} \rightarrow e^{-1} \quad \text{as } n \rightarrow \infty \quad \text{(35)}
\]

Using this value of \( v \) in (32), and the corresponding \( J \) in (31), we get:

\[
t_v = \frac{L}{C}[1 + \beta(1 + 2e)] \quad \text{(36)}
\]

This means for a stable system, the packet arrival rate \( \lambda \) should be:

\[
\lambda < \frac{1}{L[C(1+\beta(1+2e))]}, \quad \text{or} \quad \lambda < \frac{1}{L[C(1+6.44\beta)]} \quad \text{using the numerical value of } e. \quad \text{(37)}
\]

An alternative way of phrasing this is to say the maximum throughput \( \rho_{\text{max}} = \frac{1}{[1 + 6.44\beta]} \). If \( \beta = 0.1 \frac{L}{C} \), i.e., the time for propagation + detection is one-tenth of the frame transmission time, \( \rho_{\text{max}} = 0.6 \), much better than Slotted Aloha (which is 0.368). How about CSMA? It is 0.69! Wait - is CSMA/CD worse than CSMA?

Reference [4], page 317-319 describes Ethernet as unslotted CSMA/CD because a second node at the far-end can start transmitting at a short time before the first node’s signal gets to it causing us to use a collision detection time of \( 2\tau \) in our analysis. If it is slotted CSMA/CD, the second node should have started at the same time as the first node, which means its signal will reach the first node within \( \tau \) causing the collision detection time to be only \( \tau \). Result is that in slotted CSMA/CD,

\[
\rho_{\text{max}} = \frac{1}{[1 + 3.31\beta]} \quad \text{(38)}
\]

Comparing (38) with (30),

\[
3.31\beta < \sqrt{2}\beta \text{ if } \beta < 0.18, \quad \text{(39)}
\]
which means slotted CSMA/CD is better than slotted CSMA only if \( \beta < 0.18 \)? Not true; the max. throughput \( \frac{1}{1 + \sqrt{2\beta}} \) is an approximation that only hold when \( \beta \) is very small. As the matlab files show, CSMA/CD outperforms CSMA.

Point to note: CSMA and CSMA/CD become inefficient if \( \beta \) becomes a significant factor compared to packet transmission time. Increasing bus length, increasing data rate or decreasing packet size will all cause \( \beta \) to increase. Neither is a good choice if \( \beta \) is more than a few tenths is what [4] says. Consider TCP. It too becomes inefficient as MSS increases, rate increases or \( T_{prop} \) increases!

Above derivation just finds the maximum throughput possible. Reference [4], page 318 gives the equation for throughput as a function of \( n \) as with Slotted CSMA and Slotted Aloha. Drift in state \( n \) will be negative if:

\[
\lambda < \frac{g(n)e^{-g(n)}}{\beta + g(n)e^{-g(n)} + \beta[1 - (1 + g(n))e^{-g(n)}]} \quad \text{(see [4] for derivation)} \\
\]

\[
g(n) = \lambda \beta + q_r n \quad \text{(same as with slotted CSMA)} \\
\]

4. Appendix

4.1 Equation (12)

\[
(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \quad \text{and} \quad e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots
\]

\[
(1 - x)^y = 1 - xy + \frac{y(y-1)}{2!} x^2 - \frac{y(y-1)(y-2)}{3!} x^3 + \ldots
\]

\[
e^{-xy} = 1 - xy + \frac{(xy)^2}{2!} - \frac{(xy)^3}{3!} + \ldots
\]

4.2 Compare with my ppt file definitions of non-persistent, 1-persistent, p-persistent

We discussed the relation between our definition of non-persistent in mac.ppt and the [4] definition. With 1-persistent, [4] states all arrivals during a busy slot will start transmission at the end of the slot, which makes the probability of collisions rather high. In mac.ppt, we state that it keeps sensing and sends immediately upon recognizing an idle medium. [4] states that in p-persistent CSMA, new and collided packets waiting for the end of a busy period using different probabilities
for transmission on the first idle slot after the busy period. mac.ppt didn’t quite say this - it says if the channel is idle, transmit with probability $p$.

References