A bound on the capacity of backoff and acknowledgement-based protocols*

Leslie Ann Goldberg †
Dep't of Computer Science
University of Warwick

Mark Jerrum‡
Division of Informatics
University of Edinburgh

Sampath Kannan§
Dep't of Comp. and Inf. Sci.
Univ. of Pennsylvania

Mike Paterson¶
Dep't of Computer Science
University of Warwick

December 1, 2000

Abstract

We study contention-resolution protocols for multiple-access channels. We show that *every* backoff protocol is transient if the arrival rate, λ, is at least 0.42 and that the capacity of every backoff protocol is at most 0.42. Thus, we show that backoff protocols have (provably) smaller capacity than full-sensing protocols. Finally, we show that the corresponding results, with the larger arrival bound of 0.531, also hold for every acknowledgement-based protocol.

*This work was partially supported by the EPSRC grant "Sharper Analysis of Randomised Algorithms: a Computational Approach", NSF Grant CCR9820885, and the IST Programme of the EU under contract numbers IST-1999-14186 (ALCOM-FT) and IST-1999-14036 (RAND-APX).

†`leslie@dcs.warwick.ac.uk`, `http://www.dcs.warwick.ac.uk/~leslie/`, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

‡`mrj@dcs.ed.ac.uk`, `http://www.dcs.ed.ac.uk/~mrj/`, Division of Informatics, University of Edinburgh, JCMB, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom.

§`kannan@cis.upenn.edu`, `http://www.cis.upenn.edu/~kannan/`, Department of Computer and Information Science, University of Pennsylvania, 200 south 33rd st, Philadelphia, PA 19104-6389, USA

¶`msp@dcs.warwick.ac.uk`, `http://www.dcs.warwick.ac.uk/~msp/`, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

1 Introduction

A *multiple-access channel* is a broadcast channel that allows multiple users to communicate with each other by sending messages onto the channel. If two or more users simultaneously send messages, then the messages interfere with each other (collide), and the messages are not transmitted successfully. The channel is not centrally controlled. Instead, the users use a contention-resolution protocol to resolve collisions. Thus, after a collision, each user involved in the collision waits a random amount of time (which is determined by the protocol) before re-sending.

Following previous work on multiple-access channels, we work in a *time-slotted* model in which time is partitioned into discrete *time steps.* At the beginning of each time step, a random number of messages enter the system, each of which is associated with a new user which has no other messages to send. The number of messages that enter the system is drawn from a Poisson distribution with mean λ. During each time step, each message chooses independently whether to send to the channel. If *exactly* one message sends to the channel during the time step, then this message leaves the system and we call this a *success.* Otherwise, all of the messages remain in the system and the next time step is started. Note that when a message sends to the channel this may or may not result in a success, depending on whether or not any other messages send to the channel.

The quality of a protocol can be measured in several ways. Typically, one models the execution of the protocol as a Markov chain. If the protocol is good (for a given arrival rate λ), the corresponding Markov chain will be *recurrent* (with probability 1, it will eventually return to the empty state in which no messages are waiting). Otherwise, the chain is said to be *transient* (and we also say that a protocol is transient). Note that transience is a very strong form of instability. In particular, if we focus on any finite set of "good" states then if the chain is transient, the probability of visiting these states at least N times during the infinite run of the protocol is exponentially small in N. (This follows because the relevant Markov chain is irreducible and aperiodic.)

Another way to measure the quality of a protocol is to measure its *capacity.* A protocol is said to achieve *throughput* λ if, when it is run with input rate λ, the average success rate is λ. The *capacity* of the protocol [4] is the maximum throughput that it achieves.

The protocols that we consider in this paper are *acknowledgement-based* protocols. In the acknowledgement-based model, the only information that a user receives about the state of the system is the history of its own transmissions. An alternative model is the *full-sensing* model, in which *every* user listens to the channel at *every* step, regardless of whether it sends during the step.[1]

One particularly simple and easy-to-implement class of acknowledgement-based protocols is the class of *backoff protocols.* A backoff protocol is a sequence of probabilities $p_0, p_1, \ldots$. If a message has sent unsuccessfully i times before a time-step then, with probability p_i, it sends during the time-step. Otherwise, it does not send. Kelly and MacPhee [12, 13, 16] gave a formula for the *critical arrival rate,* λ^*, of a backoff protocol, which is the minimum arrival rate for which the expected number of successful transmissions that the protocol makes is finite.[2]

Perhaps the best-known backoff protocol is the *binary exponential backoff protocol* in which $p_i = 2^{-i}$. This protocol is the basis of the Ethernet protocol of Metcalfe and Boggs [17].[3] Kelly and MacPhee showed that the critical arrival rate of this protocol is $\ln 2$. Thus, if $\lambda > \ln 2$, then binary exponential backoff achieves only a finite number of successful transmissions (in expectation). Aldous [1] showed that the binary exponential backoff protocol is not a good protocol for *any* positive arrival rate λ. In particular, it is transient and the expected number of successful transmissions in t steps is $o(t)$. MacPhee [16] posed the question of whether there exists a backoff protocol which is recurrent for some positive arrival rate λ.

In this paper, we show that there is no backoff protocol which is recurrent for $\lambda \geq 0.42$. (Thus, *every* backoff protocol is transient if $\lambda \geq 0.42$.) Also, every backoff protocol has capacity at most 0.42. As far as we know, our result is the first proof showing that backoff protocols have smaller capacity than full-sensing protocols. In particular, Mosely and Humblet [19] have discovered a full-sensing protocol with capacity 0.48776.[4] Finally, we show

[1] In practice, it is possible to implement the full-sensing model when there is a single channel, but this becomes increasingly difficult in situations where there are multiple shared channels, such as optical networks. Thus, acknowledgement-based protocols are sometimes preferable to full-sensing protocols. For work on contention-resolution in the multiple-channel setting, see [6].

[2] If $\lambda > \lambda^*$, then the expected number of successes is finite, even if the protocol runs forever. They showed that the critical arrival rate is 0 if the expected number of times that a message sends during the first t steps is $\omega(\log t)$.

[3] There are several differences between the "real-life" Ethernet protocol and "pure" binary exponential backoff, but we do not describe these here.

[4] Mosely and Humblet's protocol is a "tree protocol" in the sense of Capetanakis [3] and

that *no* acknowledgement-based protocol is recurrent for $\lambda \geq 0.530045$.

1.1 Related work

Backoff protocols and acknowledgement-based protocols have also been studied in an n-user model, which combines contention-resolution with queueing. In this model, it is assumed that n users maintain queues of messages, and that new messages arrive at the tails of the queues. At each step, the users use contention-resolution protocols to try to send the messages at the heads of their queues. It turns out that the queues have a stabilising effect, so some protocols (such as "polynomial backoff") which are unstable in our model [13] are stable in the queueing model [11]. We will not describe queueing-model results here, but the reader is referred to [2, 8, 11, 21].

Much work has gone into determining upper bounds on the capacity that can be achieved by a full-sensing protocol. The current best result is due to Tsybakov and Likhanov [23] who have shown that no protocol can achieve capacity higher than 0.568. (For more information, see [4, 9, 18, 22].) In the full-sensing model, one typically assumes that messages are born at real "times" which are chosen uniformly from the unit interval. Recently, Loher [14, 15] has shown that if a protocol is required to respect these birth times, in the sense that packets must be successfully delivered in their birth order, then no protocol can achieve capacity higher than 0.4906. Intuitively, the "first-come-first-served" restriction seems very strong, so it is somewhat surprising that the best known algorithm without the restriction (that of Vvedenskaya and Pinsker) does not beat this upper bound. The algorithm of Humblet and Mosely satisfies the first-come-first-served restriction.

1.2 Improvements

We choose $\lambda = 0.42$ in order to make the proof of Lemma 10 (see the Appendix) as simple as possible. The lemma seems to be true for λ down to about 0.41 and presumably the parameters A and B could be tweaked to get λ slightly smaller.

Tsybakov and Mikhailov [24]. For a simple analysis of the protocol, see [25]. Vvedenskaya and Pinsker have shown how to modify Mosely and Humblet's protocol to achieve an improvement in the capacity (in the seventh decimal place) [26].

2 Markov Chain Background

An irreducible aperiodic Markov chain $X = \{X_0, X_1, \ldots\}$ with a countable state space Ω (see [10]) is *recurrent* if it returns to its start state with probability 1. That is, it is recurrent if for some state i (and therefore, for all i), $\mathrm{Prob}[X_t = i \text{ for some } t \geq 1 \mid X_0 = i] = 1$. Otherwise, X is said to be *transient.* X is *positive recurrent* (or ergodic) if the expected number of steps that it takes before returning to its start state is finite. We use the following theorems which we take from [5].

Theorem 1 *(Fayolle, Malyshev, Menshikov) An irreducible aperiodic time-homogeneous Markov chain X with countable state space Ω is not positive recurrent if there is a function f with domain Ω and there are constants C and d such that*

1. *there is a state x with $f(x) > C$, and a state x with $f(x) \leq C$, and*
2. *$E[f(X_1) - f(X_0) \mid X_0 = x] \geq 0$ for all x with $f(x) > C$, and*
3. *$E[\,|f(X_1) - f(X_0)| \mid X_0 = x] \leq d$ for every state x.*

Theorem 2 *(Fayolle, Malyshev, Menshikov) An irreducible aperiodic time-homogeneous Markov chain X with countable state space Ω is transient if there is a positive function f with domain Ω and there are positive constants C, d and ε such that*

1. *there is a state x with $f(x) > C$, and a state x with $f(x) \leq C$, and*
2. *$E[f(X_1) - f(X_0) \mid X_0 = x] \geq \varepsilon$ for all x with $f(x) > C$, and*
3. *if $|f(x) - f(y)| > d$ then the probability of moving from x to y in a single move is 0.*

3 Stochastic Domination and Monotonicity

Suppose that X is a Markov chain and that the (countable) state space Ω of the chain is a *partial order* with binary relation $\leq$. If A and B are random variables taking states as values, then B *dominates* A if and only if there is a joint sample space for A and B in which the value of A is always less than or equal to the value of B. Note that there will generally be other joint sample spaces in which the value of A can exceed the value of B. We write $A \leq B$ to indicate that B dominates A. We say that X is *monotonic* if for any states $x \leq x'$, the next state conditioned on starting at x' dominates the next state conditioned on starting at x. (Formally, $(X_1 \mid X_0 = x')$ dominates $(X_1 \mid X_0 = x)$.)

When an acknowledgement-based protocol is viewed as a Markov chain, the state is just the collection of messages in the system. (Each message is identified by the history of its transmissions.) Thus, the state space is countable and it forms a partial order with respect to the subset inclusion relation $\subseteq$ (for multisets). We say that a protocol is *deletion resilient* [7] if its Markov chain is monotonic with respect to the subset-inclusion partial order.

Observation 3 *Every acknowledgement-based protocol is deletion resilient.*

Proof: Consider the states x and x' with $x \subseteq x'$. Recall that each state is a set of messages, each message being identified by its transmission history. Thus, x' contains all of the messages in x and possibly others. Now consider one step of the protocol. We wish to show that the random variable denoting the next state $z' = (X_1 \mid X_0 = x')$ dominates the random variable $z = (X_1 \mid X_0 = x)$. z' does dominate z because we can draw z and z' from a joint sample space in which

- the messages in x do the same thing in both copies, and
- both copies have the same number of new arrivals, which make the same number of send attempts in both copies.

Now consider any message m which is either a new arrival or a member of x.

1. If m is silent during the step then its transmission history in z' is the same as in z.
2. If m has a collision during the transition to z then it also has a collision during the transition to z', so its transmission history in z' is the same as in z.

Thus, $z \subseteq z'$. $\square$

As we indicated earlier, we will generally assume that the number of messages entering the system at a given step is drawn from a Poisson process with mean λ. However, it will sometimes be useful to consider other message-arrival distributions. If I and I' are message-arrival distributions, we write $I \leq I'$ to indicate that the number of messages generated under I is dominated by the number of messages generated under I'.

Observation 4 *If the acknowledgement-based protocol P is recurrent under the message-arrival distribution I' and $I \leq I'$ then P is also recurrent under I.*

Proof: Let X be the Markov chain corresponding to protocol P with arrival distribution I with X_0 as the empty state. Let X' be the analogous Markov chain with arrival distribution I'. Consider the evolution of the stochastic process $(X_0, X'_0), (X_1, X'_1), \ldots$. We will choose the random variable (X_i, X'_i) from a joint probability distribution in which

1. every message which is common to X_i and X'_i does the same thing in both copies, and
2. the new arrivals which are drawn from I arrive in both copies, and make the same number of send attempts in both copies.
3. Some additional messages may arrive in X'_i (according to I').

Note that items (2) and (3) are possible since $I \leq I'$. Now we can now show by induction on t that X'_t dominates X_t. That is, when this joint distribution is used, X_t is a subset of X'_t. This holds for $t = 0$ since $X_0 = X'_0$. The inductive step is the same as the proof of Observation 3. Consider any message m which is in X_t or arrives (according to I) just before step $t+1$.

1. If m is silent during the step then its transmission history in X'_{t+1} is the same as in X_{t+1}.
2. If m has a collision during the transition to X_{t+1} then it also has a collision during the transition to X'_{t+1}.

Thus, $X_{t+1} \subseteq X'_{t+1}$. Finally, since X'_t dominates X_t, the recurrence of X'_t implies the recurrence of X_t. $\square$

4 Backoff protocols

In this section, we will show that there is no backoff protocol which is recurrent for $\lambda \geq 0.42$. Our method will be to use the "drift theorems" in Section 2. Let $p_0, p_1, \ldots$ be a backoff protocol. Without loss of generality, we can assume $p_0 = 1$, since we can ignore new arrivals until they first send. Let $\lambda = 0.42$. Let X be the Markov chain described in Section 3 which describes the behaviour of the protocol with arrival rate λ. First, we will construct a potential function (Lyapounov function) f which satisfies the conditions of Theorem 1, that is, a potential function which has a bounded positive drift. We will use Theorem 1 to conclude that the chain is not positive recurrent. Next, we will consider the behaviour of the protocol under a *truncated* arrival distribution and we will use Theorem 2 to show that the protocol is transient. Using Observation 4 (domination), we will conclude that the protocol is also transient with Poisson arrivals at rate λ or higher. Finally, we will show that the *capacity* of every backoff protocol is at most 0.42.

We will use the following technical lemma.

Lemma 5 *Let $1 \leq t_i \leq d$ for $i \in [1, k]$ and $\prod_{i=1}^{k} t_i = c$. Then $\sum_{i=1}^{k}(t_i - 1) \leq (d-1)\frac{\log c}{\log d}$.*

Proof: Let $S = \sum_{i=1}^{k} t_i$. S can be viewed as a function of $k-1$ of the t_i's, for example $S = \sum_{i=1}^{k-1} t_i + c/\prod_{i=1}^{k-1} t_i$. For $i \in \{1, \ldots, k-1\}$, the derivative of S with respect to t_i is $1 - c/(t_i \prod_{j=1}^{k-1} t_j)$. Thus, the derivative is positive if $t_i > t_k$. Thus, S is maximised (subject to c) by setting some t_i's to 1, some t_i's to d and at most one t_i to some intermediate value $t \in [1, d)$. Given this, the maximum value of $\sum_{i=1}^{k}(t_i - 1)$ is $s(d-1) + t - 1$, where $c = d^s t$ and $s = \lfloor (\log c)/(\log d) \rfloor$. Let α be the fractional part of $(\log c)/(\log d)$, that is, $\alpha = (\log c)/(\log d) - s$. We want to show that $s(d-1) + t - 1 \leq (d-1)(\log c)/(\log d)$. This is true, since

$$\begin{aligned}
(d-1)\frac{\log c}{\log d} - s(d-1) - (t-1) &= \alpha(d-1) - c/d^s + 1 \\
&= \alpha(d-1) - d^\alpha + 1 \\
&\geq \alpha(d-1) - (1 + \alpha(d-1)) + 1 \\
&= 0.
\end{aligned}$$

□

We now define some parameters of a state x. Let $k(x)$ denote the number of messages in state x. If $k(x) = 0$, then $p(x) = r(x) = u(x) = 0$. Otherwise, let $m_1, \ldots, m_{k(x)}$ denote the messages in state x, with send probabilities $\rho_1 \geq \cdots \geq \rho_{k(x)}$. Let $p(x) = \rho_1$ and let $r(x)$ denote the probability that at least one of $m_2, \ldots, m_{k(x)}$ sends on the next step. Let $u(x)$ denote the probability that exactly one of $m_2, \ldots, m_{k(x)}$ sends on the next step. Clearly $u(x) \leq r(x)$. If $p(x) < r(x)$ then we use the following (tighter) upper bound for $u(x)$.

Lemma 6 *If $p(x) < r(x)$ then $u(x) \leq \frac{r(x) - r(x)^2/2}{1 - p(x)/2}$.*

Proof: Fix a state x. We will use $k, p, r, \ldots$ to denote $k(x), p(x), r(x), \ldots$. Since $p < r$, we have $k \geq 2$.

$$u = \sum_{i=2}^{k} \frac{\rho_i}{1 - \rho_i} \prod_{i=2}^{k} (1 - \rho_i) = \sum_{i=2}^{k} (t_i - 1)(1 - r),$$

where $t_i = 1/(1 - \rho_i)$. Let $d = 1/(1 - p)$, and note that $1 \leq t_i \leq d$. By Lemma 5

$$\begin{aligned} u \leq (1 - r)(d - 1) \frac{\log(\prod_{i=2}^{k} t_i)}{\log d} &= (1 - r) \frac{p}{1 - p} \frac{\log(1/(1 - r))}{\log d} \\ &= (1 - r) \frac{p}{1 - p} \frac{(-\log(1 - r))}{(-\log(1 - p))}. \end{aligned}$$

Now we wish to show that

$$(1 - r) \frac{p}{1 - p} \frac{(-\log(1 - r))}{(-\log(1 - p))} \leq \frac{r - r^2/2}{1 - p/2},$$

i.e., that

$$(1 - r) \frac{(-\log(1 - r))}{r - r^2/2} \leq (1 - p) \frac{(-\log(1 - p))}{p - p^2/2}.$$

This is true, since the function $(1 - r)\frac{(-\log(1 - r))}{r - r^2/2}$ is decreasing in r. To see this, note that the derivative of this function with respect to r is $y(r)/(r - r^2/2)^2$, where

$$\begin{aligned} y(r) &= (1 - r + r^2/2) \log(1 - r) + (r - r^2/2) \\ &\leq (1 - r + r^2/2)(-r - r^2/2) + (r - r^2/2) = -r^4/4. \end{aligned}$$

$\square$

Let $\mathcal{S}(x)$ denote the probability that there is a success when the system is run for one step starting in state x. (Recall that a success occurs if *exactly* one message sends during the step. This single sender might be a new arrival, or it might be an old message from state x.) Let

$$g(r,p) = e^{-\lambda}[(1-r)p + (1-p)\min\{r, \frac{r-r^2/2}{1-p/2}\} + (1-p)(1-r)\lambda].$$

We now have the following corollary of Lemma 6.

Corollary 7 *For any state x, $\mathcal{S}(x) \le g(r(x), p(x))$.*

Let $s(x)$ denote the probability that at least one message in state x sends on the next step. That is, $s(x)$ is the probability that at least one *existing* message in x sends. New arrivals may also send. There may or may not be a success. (Thus, if x is the empty state, then $s(x) = 0$.) Let $A = 0.9$ and $B = 0.41$. For every $\pi \in [0,1]$, let $c(\pi) = \max(0, -A\,\pi + B)$. For every state x, let $f(x) = k(x) + c(s(x))$. The function f is the potential function alluded to earlier, which plays a leading role in Theorems 1 and 2. To a first approximation, $f(x)$ counts the number of messages in the state x, but the small correction term is crucial. Finally, let

$$h(r,p) = \lambda - g(r,p) - [1 - e^{-\lambda}(1-p)(1-r)(1+\lambda)]c(r+p-r\,p) + e^{-\lambda}p(1-r)c(r).$$

Now we have the following.

Observation 8 *For any state x, $E[\,|f(X_1) - f(X_0)|\mid X_0 = x] \le 1 + B$.*

Lemma 9 *For any state x, $E[f(X_1) - f(X_0) \mid X_0 = x] \ge h(r(x), p(x))$.*

Proof: The result follows from the following chain of inequalities, each link of which is justified below.

$$\begin{aligned}
E[f(X_1) &- f(X_0) \mid X_0 = x] = \lambda - \mathcal{S}(x) + E[c(s(X_1)) \mid X_0 = x] - c(s(x)) \\
&\ge \lambda - g(r(x), p(x)) + E[c(s(X_1)) \mid X_0 = x] - c(s(x)) \\
&\ge \lambda - g(r(x), p(x)) + e^{-\lambda}(1-p(x))(1-r(x))(1+\lambda)c(s(x)) \\
&\qquad + e^{-\lambda}p(x)(1-r(x))c(r(x)) - c(s(x)) \\
&= h(r(x), p(x)).
\end{aligned}$$

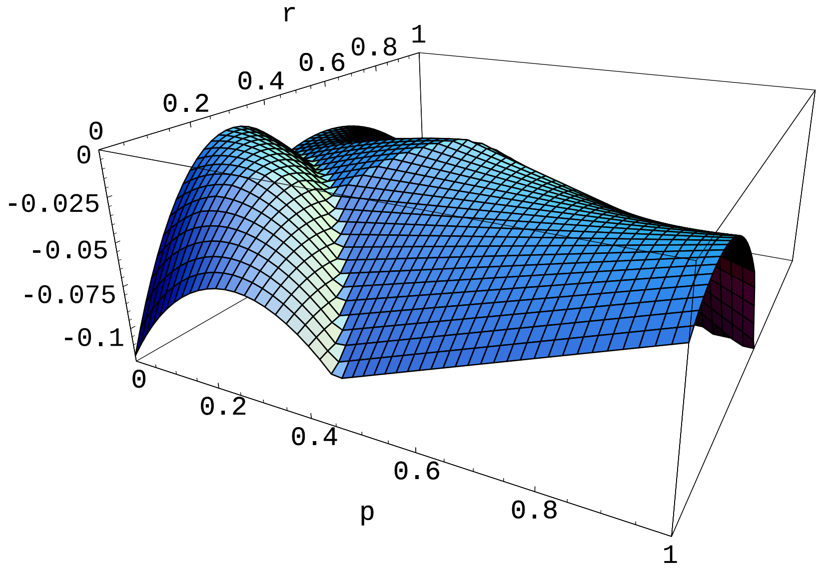

Figure 1: $-h(r,p)$ over the range $r \in [0,1], p \in [0,1]$.

The first inequality follows from Corollary 7. The second comes from substituting exact expressions for $c(s(X_1))$ whenever the form of X_1 allows it, and using the bound $c(s(X_1)) \geq 0$ elsewhere. If none of the existing messages sends and there is at most one arrival, then $c(s(X_1)) = c(s(x))$, giving the third term; if message m_1 alone sends and there are no new arrivals then $c(s(X_1)) = c(r(x))$, giving the fourth term. The final equality uses the fact that $s(x) = p(x) + r(x) - p(x)r(x)$. □

Lemma 10 *For any* $r \in [0,1]$ *and* $p \in [0,1]$, $h(r,p) \geq 0.003$.

Proof: We defer the proof of this lemma to the Appendix of the paper. Figure 1 contains a (Mathematica-produced) plot of $-h(r,p)$ over the range $r \in [0,1], p \in [0,1]$. The plot suggests that $-h(r,p)$ is bounded below zero.

We note here that our proof of the lemma (in the Appendix) involves evaluating certain polynomials at about 40 000 points, and we did this using Mathematica. □

We now have the following theorem.

Theorem 11 *No backoff protocol is positive recurrent when the arrival rate is $\lambda = 0.42$.*

Proof: This follows from Theorem 1, Observation 8 and Lemmas 9 and 10. The value C in Theorem 1 can be taken to be 1 and the value d can be taken to be $1 + B$. □

Now we wish to show that every backoff protocol is transient for $\lambda \geq 0.42$. Once again, fix a backoff protocol $p_0, p_1, \ldots$ with $p_0 = 1$. Notice that our potential function f almost satisfies the conditions in Theorem 2. The main problem is that there is no absolute bound on the amount that f can change in a single step, because the arrivals are drawn from a Poisson distribution. We get around this problem by first considering a *truncated-Poisson* distribution, $T_{M,\lambda}$, in which the probability of r inputs is $e^{-\lambda}\lambda^r/r!$ (as for the Poisson distribution) when $r < M$, but $r = M$ for the remaining probability. By choosing M sufficiently large we can have $E[T_{M,\lambda}]$ arbitrarily close to λ.

Lemma 12 *Every backoff protocol is transient for the input distribution $T_{M,\lambda}$ when $\lambda = 0.42$ and $\lambda' = E[T_{M,\lambda}] > \lambda - 0.001$.*

Proof: The proof is almost identical to that of Theorem 11, except that the first term, λ, in the definition of $h(r, p)$ (for Lemmas 9 and 10) must be replaced by λ'. The corresponding function h' satisfies $h'(r, p) \geq h(r, p) - 0.001$. Thus Lemma 10 shows that $h'(r, p) \geq 0.002$ for all $r \in [0, 1]$ and $p \in [0, 1]$.

The potential function $f(x)$ is defined as before, but under the truncated input distribution we have the property required for Theorem 2. If $|f(x) - f(y)| > M + B$ then the probability of moving from x to y in a single move is 0.

The lemma follows from Theorem 2, where the values of C, ε, and d can be taken to be 1, 0.002, and $M + B$, respectively. □

We now have the following theorem.

Theorem 13 *Every backoff protocol is transient under the Poisson distribution with arrival rate $\lambda \geq 0.42$.*

Proof: The proof is immediate from Lemma 12 and Observation 4. □

Finally, we bound the capacity of every backoff protocol.

Theorem 14 *The capacity of every backoff protocol is at most* 0.42.

Proof: Let $p_0, p_1, \ldots$ be a backoff protocol, let $\lambda'' \geq 0.42$ be the arrival rate and let $\lambda = 0.42$. View the arrivals at each step as Poisson(λ) "ordinary" messages and Poisson($\lambda'' - \lambda$) "ghost" messages. We will show that the protocol does not achieve throughput λ''. Let $Y_0, Y_1, \ldots$ be the Markov chain describing the protocol. Let $k(Y_t)$ be the number of ordinary messages in the system after t steps. Clearly, the expected number of successes in the first t steps is at most $\lambda'' t - E[k(Y_t)]$. Now let $X_1, X_2, \ldots$ be the Markov chain describing the evolution of the backoff protocol with arrival rate λ (with no ghost messages). By deletion resilience (Observation 3), $E[k(Y_t)] \geq E[k(X_t)]$. Now by Lemmas 9 and 10, $E[k(X_t)] \geq E[f(X_t)] - B \geq 0.003\, t - B$. Thus, the expected number of successes in the first t steps is at most $(\lambda'' - 0.003)t + B$, which is less than $\lambda'' t$ if t is sufficiently large. (If X_0 is the empty state, then we do not require t to be sufficiently large, because $E[f(X_t)] \geq 0.003t + B$.) □

5 Acknowledgement-based protocols

We will prove that every acknowledgement-based protocol is transient for all $\lambda > 0.531$; see Theorem 20 for a precise statement of this claim.

An acknowledgement-based protocol can be viewed a system which, at every step t, decides what subset of the old messages to send. The decision is a probabilistic one dependent on the histories of the messages held. As a technical device for proving our bounds, we introduce the notion of a "genie", which (in general) has more freedom in making these decisions than a protocol.

Since we only consider acknowledgement-based protocols, the behaviour of each new message is independent of the other messages and of the state of the system until after its first send. This is why we ignore new messages until their first send – for Poisson arrivals this is equivalent to the convention that each message sends at its arrival time. As a consequence, we impose the limitation on a genie, that each decision is independent of the number of arrivals at that step.

A *genie* is a random variable over the natural numbers, dependent on the complete history (of arrivals and sends of messages) up to time $t-1$, which gives a natural number representing the number of (old) messages to send at time t. It is clear that for every acknowledgement-based protocol there is a corresponding genie. However there are genies which do not behave like any protocol, e.g., a genie may give a cumulative total number of "sends" up to time t which exceeds the actual number of arrivals up to that time.

We prove a preliminary result for such "unconstrained" genies, but then we impose some constraints reflecting properties of a given protocol in order to prove our final results.

Let $I(t), G(t)$ be the number of arrivals and the genie's send value, respectively, at step t. It is convenient to introduce some indicator variables to express various outcomes at the step under consideration. We use i_0, i_1 for the events of no new arrival, or exactly one arrival, respectively, and g_0, g_1 for the events of no send and exactly one send from the genie. The indicator random variable $S(t)$ for a success at time t is given by $S(t) = i_0 g_1 + i_1 g_0$. Let $\mathrm{In}(t) = \sum_{j\le t} I(j)$ and $\mathrm{Out}(t) = \sum_{j\le t} S(j)$. Define $\mathrm{Backlog}(t) = \mathrm{In}(t) - \mathrm{Out}(t)$. Let $\lambda = \lambda_0 \approx 0.567$ be the (unique) root of $\lambda = e^{-\lambda}$.

Lemma 15 *For any genie and input rate $\lambda > \lambda_0$, there exists $\varepsilon > 0$ such that*

$$\mathrm{Prob}[\mathrm{Backlog}(t) > \varepsilon t \text{ for all } t \ge T] \to 1 \text{ as } T \to \infty.$$

Proof: Let $3\varepsilon = \lambda - e^{-\lambda} > 0$. At any step t, $S(t)$ is a Bernoulli variable with expectation $0, e^{-\lambda}, \lambda e^{-\lambda}$, according as $G(t) > 1, G(t) = 1, G(t) = 0$, respectively, which is dominated by the Bernoulli variable with expectation $e^{-\lambda}$. Therefore $\mathrm{E}[\mathrm{Out}(t)] \le e^{-\lambda} t$, and also, $\mathrm{Prob}[\mathrm{Out}(t) - e^{-\lambda} t < \varepsilon t \text{ for all } t \ge T] \to 1$ as $T \to \infty$. (To see this note that, by a Chernoff bound, $\mathrm{Prob}[\mathrm{Out}(t) - e^{-\lambda} t \ge \varepsilon t] \le e^{-\delta t}$ for a positive constant δ. Thus,

$$\mathrm{Prob}[\exists t \ge T \text{ such that } \mathrm{Out}(t) - e^{-\lambda} t \ge \varepsilon t] \le \sum_{t \ge T} e^{-\delta t},$$

which goes to 0 as T goes to ∞.)

We also have $\mathrm{E}[\mathrm{In}(t)] = \lambda t$ and $\mathrm{Prob}[\lambda t - \mathrm{In}(t) \le \varepsilon t \text{ for all } t \ge T] \to 1$ as $T \to \infty$, since $\mathrm{In}(t) = \mathrm{Poisson}(\lambda t)$.

Since

$$\begin{aligned}
\text{Backlog}(t) &= \text{In}(t) - \text{Out}(t) \\
&= (\lambda - e^{-\lambda})t + (\text{In}(t) - \lambda t) + (e^{-\lambda}t - \text{Out}(t)) \\
&= \varepsilon t + (\varepsilon t + \text{In}(t) - \lambda t) + (\varepsilon t + e^{-\lambda}t - \text{Out}(t)),
\end{aligned}$$

the result follows. □

Corollary 16 *No acknowledgement-based protocol is recurrent for $\lambda > \lambda_0$ or has capacity greater than λ_0.*

To strengthen the above result we introduce a restricted class of genies. We think of the messages which have failed exactly once as being contained in *the bucket*. (More generally, we could consider an array of buckets, where the jth bucket contains those messages which have failed exactly j times.) A 1-*bucket genie*, here called simply a *bucket genie*, is a genie which simulates a given protocol for the messages in the bucket and is required to choose a send value which is at least as great as the number of sends from the bucket. For such constrained genies, we can improve the bound of Corollary 16.

For the range of arrival rates we consider, an excellent strategy for a genie is to ensure that at least one message is sent at each step. Of course a bucket genie has to respect the bucket messages and is obliged sometimes to send more than one message (inevitably failing). An *eager* genie always sends at least one message, but otherwise sends the minimum number consistent with its constraints.

An eager bucket genie is easy to analyse, since every arrival is blocked by the genie and enters the bucket.

Let $\lambda = \lambda_1 \approx 0.531$ be the (unique) root of $\lambda = (1+\lambda)e^{-2\lambda}$.

Lemma 17 *For any eager bucket genie and input rate $\lambda > \lambda_1$, there exists $\varepsilon > 0$ such that*

$$\text{Prob}[\text{Backlog}(t) > \varepsilon t \text{ for all } t \geq T] \to 1 \text{ as } T \to \infty.$$

Proof: Let Eager be an eager bucket genie. Let r_i be the probability that a message in the bucket sends for the first time (and hence exits from the bucket) i steps after its arrival. Assume $\sum_{i=1}^{\infty} r_i = 1$, otherwise there is a positive probability that the message never exits from the bucket, and the result follows trivially.

The generating function for the Poisson distribution with rate λ is $e^{\lambda(z-1)}$ (i.e., the coefficient of z^k in this function gives the probability of exactly k arrivals; see, e.g., [10]). Consider the sends from the bucket at step t. Since Eager always blocks arriving messages, the generating function for messages entering the bucket i time steps in the past, $1 \le i \le t$, is $e^{\lambda(z-1)}$. Some of these messages may send at step t; the generating function for the number of sends is $e^{\lambda[(1-r_i)+r_i z-1]} = e^{\lambda r_i(z-1)}$. Finally, the generating function for all sends from the bucket at step t is the convolution of all these functions, i.e.,

$$\prod_{i=1}^{t} \exp(\lambda r_i(z-1)) = \exp\left[\lambda(z-1)\sum_{i=1}^{t} r_i\right].$$

For any $\delta > 0$, we can choose t sufficiently large so that $\sum_{i=1}^{t} r_i > 1-\delta$. The number of sends from the bucket at step t is distributed as Poisson(λ'), where $(1-\delta)\lambda < \lambda' \le \lambda$. The number of new arrivals sending at step t is independently Poisson(λ). The only situation in which a message succeeds under Eager is when there are no new arrivals and the number of sends from the bucket is zero or one. Thus the success probability at step t is $e^{-\lambda}e^{-\lambda'}(1+\lambda')$. For sufficiently small δ, we have $\lambda_1 < \lambda' \le \lambda$, and so $e^{-\lambda'}(1+\lambda') < e^{-\lambda_1}(1+\lambda_1) = e^{\lambda_1}\lambda_1 < e^{\lambda}\lambda$. Hence $e^{-\lambda}e^{-\lambda'}(1+\lambda') \le \lambda - 3\epsilon$ for ϵ sufficiently small. Thus the success event is dominated by a Bernoulli variable with expectation $\lambda - 3\epsilon$. Hence, as in the previous lemma,

$$\text{Prob}[\text{Backlog}(t) > \varepsilon t \text{ for all } t \ge T] \to 1 \text{ as } T \to \infty,$$

completing the proof. □

Let *Any* be an arbitrary bucket genie and let *Eager* be the eager bucket genie based on the same bucket parameters. We may couple the executions of Eager and Any so that the same arrival sequences are presented to each. It will be clear that at any stage the set of messages in Any's bucket is a subset of those in Eager's bucket. We may further couple the behaviour of the common subset of messages.

Let $\lambda = \lambda_2 \approx 0.659$ be the (unique) root of $\lambda = 1 - \lambda e^{-\lambda}$.

Lemma 18 *For the coupled genies* Any *and* Eager *defined above, if* Out_A *and* Out_E *are the corresponding output functions, we define*

$$\Delta\mathrm{Out}(t) = \mathrm{Out}_E(t) - \mathrm{Out}_A(t).$$

For any $\lambda \leq \lambda_2$ *and any* $\varepsilon > 0$,

$$\mathrm{Prob}[\Delta\mathrm{Out}(t) \geq -\varepsilon t \text{ for all } t \geq T] \to 1 \text{ as } T \to \infty.$$

Proof: Let c_0, c_1, c_* be indicators for the events of the number of common messages sending being 0, 1, or more than one, respectively. In addition, for the messages which are only in Eager's bucket, we use the similar indicators e_0, e_1, e_*. Let a_0, a_1 represent Any not sending, or sending, *additional* messages respectively. (Note that Eager's behaviour is fully determined.)

We write $Z(t)$ for $\Delta\mathrm{Out}(t) - \Delta\mathrm{Out}(t-1)$, for $t > 0$, so Z represents the difference in success between Eager and Any in one step. In terms of the indicators we have

$$\begin{aligned} Z(t) &= S_E(t) - S_A(t) \\ &= i_0 g_{E1}(t) + i_1 g_{E0}(t) - i_0 g_{A1}(t) - i_1 g_{A0}(t), \end{aligned}$$

where $S_E(t)$ is the indicator random variable for a success of Eager at time t and $g_{E1}(t)$ is the event that Eager sends exactly one message during step t (and so on) as in the paragraph before Lemma 15. Thus,

$$Z(t) \geq i_0 c_0 (a_0(e_0 + e_1) - a_1 e_*) - i_0 c_1 (e_1 + e_*) - i_1 c_0 a_0.$$

Note that if the number of arrivals plus the number of common bucket sends is more than 1 then neither genie can succeed. We also need to keep track of the number, ΔB, of extra messages in Eager's bucket. At any step, at most one new such extra message can arrive; the indicator for this event is $i_1 c_0 a_0$, i.e., there is a single arrival and no sends from the common bucket, so if Any does not send then this message succeeds but Eager's send will cause a failure. The number of "extra" messages leaving Eager's bucket at any step is unbounded, given by a random variable we could show as $\mathbf{e} = 1 \cdot e_1 + 2 \cdot e_2 + \cdots$. However $\mathbf{e}$ dominates $e_1 + e_*$ and it is sufficient to use the latter. The change at one step in the number of extra messages satisfies:

$$\Delta B(t) - \Delta B(t-1) = i_1 c_0 a_0 - \mathbf{e} \leq i_1 c_0 a_0 - (e_1 + e_*).$$

Next we define $Y(t) = Z(t) - \alpha(\Delta B(t) - \Delta B(t-1))$, for some positive constant α to be chosen below. Note that $X(t) = \sum_{j=1}^{t} Y(j) = \Delta\mathrm{Out}(t) - \alpha\Delta B(t)$. We also define

$$Y'(t) = i_0 c_0 (a_0(e_0 + e_1) - a_1 e_*) - i_0 c_1 (e_1 + e_*) - i_1 c_0 a_0 - \alpha(i_1 c_0 a_0 - (e_1 + e_*))$$

and $X'(t) = \sum_{j=1}^{t} Y'(j)$. Note that $Y(t) \geq Y'(t)$. That is, $Y(t)$ dominates $Y'(t)$.

We can identify five (exhaustive) cases A,B,C,D,E depending on the values of the c's, a's and e's, such that in each case $Y'(t)$ dominates a given random variable depending only on $I(t)$.

A.	c_*:	$Y'(t) \geq 0$;
B.	$(c_1 + c_0 a_1)(e_1 + e_*)$:	$Y'(t) \geq \alpha - i_0$;
C.	$(c_1 + c_0 a_1) e_0$:	$Y'(t) \geq 0$;
D.	$c_0 a_0 (e_0 + e_1)$:	$Y'(t) \geq i_0 - (1+\alpha) i_1$;
E.	$c_0 a_0 e_*$:	$Y'(t) \geq \alpha - (1+\alpha) i_1$.

For example, the correct interpretation of Case B is "conditioned on $(c_1 + c_0 a_1)(e_1 + e_*) = 1$, the value of $Y'(t)$ is at least $\alpha - i_0$." Since $\mathrm{E}[i_0] = e^{-\lambda}$ and $\mathrm{E}[i_1] = \lambda e^{-\lambda}$, we have $\mathrm{E}[Y'(t)] \geq 0$ in each case, provided that $\max\{e^{-\lambda}, \lambda e^{-\lambda}/(1 - \lambda e^{-\lambda})\} \leq \alpha \leq 1/\lambda - 1$. There exists such an α for any $\lambda \leq \lambda_2$; for such λ we may take the value $\alpha = e^{-\lambda}$, say.

Let $\mathcal{F}_t$ be the σ-field generated by the first t steps of the coupled process. Let $\hat{Y}(t) = Y'(t) - E[Y'(t) \mid \mathcal{F}_{t-1}]$ and let $\hat{X}(t) = \sum_{i=1}^{t} \hat{Y}(t)$. The sequence $\hat{X}(0), \hat{X}(1), \ldots$ forms a *martingale* (see Definition 4.11 of [20]) since $E[\hat{X}(t) \mid \mathcal{F}_{t-1}] = \hat{X}(t-1)$. Furthermore, there is a positive constant c such that $|\hat{X}(t) - \hat{X}(t-1)| \leq c$. Thus, we can apply the Hoeffding-Azuma Inequality (see Theorem 4.16 of [20]):

Theorem 19 *(Hoeffding, Azuma) Let $X_0, X_1, \ldots$ be a martingale sequence such that for each k*

$$|X_k - X_{k-1}| \leq c_k,$$

where c_k may depend upon k. Then, for all $t \geq 0$ and any $\lambda > 0$,

$$\mathrm{Prob}[|X_t - X_0| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2}\right).$$

In particular, we can conclude that

$$\mathrm{Prob}[\hat{X}_t \leq -\epsilon t] \leq 2\exp\left(-\frac{\epsilon^2 t}{2c^2}\right).$$

Our choice of α above ensured that $E[Y'(t) \mid \mathcal{F}_{t-1}] \geq 0$. Hence $Y'(t) \geq \hat{Y}(t)$ and $X'(t) \geq \hat{X}(t)$. We observed earlier that $X(t) \geq X'(t)$. Thus, $X(t) \geq \hat{X}(t)$ so we have

$$\mathrm{Prob}[X_t \leq -\epsilon t] \leq 2\exp\left(-\frac{\epsilon^2 t}{2c^2}\right).$$

Since $2\exp\left(-\frac{\epsilon^2 t}{2c^2}\right)$ converges, we deduce that

$$\mathrm{Prob}[X(t) \geq -\varepsilon t \text{ for all } t \geq T] \to 1 \text{ as } T \to \infty.$$

Since $\Delta\mathrm{Out}(t) = X(t) + \alpha\Delta B(t) \geq X(t)$, for all t, we obtain the required conclusion. □

Finally, we can prove the main results of this section.

Theorem 20 *Let P be an acknowledgement-based protocol. Let $\lambda = \lambda_1 \approx 0.531$ be the (unique) root of $\lambda = (1+\lambda)e^{-2\lambda}$. Then*

1. *P is transient for arrival rates greater than λ_1;*
2. *P has capacity no greater than λ_1.*

Proof: Let λ be the arrival rate, and suppose $\lambda > \lambda_1$. If $\lambda > \lambda_0 \approx 0.567$ then the result follows from Lemma 15. Otherwise, we can assume that $\lambda < \lambda_2 \approx 0.659$. If E is the eager genie derived from P, then the corresponding Backlogs satisfy $\mathrm{Backlog}_P(t) = \mathrm{Backlog}_E(t) + \Delta\mathrm{Out}(t)$. The results of Lemmas 17 and 18 show that, for some $\varepsilon > 0$, both $\mathrm{Prob}[\mathrm{Backlog}_E(t) > 2\varepsilon t$ for all $t \geq T]$ and $\mathrm{Prob}[\Delta\mathrm{Out}(t) \geq -\varepsilon t$ for all $t \geq T]$ tend to 1 as $T \to \infty$. The conclusion of the theorem follows. □

References

[1] D. Aldous, Ultimate instability of exponential back-off protocol for acknowledgement-based transmission control of random access communication channels, *IEEE Trans. Inf. Theory* **IT-33(2)** (1987) 219–233.

[2] H. Al-Ammal, L.A. Goldberg and P. MacKenzie, An improved stability bound for binary exponential backoff, To appear in *Theory of Computing Systems* (2000).

[3] J.I. Capetanakis, Tree algorithms for packet broadcast channels, *IEEE Trans. Inf. Theory* **IT-25(5)** (1979) 505-515.

[4] A. Ephremides and B. Hajek, Information theory and communication networks: an unconsummated union, *IEEE Trans. Inf. Theory* **44(6)** (1998) 2416–2432.

[5] G. Fayolle, V.A. Malyshev, and M.V. Menshikov, *Topics in the Constructive Theory of Countable Markov Chains,* (Cambridge University Press, 1995).

[6] L.A. Goldberg and P.D. MacKenzie, Analysis of Practical Backoff Protocols for Contention Resolution with Multiple Servers, *Journal of Computer and Systems Sciences*, **58** (1999) 232–258.

[7] L.A. Goldberg, P.D. MacKenzie, M. Paterson and A. Srinivasan, Contention resolution with constant expected delay, To appear in *J. of the ACM.*

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

[9] A.G. Greenberg, P. Flajolet and R. Ladner, Estimating the multiplicities of conflicts to speed their resolution in multiple access channels, *J. of the ACM*, **34(2)** (1987) 289–325.

[10] G.R. Grimmet and D.R. Stirzaker, *Probability and Random Processes, Second Edition.* (Oxford University Press, 1992)

[11] J. Håstad, T. Leighton and B. Rogoff, Analysis of backoff protocols for multiple access channels, *SIAM Journal on Computing* **25(4)** (1996) 740-774.

[12] F.P. Kelly, Stochastic models of computer communication systems, *J.R. Statist. Soc. B* **47(3)** (1985) 379–395.

[13] F.P. Kelly and I.M. MacPhee, The number of packets transmitted by collision detect random access schemes, *The Annals of Probability,* **15(4)** (1987) 1557–1568.

[14] U. Loher, Efficiency of first-come first-served algorithms, *Proc. ISIT* (1998) p108.

[15] U. Loher, Information-theoretic and genie-aided analyses of random-access algorithms, PhD Thesis, Swiss Federal Institute of Technology, DISS ETH No. 12627, Zurich (1998).

[16] I.M. MacPhee, On optimal strategies in stochastic decision processes, D. Phil Thesis, University of Cambridge, (1987).

[17] R.M. Metcalfe and D.R. Boggs, Ethernet: Distributed packet switching for local computer networks. *Commun. ACM,* **19** (1976) 395–404.

[18] M. Molle and G.C. Polyzos, Conflict resolution algorithms and their performance analysis, Technical Report CS93-300, Computer Systems Research Institute, University of Toronto, (1993).

[19] J. Mosely and P.A. Humblet, A class of efficient contention resolution algorithms for multiple access channels, *IEEE Trans. on Communications,* **COM-33(2)** (1985) 145–151.

[20] R. Motwani and P. Raghavan, *Randomized Algorithms* (Cambridge University Press, 1995.)

[21] P. Raghavan and E. Upfal, Stochastic contention resolution with short delays, *SIAM Journal on Computing* **28(2)** (1999) 709–719.

[22] R. Rom and M. Sidi, Multiple access protocols: performance and analysis, (Springer-Verlag, 1990).

[23] B.S. Tsybakov and N. B. Likhanov, Upper bound on the capacity of a random multiple-access system, *Problemy Peredachi Informatsii,* **23(3)** (1987) 64–78.

[24] B.S. Tsybakov and V. A. Mikhailov, Free synchronous packet access in a broadcast channel with feedback, *Probl. Information Transmission,* **14(4)** (1978) 259–280.

[25] S. Verdu, Computation of the efficiency of the Mosely-Humblet contention resolution algorithm: a simple method, *Proc. of the IEEE* **74(4)** (1986) 613–614.

[26] N.S. Vvedenskaya and M.S. Pinsker, Nonoptimality of the part-and-try algorithm, *Abstr. Papers, Int. Workshop "Convolutional Codes; Multi-User Commun.,"* Sochi, U.S.S.R, May 30 - June 6, 1983, 141–144.

Appendix. Proof of Lemma 10

Let $j(r,p) = -h(r,p)$. We will show that for any $r \in [0,1]$ and $p \in [0,1]$, $j(r,p) \leq -0.003$.

Case 1: $r + p - rp \geq r \geq B/A$ **and** $p \geq r$.

In this case we have

$$\begin{aligned} g(r,p) &= e^{-\lambda}((1-r)p + (1-p)r + (1-p)(1-r)\lambda), \\ j(r,p) &= g(r,p) - \lambda. \end{aligned}$$

Observe that

$$j(r,p) = e^{-\lambda} \sum_{i=0}^{1} \sum_{j=0}^{1} c_{i,j} p^i r^j,$$

where the coefficients $c_{i,j}$ are defined as follows.

$$\begin{aligned} c_{0,0} &= \lambda(1 - e^{\lambda}) \\ c_{1,0} &= 1 - \lambda \\ c_{0,1} &= 1 - \lambda \\ c_{1,1} &= -2 + \lambda. \end{aligned}$$

Note that the only positive coefficients are $c_{1,0}$ and $c_{0,1}$. Thus, if $p \in [p_1, p_2]$ and $r \in [r_1, r_2]$, then $j(r,p)$ is at most $\mathcal{U}(p_1, p_2, r_1, r_2)$, which we define as

$$e^{-\lambda}(c_{0,0} + c_{1,0}p_2 + c_{0,1}r_2 + c_{1,1}p_1 r_1).$$

Now we need only check that for all $r_1 \in (B/A - 0.01, 1)$ and $p_1 \in [r_1, 1)$ such that p_1 and r_1 are multiples of 0.01, $\mathcal{U}(p_1, p_1 + 0.01, r_1, r_1 + 0.01)$ is at most -0.003. This is the case. (The highest value is $\mathcal{U}(0.45, 0.46, 0.45, 0.46)$ which is -0.00366228.)

Case 2: $r + p - rp \geq r \geq B/A$ **and** $p < r$.

Now we have

$$\begin{aligned} g(r,p) &= e^{-\lambda}((1-r)p + (1-p)\frac{r - r^2/2}{1 - p/2} + (1-p)(1-r)\lambda) \\ j(r,p) &= g(r,p) - \lambda. \end{aligned}$$

Observe that

$$(1-p/2)j(r,p) = e^{-\lambda} \sum_{i=0}^{2} \sum_{j=0}^{2} c_{i,j} p^i r^j,$$

where the coefficients $c_{i,j}$ are defined as follows.

$$\begin{aligned}
c_{0,0} &= \lambda(1-e^{\lambda}) \\
c_{1,0} &= 1 - 3\lambda/2 + e^{\lambda}\lambda/2 \\
c_{0,1} &= 1-\lambda \\
c_{1,1} &= -2 + 3\lambda/2 \\
c_{2,0} &= -1/2 + \lambda/2 \\
c_{0,2} &= -1/2 \\
c_{2,1} &= 1/2 - \lambda/2 \\
c_{1,2} &= 1/2 \\
c_{2,2} &= 0.
\end{aligned}$$

Note that the only positive coefficients are $c_{1,0}$, $c_{0,1}$, $c_{2,1}$ and $c_{1,2}$. Thus, if $p \in [p_1, p_2]$ and $r \in [r_1, r_2]$, then $j(r,p)$ is at most $\mathcal{U}(p_1, p_2, r_1, r_2)$, which we define as

$$\begin{aligned}
c_{0,0} &+ c_{1,0}p_2 + c_{0,1}r_2 + c_{1,1}p_1r_1 + c_{2,0}p_1^2 + c_{0,2}r_1^2 + c_{2,1}p_2^2r_2 + c_{1,2}p_2r_2^2 \\
&+ c_{2,2}p_1^2r_1^2
\end{aligned}$$

divided by $e^{\lambda}(1-p_2/2)$.

Now we need only check that for all $r_1 \in (B/A - 0.005, 1)$ and $p_1 \in [0, r_1]$ such that p_1 and r_1 are multiples of 0.005, $\mathcal{U}(p_1, p_1+0.005, r_1, r_1+0.005)$ is at most -0.003. This is the case. (The highest value for these parameters is $\mathcal{U}(0.45, 0.455, 0.455, 0.46) = -0.00479648$.)

Case 3: $r + p - rp \geq B/A \geq r$ **and** $p \geq r$.

In this case we have

$$\begin{aligned}
g(r,p) &= e^{-\lambda}((1-r)p + (1-p)r + (1-p)(1-r)\lambda). \\
j(r,p) &= g(r,p) - \lambda - (-Ar+B)e^{-\lambda}(1-r)p.
\end{aligned}$$

Observe that

$$j(r,p) = e^{-\lambda} \sum_{i=0}^{2} \sum_{j=0}^{2} c_{i,j} p^i r^j,$$

where the coefficients $c_{i,j}$ are defined as follows.

$$\begin{aligned}
c_{0,0} &= \lambda(1-e^{\lambda}) \\
c_{1,0} &= 1-B-\lambda \\
c_{0,1} &= 1-\lambda \\
c_{1,1} &= -2+A+B+\lambda \\
c_{2,0} &= 0 \\
c_{0,2} &= 0 \\
c_{2,1} &= 0 \\
c_{1,2} &= -A \\
c_{2,2} &= 0.
\end{aligned}$$

Note that the only positive coefficients are $c_{1,0}$ and $c_{0,1}$. Thus, if $p \in [p_1, p_2]$ and $r \in [r_1, r_2]$, then $j(r,p)$ is at most $\mathcal{U}(p_1, p_2, r_1, r_2)$, which we define as

$$\frac{c_{0,0}+c_{1,0}p_2+c_{0,1}r_2+c_{1,1}p_1r_1+c_{2,0}p_1^2+c_{0,2}r_1^2+c_{2,1}p_1^2r_1+c_{1,2}p_1r_1^2+c_{2,2}p_1^2r_1^2}{e^{\lambda}}.$$

Now we need only check that for all $p_1 \in [0,1)$ and $r_1 \in [0,p_1]$ such that p_1 and r_1 are multiples of 0.01, $\mathcal{U}(p_1, p_1+0.01, r_1, r_1+0.01)$ is at most -0.003. This is the case. (The highest value is $\mathcal{U}(0.44, 0.45, 0.44, 0.45) = -0.00700507$.)

Case 4: $r+p-rp \geq B/A \geq r$ **and** $p < r$.

Now we have

$$\begin{aligned}
g(r,p) &= e^{-\lambda}\left((1-r)p+(1-p)\frac{r-r^2/2}{1-p/2}+(1-p)(1-r)\lambda\right) \\
j(r,p) &= g(r,p)-\lambda-(-Ar+B)e^{-\lambda}(1-r)p.
\end{aligned}$$

Observe that

$$j(r,p) = e^{-\lambda}(1/2)\frac{\sum_{i=0}^{2}\sum_{j=0}^{2} c_{i,j}p^i r^j}{1-p/2},$$

where the coefficients $c_{i,j}$ are defined as follows.

$$c_{0,0} = 2\lambda(1-e^{\lambda})$$

$$
\begin{aligned}
c_{1,0} &= 2 - 2B - 3\lambda + \lambda e^{\lambda} \\
c_{0,1} &= 2 - 2\lambda \\
c_{1,1} &= -4 + 2A + 2B + 3\lambda \\
c_{2,0} &= -1 + B + \lambda \\
c_{0,2} &= -1 \\
c_{2,1} &= 1 - A - B - \lambda \\
c_{1,2} &= 1 - 2A \\
c_{2,2} &= A.
\end{aligned}
$$

Note that the coefficients are all negative except $c_{1,0}$, $c_{0,1}$ and $c_{2,2}$. Thus, if $p \in [p_1, p_2]$ and $r \in [r_1, r_2]$, then $j(r,p)$ is at most $\mathcal{U}(p_1, p_2, r_1, r_2)$, which we define as

$$
\frac{c_{0,0} + c_{1,0}p_2 + c_{0,1}r_2 + c_{1,1}p_1r_1 + c_{2,0}p_1^2 + c_{0,2}r_1^2 + c_{2,1}p_1^2r_1 + c_{1,2}p_1r_1^2 + c_{2,2}p_2^2r_2^2}{2e^{\lambda}(1 - p_2/2)}.
$$

Now we need only check that for all $p_1 \in [0, 1)$ and $r_1 \in [p_1, 1)$ such that p_1 and r_1 are multiples of 0.01, $\mathcal{U}(p_1, p_1 + 0.01, r_1, r_1 + 0.01)$ is at most -0.003. This is the case. (The highest value is $\mathcal{U}(0.44, 0.45, 0.44, 0.45) = -0.00337716$.)

Case 5: $B/A \geq r + p - rp \geq r$ **and** $p \geq r$.

In this case we have

$$
\begin{aligned}
g(r,p) &= e^{-\lambda}((1 - r)p + (1 - p)r + (1 - p)(1 - r)\lambda). \\
j(r,p) &= g(r,p) - \lambda \\
&+ (-A(r + p - rp) + B)(1 - (1 - r)(1 - p)e^{-\lambda}(1 + \lambda)) \\
&- (-Ar + B)e^{-\lambda}(1 - r)p.
\end{aligned}
$$

Observe that

$$
j(r,p) = e^{-\lambda} \sum_{i=0}^{2} \sum_{j=0}^{2} c_{i,j} p^i r^j,
$$

where the coefficients $c_{i,j}$ are defined as follows.

$$
c_{0,0} = -B + Be^{\lambda} + \lambda - B\lambda - e^{\lambda}\lambda
$$

$$
\begin{aligned}
c_{1,0} &= 1 + A - Ae^{\lambda} - \lambda + A\lambda + B\lambda \\
c_{0,1} &= 1 + A + B - Ae^{\lambda} - \lambda + A\lambda + B\lambda \\
c_{1,1} &= -2 - 2A + Ae^{\lambda} + \lambda - 3A\lambda - B\lambda \\
c_{2,0} &= -A - A\lambda \\
c_{0,2} &= -A - A\lambda \\
c_{2,1} &= 2A + 2A\lambda \\
c_{1,2} &= A + 2A\lambda \\
c_{2,2} &= -A - A\lambda.
\end{aligned}
$$

Note that the only positive coefficients are $c_{1,0}$, $c_{0,1}$, $c_{2,1}$ and $c_{1,2}$. Thus, if $p \in [p_1, p_2]$ and $r \in [r_1, r_2]$, then $j(r,p)$ is at most $\mathcal{U}(p_1, p_2, r_1, r_2)$, which we define as

$$
\frac{c_{0,0} + c_{1,0}p_2 + c_{0,1}r_2 + c_{1,1}p_1r_1 + c_{2,0}p_1^2 + c_{0,2}r_1^2 + c_{2,1}p_2^2r_2 + c_{1,2}p_2r_2^2 + c_{2,2}p_1^2r_1^2}{e^{\lambda}}.
$$

Now we need only check that for all $p_1 \in [0,1)$ and $r_1 \in [0, p_1]$ such that p_1 and r_1 are multiples of 0.01, $\mathcal{U}(p_1, p_1 + 0.01, r_1, r_1 + 0.01)$ is at most -0.003. This is the case. (The highest value is $\mathcal{U}(0.19, 0.2, 0.19, 0.2) = -0.0073656$.)

Case 6: $B/A \geq r + p - rp \geq r$ **and** $p < r$.

Now we have

$$
\begin{aligned}
g(r,p) &= e^{-\lambda}((1-r)p + (1-p)\frac{r - r^2/2}{1 - p/2} + (1-p)(1-r)\lambda) \\
j(r,p) &= g(r,p) - \lambda \\
&+ (-A(r + p - rp) + B)(1 - (1-r)(1-p)e^{-\lambda}(1+\lambda)) \\
&- (-Ar + B)e^{-\lambda}(1-r)p.
\end{aligned}
$$

Observe that

$$
(1 - p/2)j(r,p) = e^{-\lambda}\sum_{i=0}^{3}\sum_{j=0}^{2} c_{i,j}p^i r^j,
$$

where the coefficients $c_{i,j}$ are defined as follows.

$$
\begin{aligned}
c_{0,0} &= -B + Be^{\lambda} + \lambda - B\lambda - e^{\lambda}\lambda \\
c_{1,0} &= 1 + A + B/2 - Ae^{\lambda} - Be^{\lambda}/2 - 3\lambda/2 + A\lambda + 3B\lambda/2 + e^{\lambda}\lambda/2
\end{aligned}
$$

$$
\begin{aligned}
c_{0,1} &= 1 + A + B - Ae^{\lambda} - \lambda + A\lambda + B\lambda \\
c_{1,1} &= -2 - 5A/2 - B/2 + 3Ae^{\lambda}/2 + 3\lambda/2 - 7A\lambda/2 - 3B\lambda/2 \\
c_{2,0} &= -1/2 - 3A/2 + Ae^{\lambda}/2 + \lambda/2 - 3A\lambda/2 - B\lambda/2 \\
c_{0,2} &= -1/2 - A - A\lambda \\
c_{2,1} &= 1/2 + 3A - Ae^{\lambda}/2 - \lambda/2 + 7A\lambda/2 + B\lambda/2 \\
c_{1,2} &= 1/2 + 3A/2 + 5A\lambda/2 \\
c_{2,2} &= -3A/2 - 2A\lambda \\
c_{3,0} &= A/2 + A\lambda/2 \\
c_{3,1} &= -A - A\lambda \\
c_{3,2} &= A/2 + A\lambda/2.
\end{aligned}
$$

Note that the only positive coefficients are $c_{1,0}$, $c_{0,1}$, $c_{2,1}$, $c_{1,2}$, $c_{3,0}$ and $c_{3,2}$. Thus, if $p \in [p_1, p_2]$ and $r \in [r_1, r_2]$, then $j(r,p)$ is at most $\mathcal{U}(p_1, p_2, r_1, r_2)$, which we define as

$$
\begin{aligned}
c_{0,0} &+ c_{1,0}p_2 + c_{0,1}r_2 + c_{1,1}p_1r_1 + c_{2,0}p_1^2 + c_{0,2}r_1^2 + c_{2,1}p_2^2r_2 + c_{1,2}p_2r_2^2 \\
&+ c_{2,2}p_1^2r_1^2 + c_{3,0}p_2^3 + c_{3,1}p_1^3r_1 + c_{3,2}p_2^3r_2^2
\end{aligned}
$$

divided by $e^{\lambda}(1 - p_2/2)$.

Now we need only check that for all $p_1 \in [0, 1)$ and $r_1 \in [p_1, 1)$ such that p_1 and r_1 are multiples of 0.005, $\mathcal{U}(p_1, p_1 + 0.005, r_1, r_1 + 0.005)$ is at most -0.003. This is the case. (The highest value is $\mathcal{U}(0.01, 0.015, 0.3, 0.305) = -0.00383814$.)