Available online at www.sciencedirect.com
SCIENCE DIRECT®

Performance Evaluation 57 (2004) 357–377

PERFORMANCE EVALUATION
An International Journal

www.elsevier.com/locate/peva

Robustness of Q-ary collision resolution algorithms in random access systems

B. Van Houdt*, C. Blondia

Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium

Received 4 September 2002; received in revised form 30 January 2004

Available online 27 April 2004

Abstract

The throughput characteristics of a random access system (RAS) which uses Q-ary tree algorithms (where Q is the number of groups into which colliding users are split) of the Capetanakis–Tsybakov–Mikhailov–Vvedenskaya type are analyzed for an infinite population of identical users generating packets. In the standard model packets are assumed to be generated according to a Poisson process. In this paper we greatly relax this assumption and consider a rich class of Markovian arrival processes, which, in general, are non-renewal. This class of arrival processes is known to lend itself very well to modeling bursty and correlated arrival processes commonly arising in computer and communication applications. Blocked and grouped channel access protocols are considered in combination with Q-ary collision resolution algorithms that exploit either binary ("collision or not") or ternary ("collision, success or idle") feedback. For the resulting RASs the corresponding maximum stable throughput is determined. It is concluded that the resulting RASs maintain their good stability characteristics under the wide range of arrival processes considered, thereby further extending the theoretical foundations of tree algorithms.

Keywords: Random access; Maximum stable throughput; D-BMAP arrivals; Tree algorithms

1. Introduction

Random access systems (RAS) of the Capetanakis–Tsybakov–Mikhailov–Vvedenskaya (CTMV) type have been studied extensively over the past 20 years [1–7]. Underlying most of the theoretical work done in this area are the following key assumptions [8,9]:

1. New arrivals occur according to a Poisson process with rate λ.
2. The number of nodes or stations is assumed to be infinite. In practice, the number of nodes is always finite. Assuming an infinite number provides us with pessimistic estimates for finite populations [9,10]. In particular, each finite set of nodes can regard itself as an infinite set of virtual stations, one for each

* Corresponding author.
E-mail addresses: benny.vanhoudt@ua.ac.be (B. Van Houdt), chris.blondia@ua.ac.be (C. Blondia).

0166-5316/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.peva.2004.01.002

arriving packet. This situation is equivalent to the infinite node assumption and allows a station with backlogged packets to compete with itself.
3. A single error-free channel provides immediate—that is, at the end of the slot—binary (collision or not) or ternary (collision, success or empty) feedback.
4. If two or more stations transmit simultaneously, then there is a collision, meaning that the transmissions interfere destructively so that none succeeds.
5. Time is *slotted* and may be considered discrete. Users are synchronized with respect to the time slots. Each slot has a fixed duration equal to the time required to transmit a packet.

A number of algorithms belonging to the class of the CTMV type have been studied with some of the assumptions weakened. For instance, Polyzos and Molle [11] have considered finite population models for the grouped access strategy, which they refer to as window access. In case of a finite population, one generally assumes that the new arrivals occur according to a Bernouilli process instead of a Poisson process (in which case the number of arrivals in consecutive slots is still independent). Kessler, Seri and Sidi [12,13] have relaxed the third and fourth assumption and studied the performance of splitting algorithms in noisy channels with memory and Markovian capture. Many researchers have also considered different types of feedback, e.g., "success–failure" and "something–nothing", and early/delayed feedback. A comprehensive overview of most of the extensions made to a non-standard environment can be found in [9, Section 6].

What is apparent from this overview is that almost all researchers assume Poisson arrivals, except for some of the results on blocked access algorithms and a limited number of finite population studies that consider Bernouilli arrivals. This might seem like an obvious choice, especially in case of an infinite population, because the traffic generated by a very large population with independent users approaches a Poisson process. Nevertheless, studying the performance of an algorithm with an infinite population under a broad set of arrival processes might be very useful because such an infinite population model is a pessimistic estimate for a finite population. Thus, we can further extend the theoretical foundation of algorithms of the CTMV type by proving that these algorithms have good stability characteristics in such an environment. In 1998, during the 50th birthday of the IEEE Transactions on Information Theory Society, there was a survey article by Ephremides and Hajek [14] that stated that the union between information theory and communication networks has been only partially successful. The following statement on this unconsummated union was made: "The principle reason for this failure is twofold. First, by focusing on the classic point-to-point, source-channel-destination model of communication, information theory has ignored the bursty nature of real sources. Early on there seemed to be no point in considering the idle periods of source silence or inactivity. However, in networking, source burstiness is the central phenomenon that underlies the process of resource sharing for communication." It is the bursty nature of the arrivals that separates D-BMAPs from Poisson arrivals.

In this paper we leave the last four above-mentioned assumptions unchanged, instead we greatly relax the assumption made on the arrival process. That is, instead of assuming Poisson arrivals with a mean rate λ, we consider a rich class of arrival processes commonly known as discrete-time batch Markovian arrival processes (D-BMAPs). This class of arrival processes is known to lend itself very well to modeling bursty and correlated arrival processes commonly arising in computer and communication applications [15–23]. The aim of this work is to demonstrate that the good efficiency characteristics of RASs of the CTMV type with blocked or grouped access (see Section 2 for definitions) remain valid when we replace the Poisson arrival process by a general D-BMAP arrival process. Earlier work [24,25] focussed on RASs with free access and D-BMAP arrivals.

1.1. Maximum stable throughput or efficiency

A lot of attention has gone into determining the maximum stable throughput—which is sometimes called the efficiency or capacity—of random access systems. In the standard model, that is, under assumptions 1–5, the maximum stable throughput can be thought of as the highest possible arrival rate λ, denoted as λ_{crit}, for which each packet is successfully transmitted with a finite delay with probability one. For the blocked access algorithms—which are sometimes called gated access algorithms—the maximum stable throughput is generally known with sufficient precision for any practical purpose (meaning that an interval of size 0.001 or less is known to hold λ_{crit}). For most free and grouped access schemes these values are known exactly.

If we relax the first assumption and allow the new arrivals to occur according to a D-BMAP process, we define the maximum stable throughput as follows. A D-BMAP process is characterized by an infinite set of matrices B_n. For each primitive D-BMAP—see Section 3 for its definition—we can easily calculate the mean arrival rate λ, that is, the expected number of new packets generated in a time slot. For each RAS considered, the set of all primitive D-BMAPs $\mathbf{S}$ can be subdivided into two subsets $\mathbf{S}_1$ and $\mathbf{S}_2$, with $\mathbf{S}_1 \cup \mathbf{S}_2 = \mathbf{S}$ and $\mathbf{S}_1 \cap \mathbf{S}_2 = \emptyset$, such that the RAS is stable if and only if the new arrivals are generated according to a D-BMAP belonging to $\mathbf{S}_1$. From now on, unless otherwise stated, we simply refer to the set of all primitive D-BMAPs as the set of all D-BMAPs.

Ideally, there exists a λ_{crit} such that $\lambda < \lambda_{\text{crit}}$ if and only if the D-BMAP belongs to $\mathbf{S}_1$. In such case we define λ_{crit} as the maximum stable throughput. Unfortunately, such a λ_{crit} does not seem to exist. However, if we exclude a trivial subset of the D-BMAPs, we find, for the blocked access algorithms, a λ_{crit} and a δ small such that all D-BMAPs with $\lambda < \lambda_{\text{crit}} - \delta$ are a part of $\mathbf{S}_1$ and all D-BMAPs with $\lambda > \lambda_{\text{crit}} + \delta$ are a part of $\mathbf{S}_2$. For the grouped access algorithms we will show that there exists a $\lambda_{\min}$ and a $\lambda_{\max}$ such that all D-BMAPs with $\lambda < \lambda_{\min}$ belong to $\mathbf{S}_1$ and all D-BMAPs with $\lambda > \lambda_{\max}$ belong to $\mathbf{S}_2$. The difference between $\lambda_{\min}$ and $\lambda_{\max}$ depends on the length of the grouping interval Δ—the length of this interval is sometimes referred to as the window size. Moreover, the difference between $\lambda_{\min}$ and $\lambda_{\max}$ decreases rapidly as Δ increases. We also prove that the value of $\lambda_{\min}$, respectively, $\lambda_{\max}$, can hardly be increased, respectively, decreased, meaning that we can find a δ small such that there exits a D-BMAP with an arrival rate $\lambda_{\min} + \delta$ that is part of $\mathbf{S}_2$ and a D-BMAP with an arrival rate $\lambda_{\max} - \delta$ that is part of $\mathbf{S}_1$.

In order to get an idea of the statistical nature of the arrival processes that result in the worst and the best stability results, we examine a number of simple subclasses of the D-BMAPs. To achieve this, we have developed a procedure that allows us to study the stability of each of the RASs considered under any primitive D-BMAP arrival process. These explorations indicate that the stability results are the poorest under bursty and correlated arrival processes. Nevertheless, the obtained results are, in general, not far below the well known Poisson results, indicating that the good efficiency of the blocked and grouped access algorithms is maintained under D-BMAP arrivals. We believe that this result contributes to the theoretical foundation of algorithms of the CTMV type.

The paper is organized as follows. Section 2 presents the conflict resolution algorithms (CRAs) and channel access protocols (CAPs) considered. In Section 3 we introduce the discrete-time batch Markovian arrival processes (D-BMAPs). Section 4 focusses on the blocked access algorithms, whereas Section 5 evaluates the grouped access schemes. Conclusions are drawn in Section 6.

2. Specification of the random access systems

Random access systems are formed by combining a conflict resolution algorithm with a channel access protocol. A CAP states the rules for the first transmission attempt of a newly arrived packet, whereas the CRA indicates how collisions are resolved. We consider the following three well known channel access protocols:

1. *Blocked access.* After an initial collision of n stations, all new packets postpone their first transmission attempt until the n initial stations have resolved their collision. The time elapsed from the initial collision until the point where the n stations have transmitted successfully, and are aware of this, is called the collision resolution interval (CRI). Suppose that m new packets are generated during the CRI. Then, a new CRI starts (with m participants) when the previous CRI (with n stations involved) ends. In conclusion, when the blocked access mode is used new arrivals are blocked until the CRI during which they arrived has ended. They will participate in the next CRI.
2. *Grouped access.* All packets are subdivided into an infinite number of groups based on their arrival time, that is, the arrival-time axis is subdivided into an infinite number of fixed length intervals of size Δ and a packet belongs to group i if its arrival time is part of the ith interval (the left boundary of the ith interval coincides with the right boundary of the $i-1$th). Packets part of group i attempt a first transmission if all prior groups, i.e., group 1 to $i-1$, have been resolved and if the ith grouping interval has ended (that is, if the first grouping interval starts at time $t=0$, the current time must be equal to or larger than $t=i\Delta$). Hence, a possible collision of the $i-1$th group is resolved before group i packets are allowed to transmit.
3. *Free access.* New packets are transmitted immediately at the beginning of the next slot following their arrival. RASs with free access CAP are not considered within this paper. Results on free access schemes with D-BMAP arrivals can be found in [24,25]; therefore, we do not consider this CAP here.

The blocked access protocol is sometimes referred to as the gated access protocol, whereas the grouped access protocol is also known as the simplified window protocol [9]. The following two conflict resolution algorithms are considered in this paper:

1. *The basic Q-ary CTM algorithm.* Consider an arbitrary time slot t. Stations are allowed to transmit in slot t whenever their *index* for slot t is equal to one. Hence, new packets that are allowed to make their first transmission attempt in time slot t by the CAP initialize their index for slot t at one. The index values are assigned as follows. Whenever $n \geq 2$ stations collide in slot t, each station involved flips a "Q-sided coin" with values $1, \ldots, Q$ (the Q-ary coins need not to be fair, although we focus on fair coins unless otherwise stated). This splits the set of n colliding stations into Q subsets. To each of these subsets, we assign as an index the value which was flipped (some subsets may be empty). Stations part of the ith subset set their index for slot $t+1$ equal to i. Stations that were not involved in the collision at slot t, but who have an index for slot t equal to $i > 1$, set their index for slot $t+1$ at $i+(Q-1)$. If, on the other hand, the outcome of slot t is "no collision", all stations who have an index $i > 1$ for slot t, set their index for slot $t+1$ at $i-1$ (if there was a single station with an index for slot t equal to one, then a successful transmission occurred and the station deletes the index). The basic Q-ary CTM algorithm distinguishes only "collisions" and "no collisions", therefore, binary feedback suffices.
2. *The modified Q-ary CTM algorithm.* The mechanism used to resolve the collisions is the same as in the basic algorithm, except that the algorithm tries to improve the basic algorithm using ternary

feedback, namely, a successful transmission is distinguished from an idle slot. If, after a collision, the next $Q-1$ slots turn out to be empty, then the next slot must hold a collision when the basic algorithm is applied (because all the stations involved in the last collision must have chosen the Qth subset). This *doomed* slot can be skipped by having all stations immediately act as if it had occurred.

A detailed description of both algorithms, examples of their transmission process, their motivation, and many of their properties can be found in a variety of papers, for example [2,4,6,9,10].

Combining the three CAPs with the two CRAs results in six different random access systems. We will study the stability of four of these algorithms when the Poisson arrivals of the standard model are replaced by D-BMAP arrivals (that is, the blocked and grouped access RASs).

3. Discrete-time batch Markovian arrival processes (D-BMAPs)

The D-BMAP is the discrete time counterpart of the BMAP [26,27] and was first introduced in [28]. Formally, a D-BMAP is defined by an infinite set of positive $l \times l$ matrices $(B_n)_{0\leq n<\infty}$, with the property that

$$B = \sum_{n=0}^{\infty} B_n \tag{1}$$

is a transition matrix. By definition the Markov chain J_t associated with B and having $\{i; 1 \leq i \leq l\}$ as its state space, is controlling the actual arrival process as follows. Suppose J is in state i at time t. By going to the next time instance $t+1$, there occurs a transition to another or possibly the same state, and a batch arrival may or may not occur. The entries $(B_n)_{i,j}$ represent the probability of having a transition from state i to j and a batch arrival of size n. So, a transition from state i to j without an arrival will occur with probability $(B_0)_{i,j}$. Define by X_t the number of arrivals generated at time t.

We assume that the transition matrix B is an aperiodic irreducible matrix [29]. Notice that the stability of an algorithm under reducible D-BMAP arrivals depends on the initial state at time $t=0$, whereas this is not the case for irreducible D-BMAPs. Aperiodic irreducible matrices are often referred to as primitive matrices. Thus, whenever we refer to a primitive D-BMAP we mean to say that its transition matrix B is aperiodic and irreducible. For B primitive the Markov chain J_t has a unique stationary distribution. Let β be the stationary probability vector of the Markov chain J_t, i.e., $\beta B = \beta$ and $\beta e = 1$ with e a column vector of 1's. The mean arrival rate $\lambda = E[X_t]$ of the D-BMAP $(B_n)_n$ is given by

$$\lambda = \beta\left(\sum_{n=1}^{\infty} nB_n\right)e. \tag{2}$$

Due to the Ergodic theorem for primitive Markov Chains [29] we have

$$\lim_{L\to\infty} \frac{E[\sum_{i=0}^{L-1} X_{t+i}|J_t = j]}{L} = \lambda \tag{3}$$

for $1 \leq j \leq l$. D-BMAPs for which $B_n = 0$, for $n \geq 2$, are referred to as discrete time Markovian arrival processes (D-MAPs). Many properties like the autocorrelation function or the index of dispersion for count (IDC) can be found in [15,28,30]. Another important characteristic of D-BMAPs is that any finite superposition of D-BMAPs is again a D-BMAP. A few very simple D-BMAP subclasses, used for later discussions, are presented in Appendix A. The D-BMAPs used to model nowadays communication network traffic are, of course, much more complicated and often have 50 or more states, e.g., in [22,23] an individual video source (based on a Bond movie trace) is modeled by 65 states.

4. Analysis of the RASs with blocked access

It is well known [1,2,4] that if the input traffic is Poisson with a mean λ and if a CRA has an expected running time $T(n)$, to resolve n participants, then the corresponding RAS with blocked access is stable for all $\lambda < \liminf n/T(n)$; unstable for $\lambda > \limsup n/T(n)$. The expression for $T(n)$ depends upon the CRA. Therefore, it is sufficient to study the asymptotic behavior of $n/T(n)$ for n to infinity in order to determine the stability of a blocked access algorithm under Poisson input. This behavior is, obviously, independent of the arrival process. It is fairly easy to generalize this to the following theorem:

Theorem 1. *A RAS with blocked access, corresponding to a CRA that resolves conflicts of multiplicity n in an expected time $T(n)$, is stable under primitive D-BMAP traffic if*

1. $\lambda < \liminf n/T(n)$

and unstable if

1. $\lambda > \limsup n/T(n)$,
2. $n > 1$ *and* $1 \leq i, j \leq l$ *exist such that* $(B_n)_{i,j} \neq 0$, *that is, the D-BMAP is not a D-MAP.*

The stability part of this theorem is straightforward as Cidon and Sidi [4, Theorem 8] have proven the following theorem. Let $\sigma = \liminf n/T(n)$ and let $N_{t,t+L}$ be the number of new packets arriving to the system in the interval $(t, t + L]$. Then, the system is stable if there exists a $\delta > 0$ and an L^* such that $E[N_{t,t+L}] < (\sigma - \delta)L$ for all t and $L > L^*$. From Eq. (3) we know that the expected number of arrivals of a primitive D-BMAP that occur in an interval of length L approaches λL as L approaches infinity, where λ is the mean arrival rate (whichever the state at the start of the interval is). Because the number of states of the D-BMAP l is finite, we find that for any $\epsilon > 0$ there exists an L^* such that $E[N_{t,t+L}] < (\lambda + \epsilon)L$ for all t and $L > L^*$. Thus, whenever $\lambda + \epsilon < \sigma$, it suffices to choose δ between 0 and $(\sigma - \lambda) - \epsilon > 0$ to fulfill the required equation.

We did not manage to find a formal proof of a result similar to that of Cidon and Sidi, in the existing literature, that allows a short proof of the instability condition. Therefore, we have extended Massey's argument [2] for the Poisson result to D-BMAPs. Because it requires a number of Epsilons we give a rigorous proof in Appendix B.

Recall, the expression $T(n)$ is the expected time required by the CRA to resolve a collision of n participants. Thus, it does not depend on the arrival process. As a result, the all well known efficiency results for the CRAs presented in Section 2 (with fair and biased coins) and Poisson arrivals are also valid for D-BMAP arrivals. For example, if the basic binary CTM is used as the CRA, we have a stable system for any D-BMAP arrival process if the arrival rate $\lambda < 0.3464$. Similarly, the system is always unstable if $\lambda > 0.3471$. This immediately closes the discussion of the blocked access algorithms. Next, we consider grouped access systems, which are more difficult to analyze.

5. Analysis of the RASs with grouped access

We restrict ourselves to the case where Δ, the length of the grouping interval, is an integer value and Q, the splitting factor, is equal to 2. The arguments presented in this section can easily be extended to the case where $Q > 2$. Most of the researchers working on RASs with grouped access focus on CRAs with $Q = 2$ because the basic idea behind grouping is to form groups with a small number of contenders (ideally, [slightly more than] one in each group [10]). Therefore, it is important to have a CRA that performs well for groups with very few contenders. The basic Q-ary CTM algorithm is known to perform best in resolving groups with $n \leq 3$ participants for $Q = 2$. The same can be said about the modified CTM algorithm for $n \leq 7$ [6]. Moreover, the resulting RASs achieve the highest maximum stable throughput when the length of the grouping interval Δ is optimized (under Poisson arrivals). The fact that the binary CRAs are the best in resolving small groups does not depend on the statistics of the arrival process; hence, we also expect the binary CRAs of Section 2 to perform best under BMAP input traffic if Δ is small (except for some of the more artificial processes belonging to the class of the D-BMAPs).

We start by proving that a RAS that uses a grouping algorithm as its CAP is stable under primitive D-BMAP traffic if the expected time required to resolve an arbitrary group $E[G]$ is smaller than Δ and unstable if $E[G] > \Delta$. Afterwards we indicate how to obtain tight upper and lower bounds on $E[G]$. For the two CRAs introduced in Section 2, with $Q = 2$ and fair coins, these bounds allow us to determine the maximum stable throughput with sufficient accuracy.

5.1. A stability condition for D-BMAP input

A RAS that applies a grouping strategy as its CAP under primitive D-BMAP input traffic can be seen as a queue with the following characteristics. We assumed that Δ is an integer. The customers arriving in the queue correspond to the groups produced by the algorithm. Thus, every Δ time slots a new customer arrives—that is, we have a deterministic arrival process. The queue has an infinite waiting room and a single server. A customer is said to be of type j with $1 \leq j \leq l$ if the state of the input D-BMAP at the start of the corresponding grouping interval was j. The group types are therefore determined by a primitive discrete time Markov chain with transition matrix B^{Δ}, where B is the transition matrix of the input D-BMAP, i.e., $B = \sum_n B_n$. Thus, if the type of customer n is i than the type of customer $n + 1$ is j with probability $(B^{\Delta})_{i,j}$. The service time of a customer—that is, the time required to resolve the corresponding group—depends upon the type of the customer. Hence, the service time of a customer of type j is t with some probability $G_j(t)$. Meaning, the service time distribution of a customer depends on the state of the D-BMAP at the start of the corresponding grouping interval. For l the number of states of the D-BMAP, or else the number of customer types, equal to one the above-mentioned queue reduces to a $D/G/1$ queue and such a queue is known to be stable for $\rho < 1$ [31]. This condition is obviously equivalent to $E[G] < \Delta$. Another way to prove that $E[G] < \Delta$ is a sufficient condition for stability when $l = 1$ is to use the stability lemma of Pakes [10, p. 264]. For $l > 1$, things are slightly more complicated.

The arrival process of our queue can be seen as a special case of the discrete time version of a Markovian arrival process with marked arrivals [32,33], denoted as $MMAP[K]$. Such a Markov arrival process is characterized by a set of $m \times m$ matrices M_0 and M_J with J a string of integers, where each integer is part of $[1, K]$. The i, jth element of M_J, with $J = j_1, \dots j_n$, $n > 0$, represents the probability that a transition is made from state i to j and that n arrivals occur. The type of these n arrivals is as follows: the kth customer that arrives is a customer of type j_k. The matrix M_0 characterizes the transitions when no new arrivals occur. For $K = 1$ the $MMAP[K]$ arrival process reduces to a D-BMAP arrival process (if we identify the matrix B_n with M_J where J is a string that consists of n ones). It is easily seen that the arrival process of our queue of interest is actually a $MMAP[K]$ process with $K = l$ and $m = \Delta l$. The matrix M_0 has the following form:

$$M_0 = \begin{pmatrix} 0 & I & 0 & \dots & 0 & 0 \\ 0 & 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & I & 0 \\ 0 & 0 & 0 & \dots & 0 & I \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \tag{4}$$

where I is the $l \times l$ unity matrix. The matrices M_k, $1 \leq k \leq l$, obey the following equation:

$$M_k = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ B^{\Delta}(k) & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \tag{5}$$

where $B^{\Delta}(k)$ is obtained from B^{Δ} by keeping the kth column of the matrix B^{Δ} and setting all other elements to zero. The entries of the matrices M_J with J a string of length 2 or more are all zero. Now that we know that the input is a $MMAP[K]$, the queue we are interested in is a special case of a $MMAP[K]/G[K]/1$ queue.

He [32] has shown that a $MMAP[K]/G[K]/1$ queue with a work conserving service discipline is positive recurrent if $\rho = \lambda_1 E[G_1] + \cdots + \lambda_K E[G_K] < 1$ and is transient if $\rho > 1$, where λ_i corresponds to the average number of type i customers arriving in the queueing system (per time unit) and $E[G_i]$ to the expected service time of a type i customer. In our case, the vector $(\lambda_1, \dots, \lambda_K)$ is nothing but the vector β/Δ, where $\beta B = \beta$ and $\beta e = 1$ (because β is also the invariant vector of B^{Δ}). Thus, $\Delta\rho$ is equal to the expected service time of an arbitrary customer—that is, the expected time required to resolve an arbitrary group. This proves that we get a stable, respectively, unstable, system whenever $E[G] < \Delta$, respectively, $E[G] > \Delta$. Notice, that the stability of a grouped access RAS corresponds to saying that the number of waiting groups does not grow to infinity.

5.2. Tight bounds on $E[G]$

Following Massey's approach [2], it is fairly straightforward to obtain a tight upper and lower bound on $E[G]$ when the basic or modified binary CTM algorithm is used to resolve the groups. First, we determine the probability that a group contains n contenders—that is, n arrivals occur in the corresponding interval of length Δ. The probability that the state of the D-BMAP is j, $1 \leq j \leq l$, at the start of a grouping interval is equal to β_j, where β_j is the jth component of the stationary vector β corresponding to the input D-BMAP because β is also an invariant vector of B^{Δ}. The probability of having n arrivals in an interval of length Δ provided that the state is j at the start of the interval, say $P_j(n)$, is easily computed as follows. Define the matrices $B_{n,i}$, $i > 1, n \geq 0$, as

$$B_{n,i} = \sum_{j=0}^{n} B_{j,i-1} B_{n-j} \tag{6}$$

with $B_{n,1}$ equal to B_n. Then, $P_j(n)$ is found as the jth component of $B_{n,\Delta} e$. Therefore, the probability that a group contains n arrivals, say $P(n)$, is nothing but $\sum_{j=1}^{l} \beta_j P_j(n)$.

The expected time required to resolve an arbitrary group $E[G]$ is found as $E[G] = \sum_n P(n)T(n)$, where $T(n)$ represents the expected time required by the CRA to resolve a set of n contenders. Massey [2] obtained the following upper and lower bounds on $T(n)$ for the basic and modified binary CTM algorithm. In order to distinguish both algorithms we write $T_{\mathrm{b}}(n)$ for the expected time required by the basic binary CTM algorithm and $T_{\mathrm{m}}(n)$ as the expected time required by the modified binary CTM algorithm. For the basic binary CTM algorithm we have

$$T_{\mathrm{b}}(n) \leq a_1 n - 1 + 2\delta_{0,n} + (2 - a_1)\delta_{1,n} + (6 - 2a_1)\delta_{2,n} + (\tfrac{26}{3} - 3a_1)\delta_{3,n} \tag{7}$$

with $a_1 \approx 2.8867$ and $\delta_{i,j} = 0$ if $i \neq j$ and 1 if $i = j$. Moreover,

$$T_{\mathrm{b}}(n) \geq a_2 n - 1 + 2\delta_{0,n} + (2 - a_2)\delta_{1,n} + (6 - 2a_2)\delta_{2,n} + (\tfrac{26}{3} - 3a_2)\delta_{3,n} \tag{8}$$

with $a_2 \approx 2.8810$. Whereas for the modified binary CTM we find

$$T_{\mathrm{m}}(n) \leq b_1 n - 1 + 2\delta_{0,n} + (2 - b_1)\delta_{1,n} + (\tfrac{11}{2} - 2b_1)\delta_{2,n} + (8 - 3b_1)\delta_{3,n} \tag{9}$$

with $b_1 \approx 2.6651$ and

$$T_{\mathrm{m}}(n) \geq b_2 n - 1 + 2\delta_{0,n} + (2 - b_2)\delta_{1,n} + (\tfrac{11}{2} - 2b_2)\delta_{2,n} + (8 - 3b_2)\delta_{3,n} \tag{10}$$

with $b_2 \approx 2.6607$. If we calculate $E[G] = \sum_n P(n)T(n)$ and replace $T(n)$ by its lower, respectively, upper, bound we obtain a lower, respectively, upper, bound on $E[G]$. Whenever the lower bound is larger than Δ we know from Section 5.1 that the RAS is unstable, whereas if the upper bound is smaller than Δ we have a stable RAS. For those cases that produce an upper bound larger than Δ and a lower bound that is smaller we know nothing about the stability.

5.3. Numerical results

In this section we will show that there exists a $\lambda_{\min}$ and a $\lambda_{\max}$ for each RAS considered such that the RAS is stable for all primitive D-BMAPs with $\lambda < \lambda_{\min}$ and unstable for all primitive D-BMAPs with $\lambda > \lambda_{\max}$ The difference between $\lambda_{\min}$ and $\lambda_{\max}$ depends on the length of the grouping interval Δ. We also prove that the value of $\lambda_{\min}$, respectively, $\lambda_{\max}$, can hardly be increased, respectively, decreased, meaning that there exists a δ small such that there exits a D-BMAP with an arrival rate $\lambda_{\min} + \delta$ for which the RAS is unstable and a D-BMAP with an arrival rate $\lambda_{\max} - \delta$ for which the RAS is stable.

Theorem 2. *A RAS with grouped access, that uses the basic binary CTM algorithm as its CRA, is stable under primitive D-BMAP traffic if*

$$\lambda < \frac{1}{a_1}\left(1 - \frac{1}{\Delta}\right) \approx 0.3464\left(1 - \frac{1}{\Delta}\right), \tag{11}$$

where $a_1 \approx 2.8867$, *and unstable if*

$$\lambda > \frac{1}{a_2}\left(1+\frac{a_2-1}{\Delta}\right) \approx 0.3472\left(1+\frac{1.8810}{\Delta}\right), \tag{12}$$

where $a_2 \approx 2.8810$.

Proof. Using Eq. (7) and $T_\mathrm{b}(n) \le a_1 n$ for $n > 0$, we have $E[G] = \sum_{n\ge 0} P(n)T_\mathrm{b}(n) \le \sum_{n>0} a_1 nP(n) + P(0) = a_1\lambda\Delta + P(0)$. Hence, $\lambda < (1/a_1)(1 - P(0)/\Delta)$ is a sufficient condition for having a stable algorithm. Obviously, $P(0) < 1$ if $\lambda > 0$. As a result we have stability if $\lambda < (1/a_1)(1 - 1/\Delta)$ for any primitive D-BMAP input traffic. Using Eq. (8), we have $E[G] = \sum_n P(n)T_\mathrm{b}(n) \ge a_2 \sum_n nP(n) - \sum_n P(n) + 2P(0) + (2-a_2)P(1) + (6-2a_2)P(2) + (26/3 - 3a_2)P(3) \ge a_2\lambda\Delta - 1 + (2-a_2) = a_2\lambda\Delta - (a_2 - 1)$. Thus, the grouping algorithm that uses the basic binary CTM algorithm is unstable if $\lambda > (1/a_2)(1 + (a_2-1)/\Delta)$. □

From the proof of Theorem 2 it should be clear that only the presence of empty groups might reduce the maximum stable throughput below $1/a_1 \approx 0.3464$. Similarly, we can prove the following theorem for the modified binary CTM algorithm:

Theorem 3. *A RAS with grouped access, that uses the modified binary CTM algorithm as its CRA, is stable under primitive D-BMAP traffic if*

$$\lambda < \frac{1}{b_1}\left(1-\frac{1}{\Delta}\right) \approx 0.3752\left(1-\frac{1}{\Delta}\right), \tag{13}$$

where $b_1 \approx 2.6651$, *and unstable if*

$$\lambda > \frac{1}{b_2}\left(1+\frac{b_2-1}{\Delta}\right) \approx 0.3758\left(1+\frac{1.6607}{\Delta}\right), \tag{14}$$

where $b_2 \approx 2.6607$.

Numerical results for different values of Δ are presented in Table 1. Obviously, for Δ large we find that the interval reduces to $[1/a_1, 1/a_2]$, respectively $[1/b_1, 1/b_2]$. Both these intervals are rather small and contain the maximum stable throughput of the corresponding RASs with blocked access (see Section 4). Thus, whether the basic, respectively modified, binary CTM algorithm uses a blocked access strategy or a grouping strategy (with Δ large) makes little difference as far as the stability under primitive D-BMAP traffic is concerned. The fact that these blocked access and grouped access, with Δ large, algorithms perform similar was already known for Poisson input traffic [9]. The following property proves that the size of the intervals in Table 1 can hardly be decreased.

Property 1. *Consider a RAS with grouped access, that uses the basic binary CTM algorithm as its CRA. Then, for each* $\epsilon > 0$ *it is possible to find a primitive D-BMAP with an arrival rate* $1/a_2(1 - 1/\Delta) + \epsilon$ *for which the RAS is unstable, and a primitive D-BMAP with an arrival rate* $1/a_1(1 + (a_1 - 1)/\Delta) - \epsilon$ *for which the RAS is stable.*

Table 1
Numerical values for the four quantities in Theorems 2 and 3

Δ	Basic binary		Modified binary	
1	0	1	0	1
2	0.1732	0.6736	0.1876	0.6879
3	0.2309	0.5647	0.2501	0.5839
4	0.2598	0.5103	0.2814	0.5319
5	0.2771	0.4777	0.3002	0.5007
10	0.3118	0.4124	0.3377	0.4383
20	0.3291	0.3797	0.3565	0.4070
50	0.3395	0.3602	0.3677	0.3883
100	0.3430	0.3536	0.3715	0.3821
1000	0.3461	0.3478	0.3748	0.3765
10000	0.3464	0.3472	0.3752	0.3759
∞	0.3464	0.3471	0.3752	0.3758

Proof. For $\Delta = 1$ the proof is trivial. For $\Delta \geq 2$, consider the D-BMAP with $(2\Delta - 1)$ states and the following matrix B_0, where I_j represents the $j \times j$ unity matrix

$$B_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\Delta-2} & 0 & 0 \\ 1-\dfrac{1}{p} & 0 & 0 & \dfrac{1}{p} & 0 \\ 0 & 0 & 0 & 0 & I_{\Delta-2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{15}$$

The symbol 0 in the matrix B_0 is sometimes used for a vector of zeros, a matrix of zeros or simply a zero. The dimension of each of these 0 entries should be clear from the fact that the D-BMAP has $2\Delta - 1$ states. Notice that p is a parameter that we fix later on. The matrix B_1 is a matrix with all its entries equal to zero, except for $(B_1)_{1,2} = 1$. All the other B_i matrices with $i > 1$ are zero, except for B_{m}, where m is a parameter. As for the matrix B_1, the matrix B_{m} has only one entry differing from zero: $(B_{\mathrm{m}})_{2\Delta-1,1} = 1$. Thus, our D-BMAP is completely determined once we fix the parameters m, p and Δ. It is fairly straightforward to prove that this D-BMAP is a primitive one (actually, it follows from the fact that the greatest common divisor of Δ and $2\Delta - 1$ equals one). Moreover, the invariant vector β can be obtained explicitly as a function of m, p and Δ. With this vector one finds the following arrival rate λ and probabilities $P(n)$:

$$\lambda = \frac{m+p}{\Delta p + \Delta - 1}, \qquad P(1) = \frac{\Delta p - 1}{\Delta p + \Delta - 1}, \qquad P(m) = \frac{\Delta - 1}{\Delta p + \Delta - 1},$$
$$P(m+1) = \frac{1}{\Delta p + \Delta - 1},$$

the other probabilities $P(n)$ are equal to zero. With these probabilities, and Eq. (8), we can find the following upper bound for the expected service time $E[G]$ of an arbitrary group:

$$\frac{\Delta(p-1) + a_1 m\Delta + a_1 - 1}{\Delta p + \Delta - 1}. \tag{16}$$

Provided that this upper bound is smaller than Δ we have a stable algorithm for this D-BMAP input traffic. This stability condition can be rewritten as

$$\frac{1}{\Delta - 1}\left(\frac{a_1 - 1}{\Delta} + a_1 m - \Delta\right) < p. \tag{17}$$

Hence, if we choose p as $\epsilon > 0$ plus the left-hand side of Eq. (17) we have a stable algorithm. Having done this the arrival rate λ can be written as

$$\lambda = \frac{m(a_1 + \Delta - 1) + (a_1 - 1)/\Delta - \Delta + \epsilon(\Delta - 1)}{a_1 \Delta m + a_1 - 2\Delta + \epsilon\Delta(\Delta - 1)}. \tag{18}$$

This value can be optimized by letting m go to infinity, in which case λ approaches $1/a_1(1+(a_1-1)/\Delta)$. This proves half of the theorem. The construction of the D-BMAP used to prove the other half is very similar and aims at finding a D-BMAP that either generates empty groups or very large groups. □

Using the same primitive D-BMAP arrival processes we can prove the following property as well.

Property 2. *Consider a RAS with grouped access, that uses the modified binary CTM algorithm as its CRA. Then, for each* $\epsilon > 0$ *it is possible to find a primitive D-BMAP with an arrival rate* $1/b_2(1-1/\Delta)+\epsilon$ *for which the RAS is unstable, and a primitive D-BMAP with an arrival rate* $1/b_1(1+(b_1-1)/\Delta)-\epsilon$ *for which the RAS is stable.*

Numerical results for different values of Δ are presented in Table 2. If we compare these values with Table 1, it is clear that the values in Theorems 2 and 3 can hardly be increased, respectively, decreased, that is, the difference is at most 0.0007 for any $\Delta \geq 1$. As with many of the arrival processes belonging to the class of the D-BMAPs, the processes used to prove Properties 1 and 2 are quite artificial and are not very likely to be used as a model for the input traffic of a communication network. To get an idea of the statistical properties that influence the maximum stable throughput, we study the maximum stable throughput of the RASs for a number of very simple D-BMAP arrival processes described in Appendix A.

Table 2
Numerical values for the four quantities in Properties 1 and 2

Δ	Basic binary		Modified binary	
1	0	1	0	1
2	0.1736	0.6732	0.1879	0.6876
3	0.2315	0.5643	0.2506	0.5835
4	0.2604	0.5098	0.2819	0.5314
5	0.2778	0.4771	0.3007	0.5002
10	0.3125	0.4118	0.3383	0.4377
20	0.3298	0.3791	0.3570	0.4077
50	0.3403	0.3595	0.3683	0.3877
100	0.3437	0.3530	0.3721	0.3815
1000	0.3469	0.3471	0.3755	0.3758
10000	0.3472	0.3465	0.3758	0.3753
∞	0.3472	0.3465	0.3758	0.3753

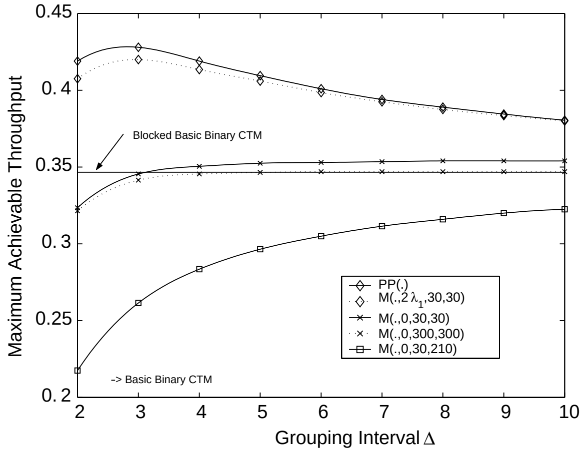

Fig. 1. The impact of Δ on the maximum stable throughput (basic).

5.3.1. *Two-state Markov modulated Poisson processes (TMMPP)*

We start with a discussion of the two-state Markov modulated Poisson processes described in Appendix A.3. As with the class of the primitive D-BMAPs, there exists no λ_{crit} such that all two-state Markov modulated Poisson process with a $\lambda < \lambda_{crit}$ result in a stable algorithm; $\lambda > \lambda_{crit}$ result in an unstable algorithm. However, if we fix the parameters a and b and make λ_2 a function of λ_1, such a λ_{crit} can be found. We refer to this λ_{crit} as the maximum stable throughput of the $M(\cdot, f(\lambda_1), a, b)$ process, where the dot indicates that the arrival rate of state 1 is the variable of the process.

Fig. 1, respectively, Fig. 2, compares the maximum stable throughput as a function of Δ $(2 \leq \Delta \leq 10)$ for a few TMMPPs when the basic, respectively, modified, binary CTM algorithm is combined with a grouping strategy. We denote x, a multiple of 0.0005, as the maximum stable throughput if the interval $[x, x + 0.0015]$ holds the maximum stable throughput of the arrival process considered. Both figures are almost identical, except that the modified scheme supports throughputs which are a few percentages higher.

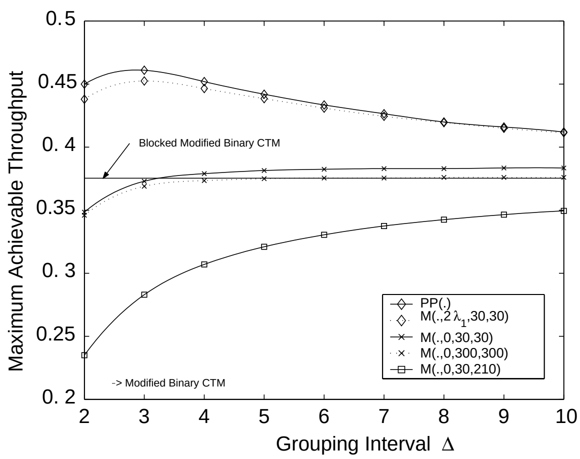

Fig. 2. The impact of Δ on the maximum stable throughput (modified).

A first conclusion drawn from both figures is that a rather serious degradation of the maximum stable throughput might occur if the burstiness of the arrival processes—that is, the variance of the number of arrivals in a time slot—increases, especially if Δ is very small. The reason for this is the presence of the empty groups (this follows from the proof of Theorem 2). Although the probability $P(0)$ of having an empty group does not decrease that rapidly when increasing Δ, the throughput degradation does disappear rather quickly. This is due to the fact that the throughput can be written as a weighted sum of the throughputs T_i, where T_i is the throughput associated with a collision resolution interval (CRI) that corresponds to a grouping interval starting in state i. We refer to such a CRI as a type i CRI. Denote C_i as the expected time required by the CRA algorithm to resolve a CRI of type i. Then, the weight w_i that corresponds to T_i equals

$$w_i = \frac{\beta_i C_i}{\sum_j \beta_j C_j}. \tag{19}$$

In the case of our $M(\cdot, 0, a, b)$ processes, $T_2 \approx 0$ because $\lambda_2 = 0$. Increasing Δ results in a decreasing $P(0)$. However, $P(0)$ decreases slowly (especially if b is large). Nevertheless, the maximum stable throughput recovers quickly when increasing Δ. This follows from the fact that w_2 decreases rapidly when increasing Δ. Indeed, we find that the expected number of contenders associated with a type 1 CRI increases rapidly as Δ increases; hence, C_1 increases rapidly. Whereas the expected number of contenders in a type 2 CRI remains close to zero (for $\Delta \ll b$); hence, C_2 remains small. This implies that the weight w_2 associated with $T_2 \approx 0$ decreases rapidly when Δ increases, which explains the rapid restoration of the maximum stable throughput when Δ is increased.

Figs. 1 and 2 also indicate that correlation is of lesser importance. For instance, the $M(\cdot, 0, 30, 30)$, the correlation function $r(k)$ of which decays as 0.9333^k, performs only slightly better than the $M(\cdot, 0, 300, 300)$, which has a correlation function $r(k)$ that decays as 0.9933^k. Moreover, the results for the $M(\cdot, 0, 3000, 3000)$ arrival process, which are not included in the figures, are almost identical to those of the $M(\cdot, 0, 300, 300)$ process. This comes as no surprise because the grouping mechanism breaks the correlation (Moreover, the order in which the groups are resolved is of no importance with respect to the efficiency, indeed, He's theorem used to obtain the stability condition in Section 5.1 is valid for any work conserving scheduling discipline.)

Notice, the maximum stable throughput under $M(\cdot, 0, 30, 210)$ input traffic is only a few percentages higher than $1/a_1(1 - 1/\Delta)$, respectively, $1/b_1(1 - 1/\Delta)$ (see Table 1). We can easily define a TMMPP for which the maximum stable throughput is even closer to $1/a_1(1 - 1/\Delta)$, respectively, $1/b_1(1 - 1/\Delta)$. For instance, the basic, respectively, modified, CTM algorithm with grouping has a maximum stable throughput under $M(\cdot, 0, 30, 3000)$ input traffic of $\approx$0.1770, respectively, $\approx$0.1915. The $M(\cdot, 0, 30, 3000)$ process is very bursty: the average sojourn time in the silent state is 3000 slots, whereas the average time in the active state is only 30 slots. Therefore, all the traffic is more or less concentrated in 1% of the grouping intervals of length Δ.

5.3.2. Erlang arrival process

As with the TMMPPs, there exists no λ_{crit} for all Erlang processes (see Appendix A.2 for a definition), however, fixing the parameter k results in a unique maximum stable throughput. Fig. 3 and 4 present the results for the Erlang arrival processes. As expected we get a higher maximum stable throughput if k is increased, i.e., if the process becomes more deterministic. Also, the results for the $\text{ER}(\cdot, 10)$ process are only a few percentages below $1/a_2(1 + (a_2 - 1)/\Delta)$, respectively, $1/b_2(1 + (b_2 - 1)/\Delta)$. For $k = 50$

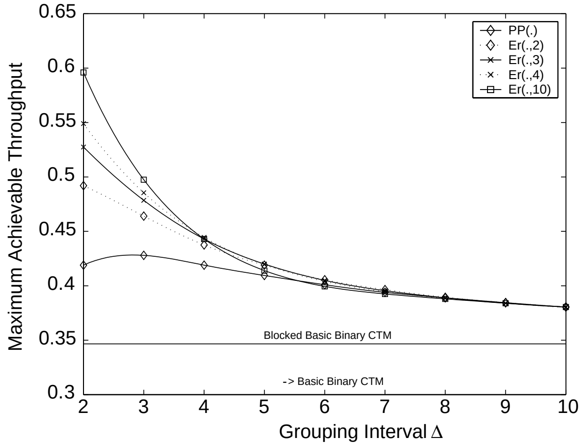

Fig. 3. The impact of Δ on the maximum stable throughput (basic).

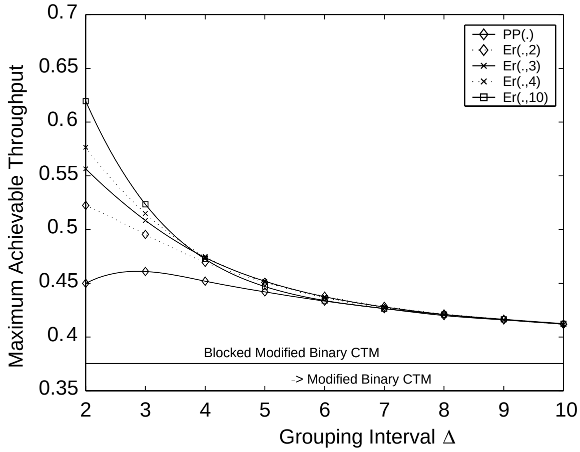

Fig. 4. The impact of Δ on the maximum stable throughput (modified).

we found a maximum stable throughput for $\Delta = 2$ of $\approx$0.623, respectively, $\approx$0.6415. It is easy to prove that the maximum stable throughput for $\Delta = 2$ converges to 0.625, respectively, 0.6429, as k approaches infinity. Erlang arrival processes with k large can be used to model constant bite rate (CBR) traffic sources in communication networks.

Similar figures can be obtained for other arrival processes belonging to the D-BMAP class.

6. Conclusions

This paper examines the maximum stable throughput of a random access system which uses Q-ary tree algorithms (where Q is the number of groups into which colliding users are split) of the Capetanakis–

Tsybakov–Mikhailov–Vvedenskaya type for an infinite population of identical users generating packets according to a discrete-time batch Markovian arrival process (D-BMAP). Blocked and grouped channel access protocols have been considered in combination with Q-ary collision resolution algorithms that exploit either binary (“collision or not”) or ternary (“collision, success or idle”) feedback. For the resulting RASs the corresponding maximum stable throughput is determined.

For the RAS with blocked access, it was shown that the well known stability results for the Poisson traffic also apply to D-BMAP arrival processes. Thus, for each CRA, there exists a λ_{crit} and a δ small such that the RAS with blocked access is stable under all primitive D-BMAPs with a mean arrival rate $\lambda < \lambda_{\text{crit}} - \delta$ and unstable if $\lambda > \lambda_{\text{crit}} + \delta$.

For the grouped access RAS it was shown that for each CRA there exists a $\lambda_{\min}$ and a $\lambda_{\max}$ such that the RAS with grouped access is stable for $\lambda < \lambda_{\min}$ and unstable for $\lambda > \lambda_{\max}$. The difference between $\lambda_{\min}$ and $\lambda_{\max}$ depends on the length of the grouping interval Δ. We have also proven that the value of $\lambda_{\min}$, respectively, $\lambda_{\max}$, can hardly be increased, respectively, decreased, meaning that there exists a δ small such that there exits a D-BMAP with an arrival rate $\lambda_{\min} + \delta$ for which the RAS is unstable and a D-BMAP with an arrival rate $\lambda_{\max} - \delta$ for which the RAS is stable. Additional numerical explorations, in Section 5.3, have indicated that the stability results are the worst if the arrival process is bursty, and to a lesser extend, highly correlated. However, provided that the length of the grouping interval Δ is not too small, i.e., $\Delta \geq 10$, the stability remains high, i.e., above 30%.

In general, it is concluded that the RASs considered maintain their good stability characteristics under the wide range of D-BMAP arrival processes—with the exception perhaps of grouped access RASs with Δ, the length of the grouping interval, small—thereby further extending the theoretical foundations of tree algorithms.

Acknowledgements

We would like to thank the reviewers for their valuable comments, which have improved the presentation of this paper.

Appendix A. Some D-BMAP subclasses

A.1. The discrete time Poisson process

The discrete time Poisson process is obtained by observing the continuous time Poisson process at the slot boundaries. Arrivals that occurred in the interval $(t, t+1]$ are now assumed to arrive on the boundary of slot t and $t+1$, i.e., at time $t+1$. We can model the discrete time Poisson process as a D-BMAP with a single state by letting $B_n = \mathrm{e}^{-\lambda}\lambda^n/n!$ for $n \geq 0$. For later reference, we abbreviate the Poisson process as $\mathrm{PP}(\lambda)$.

A.2. The discrete time Erlang process

We define the continuous time Erlang process as follows. The continuous time Erlang process has independent and identically distributed interarrival times that obey an Erlang distribution with parameters

k and λ_e (this λ_e is not to be confused with the arrival rate λ of the corresponding D-BMAP). Clearly, for $k = 1$ the Erlang process is reduced to the Poisson process. By observing the Erlang process at the slot boundaries we obtain the discrete time Erlang process (arrivals are assumed to occur on slot boundaries). The discrete time Erlang process can be modeled as a D-BMAP in the following way. Let $\gamma_n = \mathrm{e}^{-\lambda}\lambda^n/n!$, $n \geq 0$, and let B_n, $n \geq 0$, be $k \times k$ matrices defined as

$$(B_n)_{i,j} = \gamma_{nk+j-i}, \quad nk \geq j - i, \tag{A.1}$$

$$(B_n)_{i,j} = 0, \quad nk < j - i. \tag{A.2}$$

The arrival rate λ of this D-BMAP is λ_e/k. For later reference, we abbreviate the Erlang k process as $\mathrm{ER}(\lambda_e, k)$.

A.3. The two-state discrete time Markov modulated Poisson process

Two-state discrete time Markov modulated Poisson processes are characterized by two parameters λ_1, λ_2 and a 2×2 matrix T. The process will generate arrivals according to a Poisson process with a mean rate λ_i when the current state is i. Transitions from one state to another can occur at the end of each time slot according to a 2×2 transition matrix T:

$$T = \begin{pmatrix} 1 - \dfrac{1}{a} & \dfrac{1}{a} \\ \dfrac{1}{b} & 1 - \dfrac{1}{b} \end{pmatrix}. \tag{A.3}$$

The expected sojourn time in state 1, respectively, state 2, is a, respectively, b, time slots. The matrices B_n are found as

$$B_n = \begin{pmatrix} \dfrac{\lambda_1^n\,\mathrm{e}^{-\lambda_1}}{n!}\left(1 - \dfrac{1}{a}\right) & \dfrac{\lambda_1^n\,\mathrm{e}^{-\lambda_1}}{n!}\dfrac{1}{a} \\ \dfrac{\lambda_2^n\,\mathrm{e}^{-\lambda_2}}{n!}\dfrac{1}{b} & \dfrac{\lambda_2^n\,\mathrm{e}^{-\lambda_2}}{n!}\left(1 - \dfrac{1}{b}\right) \end{pmatrix}. \tag{A.4}$$

Notice, $\sum_n B_n = B = T$. The arrival rate λ is calculated as $(\lambda_1 a + \lambda_2 b)/(a + b)$. For later reference, we abbreviate the two-state Markov modulated Poisson process with parameters λ_1, λ_2, a and b as $M(\lambda_1, \lambda_2, a, b)$.

Appendix B. Proof of Theorem 1

Let Y_i and X_i denote the length and the number of participants of the ith collision resolution interval (CRI), where X_0 and Y_0 correspond to the CRI beginning at time $t = 0$. Let Z_i denote the state of the primitive D-BMAP at the start of the ith CRI, where Z_0 is the state at time $t = 0$. Let $T(n)$ be the expected time required by the CRA to resolve a set of n contenders, i.e., $T(n) = E[Y_i | X_i = n]$. Using the law of total probability, we have

$$E[Y_i] = \sum_{n=0}^{\infty} P[X_i = n]E[Y_i|X_i = n]. \tag{B.1}$$

Let $\tau = \limsup n/T(n)$, then for any $\epsilon_1 > 0$ there exists an N such that $n/T(n) \leq \tau + \epsilon_1$ for $n > N$. In other words, $T(n) \geq n/(\tau + \epsilon_1)$ for $n > N$. Therefore, we can write Eq. (B.1) as

$$E[Y_i] \geq \frac{1}{\tau + \epsilon_1} \sum_{n>N} nP[X_i = n] + \sum_{n \leq N} E[Y_i|X_i = n]P[X_i = n]. \tag{B.2}$$

Let $T(n) = n/(\tau + \epsilon_1) + g(n)$, where $g(n)$ is a correction that can be either positive or negative. Therefore,

$$E[Y_i] \geq \frac{1}{\tau + \epsilon_1} E[X_i] + \sum_{n \leq N} g(n)P[X_i = n]. \tag{B.3}$$

Whenever $g(n) \geq 0$ we use 0 as a lower bound for $g(n)P[X_i = n]$; otherwise, we use $g(n)$ as an lower bound for $g(n)P[X_i = n]$. Hence,

$$E[Y_i] \geq \frac{1}{\tau + \epsilon_1} E[X_i] + e, \tag{B.4}$$

where $\epsilon_1 > 0$, $e \leq 0$ is a fixed number that depends upon the value of ϵ_1 and the CRA, but that does not depend upon i. We know from Eq. (3) that for any primitive D-BMAP the expected number of arrivals in an interval of length L approaches λL as L approaches infinity, where λ is the arrival rate of the D-BMAP (independent of the state at the start of the interval). Thus, because the number of states of a D-BMAP is finite, we have that for any $\epsilon_2 > 0$ there exists a K such that $E[X_{i+1}|Y_i = L] \geq (\lambda - \epsilon_2)L$ for $L > K$. Hence, by means of the law of total probability

$$E[X_{i+1}] \geq (\lambda - \epsilon_2) \sum_{L>K} LP[Y_i = L] + \sum_{L \leq K} P[Y_i = L]E[X_{i+1}|Y_i = L]. \tag{B.5}$$

Recall Z_i is the state of the D-BMAP at the start of the ith CRI. Obviously,

$$E[X_{i+1}|Y_i = L] \geq \min_j E[X_{i+1}|Y_i = L \cap Z_i = j]. \tag{B.6}$$

The expression $\min_j E[X_{i+1}|Y_i = L \cap Z_i = j]$ is nothing but the expected number of arrivals generated by the input D-BMAP during an interval of length L, provided that the state at the start of the interval is j. Hence, we can write it as $(\lambda - \epsilon_2)L + h(L)$, where $h(L)$ is a correction that is either positive or negative, to obtain

$$E[X_{i+1}] \geq (\lambda - \epsilon_2)E[Y_i] + \sum_{L \leq K} h(L)P[Y_i = L]. \tag{B.7}$$

For $h(L)$ negative, respectively, positive, we replace $h(L)P[Y_i = L]$ by $h(L)$, respectively, 0, to find that

$$E[X_{i+1}] \geq (\lambda - \epsilon_2)E[Y_i] + f, \tag{B.8}$$

where $f \le 0$ is a fixed number that depends upon ϵ_2, the input D-BMAP and CRA, but that does not depend upon i. Combining Eqs. (B.4) and (B.8) provides us with the following equation:

$$E[X_{i+1}] \ge \frac{\lambda - \epsilon_2}{\tau + \epsilon_1} E[X_i] + (\lambda - \epsilon_2)e + f \tag{B.9}$$

for $i \ge 0$. When the equality is taken in Eq. (B.9), we have a first-order linear recursion whose solution for the initial condition $X_0 = N$ and $Z_0 = j$ is a lower bound on $E[X_i]$. This lower bound can be rearranged to the following form:

$$E[X_i] \ge \left(N - \frac{[(\lambda - \epsilon_2)e + f]}{1 - (\lambda - \epsilon_2)/(\tau + \epsilon_1)}\right)\left(\frac{\lambda - \epsilon_2}{\tau + \epsilon_1}\right)^i + \frac{[(\lambda - \epsilon_2)e + f]}{1 - (\lambda - \epsilon_2)/(\tau + \epsilon_1)} \tag{B.10}$$

with $e \le 0$ and $f \le 0$. Define $[(\lambda - \epsilon_2)e + f]/(1 - (\lambda - \epsilon_2)/(\tau + \epsilon_1))$ as q. For $(\lambda - \epsilon_2) > (\tau + \epsilon_1)$ we find $q \ge 0$. Thus, for $(\lambda - \epsilon_2) > (\tau + \epsilon_1)$ the lower bound for $E[X_i]$ presented in Eq. (B.10) grows without a bound as i goes to infinity if N is large enough—that is, larger than q. For N smaller than q the lower bound for $E[X_i]$ decreases to minus infinity and we know nothing from Eq. (B.10).

Notice, Eq. (B.10) actually states that if a CRI with more than q participants occurs, $E[X_i]$ grows without bound—that is, the algorithm is unstable under D-BMAP traffic—for $\lambda > \tau$. It is not difficult to prove, by means of a finite Markov chain with an absorbing state, that a CRI with more than q contenders occurs, when the CRAs of Section 2 are used, with probability one if the input D-BMAP is not a D-MAP.

References

[1] A. Greenberg, P. Flajolet, R. Ladner, Estimating the multiplicity of conflicts to speed their resolution in multiple access channels, J. Assoc. Comput. Mach. 34 (2) (1987) 289–325.

[2] J. Massey, Collision resolution algorithms and random-access communication, in: G. Longo (Ed.), Multi-Users Communication Networks, CISM Courses and Lectures No. 256, Springer, Wien, 1981, pp. 73–137.

[3] J. Massey, Some new approaches to random-access communications, in: P. Courtois, G. Latouche (Eds.), Proceedings of the Performance'87, North Holland, Amsterdam, 1987.

[4] I. Cidon, M. Sidi, Conflict multiplicity estimation and batch resolution algorithms, IEEE Trans. Inform. Theory IT-34 (1) (1988) 101–110.

[5] G. Fayolle, P. Flajolet, M. Hofri, P. Jacquet, Analysis of a stack algorithm for random multiple-access communication, IEEE Trans. Inform. Theory IT-31 (2) (1985) 244–254.

[6] P. Mathys, P. Flajolet, Q-ary collision resolution algorithms in random-access systems with free or blocked channel access, IEEE Trans. Inform. Theory IT-31 (2) (1985) 217–243.

[7] P. Humblet, On the throughput of channel access algorithms with limited sensing, IEEE Trans. Comm. COM-34 (1986) 345–347.

[8] A. Tanenbaum, Computer Networks, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[9] M.L. Molle, G. Polyzos, Conflict resolution algorithms and their performance analysis, Technical Report, CS93-300, University of Toronto, 1993.

[10] D. Bertsekas, R. Gallager, Data Networks, Prentice-Hall, Englewood Cliffs, NJ, 1992.

[11] G. Polyzos, Molle, Performance analysis of finite nonhomogeneous population tree conflict resolution algorithms using constant size window access, IEEE Trans. Commun. 35 (11) (1987) 1124–1138.

[12] M. Seri, M. Sidi, Splitting algorithms in channels with Markovian capture, Eur. Trans. Telecommun. Rel. Technol. 5 (1) (1994) 19–26.

[13] I. Kessler, M. Sidi, Splitting algorithms in noisy channels with memory, IEEE Trans. Inform. Theory IT-35 (5) (1989) 1034–1043.

[14] A. Ephremides, B. Hajek, Information theory and communication networks: an unconsummated union, IEEE Trans. Inform. Theory 44 (6), 2416–2634.
[15] C. Blondia, O. Casals, Statistical multiplexing of VBR sources: a matrix-analytical approach, Perform. Eval. 16 (1992) 5–20.
[16] M. Neuts, Structured stochastic matrices of M/G/1 type and their applications, Marcel Dekker, New York, 1989.
[17] M. Neuts, Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach, John Hopkins University Press, Baltimore, MD, 1981.
[18] I. Widjaja, A. Elwalid, Performance issues in VC-merge capable switches for multiprotocol label switching, IEEE J. Select. Areas Commun. 17 (6) (1999) 1178–1189.
[19] J. Silvester, N. Fonseca, S. Wang, D-BMAP models for the performance evaluation of multimedia traffic in ATM networks, in: D. Kouvatsos (Ed.), Performance Modelling and Evaluation of ATM Networks, Chapman & Hall, London, 1994, pp. 325–346.
[20] A. Kamal, Efficient solution of multiple server queues with applications to the modeling of ATM concentrators, in: Proceedings of the IEEE Infocom, San Francisco, CA, 1996, pp. 248–254.
[21] B. Van Houdt, C. Blondia, O. Casals, J. García, Performance evaluation of a MAC protocol for broadband wireless ATM networks with QoS provisioning, J. Interconnection Netw. 2 (1) (2001) 103–130.
[22] K. Spaey, C. Blondia, Circulant matching method for multiplexing ATM traffic applied to video sources, in: Proceedings of the Performance of Information and Communication Systems (PICS), Lund, Sweden, 1998, pp. 234–245.
[23] E. Aarstad, S. Blaabjerg, F. Cerdan, S. Peeters, K. Spaey, Cac investigation for video and data, in: Proceedings of the Broadband Communications, Stuttgart, Germany, 1998, pp. 356–368.
[24] B. Van Houdt, C. Blondia, Stability and performance of stack algorithms for random access communication modeled as a tree structured QBD Markov chain, Stochast. Models 17 (3) (2001) 247–270.
[25] B. Van Houdt, C. Blondia, Throughput of q-ary splitting algorithms for contention resolution in communication networks, Adv. Perform. Anal., in press.
[26] D. Lucantoni, New results on the single server queue with a batch Markovian arrival process, Stochast. Models 7 (1) (1991) 1–46.
[27] D. Lucantoni, K. Meier-Hellstern, M. Neuts, A single server queue with server vacations and a class of non-renewal arrival processes, Adv. Appl. Prob. 22 (1990) 676–705.
[28] C. Blondia, A discrete-time batch Markovian arrival process as B-ISDN traffic model, Belgian J. Oper. Res., Statist. Comput. Sci. 32 (3–4) 3–23.
[29] E. Seneta, Non-negative Matrices and Markov Chains, Springer, New York, 1981.
[30] C. Blondia, F. Geerts, The correlation structure of the output of an ATM multiplexer, in: Proceedings of the Fifth IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Kluwer Academic Publishers, Dordrecht, 1999, pp. 235–250.
[31] D. Gross, C. Harris, Fundamentals of Queueing Theory, Wiley, New York, 1974.
[32] Q. He, Classification of Markov processes of matrix M/G/1 type with a tree structure and its applications to the MMAP[K]/G[K]/1 queue, Stochast. Models 16 (5) (2000) 407–434.
[33] Q. He, M. Neuts, Markov chains with marked transitions, Stochast. Process. Appl. 74 (1998) 37–52.

B. Van Houdt received his M.Sc. degree in Mathematics and Computer Science, and a Ph.D. in Science from the University of Antwerp (Belgium) in July 1997, and May 2001, respectively. From August 1997 until September 2001 he held an Assistant position at the University of Antwerp. Since October 2001, he is a postdoctoral fellow of the FWO-Flanders.

His main research interest goes to the performance evaluation and stochastic modelling of wired and wireless telecommunication networks. He has published a several papers, containing both theoretical and practical contributions, in a variety of international journals (e.g., IEEE JSAC, Performance Evaluation, Journal of Applied Probability, Stochastic Models, etc.) and in conference proceedings.

C. Blondia obtained his Ph.D. in Mathematics from the University of Ghent (Belgium) in 1982. Between 1986 and 1991 he was a researcher at the Philips Research Laboratory and from 1991 to 1994 he was an Associate Professor at the Computer Science Department of the University of Nijmegen (The Netherlands). In 1995 he joined the Department of Mathematics and Computer Science of the University of Antwerp, where he is currently a professor, head of the research group "Performance Analysis of Telecommunication Systems" (PATS) and Head of the Department. His main research interests are related to both theoretical methods for stochastic modelling (in particular queueing systems) and to the design, the modelling and performance evaluation of telecommunication systems, in particular, related to traffic management in broadband networks, IP mobility management, medium access control for wireless and wired access networks, etc. He has published a substantial number of papers in international journals and conference proceedings. He is editor of the "Journal of Network and Computer Applications" and member of IFIP W.G. 6.3 on "Performance of Computer Networks". He is and has been involved in several European research programs (RACE, ACTS, IST and COST).