Towards Computable Stability Criteria for Some Multidimensional Stochastic Processes Arising in Queueing Models

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
89-879
TOWARDS COMPUTABLE STABILITY CRITERIA
FOR SOME MULTIDIMENSIONAL STOCHASTIC
PROCESSES ARISING IN QUEUEING MODELS

Wojciech Szpankowski

CSD-TR-879
April 1989
Primary motivation for this research is the need for firmly based but also easily computable methods applicable to the study of stability of stochastic models arising in the analysis of computer and communication systems. The stability definition adopted in this work is broad enough to cover such problems as existence of stationary distribution, ergodicity and nonergodicity, finiteness of some quantities of interest and so forth. In this article, though we mainly discuss multidimensional Markovian models, the stability of such models is ascertained by Markovian and non-Markovian methods. In the first category, we concentrate on the Lyapunov (test) function approach. The latter methodology is based on Loynes' result regarding stability of a general (non-Markovian) GIGI queue. A variety of approaches are used to obtain ultimate stability conditions for practical systems such as token passing rings, coupled-processor systems, buffered ALOHA systems and a decentralized dynamic control protocol for broadcast communications.
1. INTRODUCTION

*Stability* is a fundamental issue in the performance of real life systems, since only stable systems can operate in practice. There are many concepts of stability, but all of them fall into the following definition. A system is stable if it preserves *required properties* in the presence of some perturbations (disturbances). In a stochastic approach to the analysis of computer and communication systems, a source of disturbances is usually the arrival process. Then the concept of stability depends on the nature of the required properties. Existence of a steady-state distribution leads to stability in the sense of *ergodicity* and (for general, not necessary Markovian methods) *stationary distributions*. *Finiteness* of some moments of a quantity of interest is another concept of stability. Investigating small changes in the output distribution (e.g., queue length) subject to small changes in the arrival process, we must deal with stability in the sense of *robustness* and *continuity*. Studying bistability, we require that a desired property is a particular shape of a steady state distribution, and so forth. In this article we restrict our interest to stability in the sense of existence of stationary distribution (ergodicity) and finiteness of moments. But we focus on *multidimensional* stochastic processes, which makes the study interesting. While one-dimensional (also in some cases, two-dimensional) Markov chains have been studied extensively over the last twenty years and stability criteria are well known, the multidimensional case is an open area of continuing research. We shall present a state-of-the-art in establishing easily computable stability criteria for multidimensional Markov chains that arise in the analysis of computer and communication systems. We must warn the reader, however, that our exposition is biased by the author's taste, and his own involvement in the area.

Before we present a plan of the article, we briefly discuss a history of stability criteria for stochastic models that have been influenced by the rapid growth in the development of computer and communication systems. We can group relevant papers into three categories: ergodicity conditions for Markov processes, stability criteria for non-Markovian processes (or non-
Markov analysis of Markovian processes) and stability analyses for some specific stochastic systems, such as token passing rings, ALOHA systems, exponential back-off protocols, etc. In the first category, we restrict our attention to Markov chains and focus on the classification of states in such a process, i.e., ergodicity and nonergodicity problems. The first paper to present easily verifiable ergodicity conditions for Markov chains with a countable number of states, was due to Foster [FOS53]. Under his influence, in 1969, Pakes derived the so called Pakes’ Lemma, a result which is probably the most often used in establishing stability for a one-dimensional Markov chain. Later, Tweedie in [TWE76, TWE81, TWE82] (and in many other papers of his own or with his collaborators) extended Foster’s criteria to uncountable Markov chains. Another line of research is visible in the papers of Malyshev [MAL72], Mensikov [MEN74] and Malyshev and Mensikov [MaM81]. Although they have been able to present, for some particular cases, sufficient and necessary conditions for ergodicity of a multidimensional Markov chain, unfortunately their criteria are very difficult to verify in practice, except for two-dimensional Markov chains. In the latter case, however, we should mention a contribution recently reported by Rozenkrantz [ROS89], and Vaninskii and Lazareva in [VaL88]. These authors relaxed the assumption of bounded jumps required by Malyshev in [MAL72]. Hajek in [HAJ82] studied bounds of exponential type for the first-hitting time and occupation times of a real valued random sequence. These bounds present a flexible technique for providing stability of processes frequently encountered in the control of queues (e.g., geometric ergodicity for a certain two-dimensional Markov chain which arises in the decentralized control of a multiaccess system). In [SZP88] Szpankowski introduced some other criteria for multidimensional Markov chains. Finally in 1979 Kaplan [KAP79] initiated studies in (practical) criteria for the nonergodicity of Markov chains. This work was extended in the research of Sennott et al. [SMT83], Szpankowski [SZP85] and Szpankowski and Rego [SzR88].

Another approach was adopted by Loynes in [LOY62] who derived stability conditions for
a non-Markovian stochastic process, arising in the analysis of the GlGls queue. He proved that
the ergodicity condition of a Markovian queue (GlGls) is identical to the stability condition for
a non-Markovian queue (i.e., GlGls). His work was extended by Borovkov in [BOR76,
BOR78], Rolski [ROL81] and Baccelli and Bremaud [BaB87]. Recently, Szpankowski and
Rego [SzR87] applied Loynes’ result to obtain sufficient and necessary conditions for a multi­
queue system arising in the analysis of computer and communication systems.

The third category of research in stability problems is motivated by the proliferation of
computer and communication systems, and distributed computing environments. Authors of
papers in this category have studied stability conditions arising in the analysis of particular sys­
tems. For example, Kuehn [KUE79] presented stability criteria for a class of token passing sys­
tems, but however without (formal) proof (see also [WAT84]). Other stability criteria are met
in the analysis of couple-processor systems [FaL79, CoB83]. Unfortunately, the analyses of
[FaL79, CoB83] are restricted to the two users case, and based on rather sophisticated tools,
namely the Riemann-Hilbert problem approach. A large class of stability problems arises in the
evaluation of multiaccess protocols with buffered or unbuffered (unit-capacity) users. The ergo­
dicity condition for slotted buffered ALOHA systems was initiated by Tsybakov and Mikhailov
[TsM79]. This research was continued by Rao and Ephremides [RaE89], Sharma [SHA89],
Szpankowski [SZP88], Tsybakov [TSY85], Tsybakov and Bakirov [TsB84] (see also [KLA86,
RoT83 and SaE81]). Finally exponential back-off algorithms gave another "push" into
research on stability (see [ALD87, HaL82, HAJ82, KEL85, ROS84, SzR88]). The contribution
of computer scientists to that problem is well established in two excellent papers by Goodman,
Greenberg, Madras and Mardi [GGM88], and Halstads, Leighton and Rogoff [HLR87].

Our presentation of stability criteria follows the above sketched "historical paradigm". In
the next section, we discuss a variety of stability concepts. Section 3 is entirely devoted to
ergodicity and finiteness of moments for multidimensional Markov chains. In this section, we
study stability criteria via a Markovian approach (i.e., Lyapunov test function, drift, etc.). However, we do believe that to obtain ultimate stability conditions for Markov processes, we cannot stay within the frame of Markovian analysis. Therefore, Section 4 presents a non-Markovian analysis applied to Markovian systems, and it is based on Loynes' result. Finally, the last section applies to various criteria derived in the previous sections to obtain stability criteria for token passing rings, coupled-processor systems, buffered ALOHA systems and decentralized dynamic control protocols for broadcast communications.

2. THE MYRIAD VIRTUES OF STABILITIES

A non-trivial problem is to design stable systems and to recognize whether a system is stable or not. Moreover, a system may be stable in one sense and unstable in another sense. A sense of stability depends on what one understands by required properties and perturbations. Below we attempt to present various definitions of stability that can be applied to the analysis and design of some real-time communication and computing systems.

Let a system be described by an $M$-dimensional stochastic process $N^t = (N_1^t, N_2^t, \ldots, N_M^t)$ defined over a denumerable state space $\mathcal{E}$. For example, $N_i^t$ may represent queue length in the $i$-th buffer of a network of queues. Without loss of generality we further assume that the state space $\mathcal{E}$ comprises $M$-tuples of nonnegative integers, and time $t$ is discrete, that is, $t = 1, 2, \ldots$. Various stability concepts may be studied. From the practical viewpoint, the following concepts seem to be the most important.

Stationary distribution. By stability in this case we mean that the distribution of $N^t$ as $t \to \infty$ exists and the distribution is honest. In other words, $N^t$ is stable if for any $x = (x_1, \ldots, x_M)$ the following holds

$$\lim_{t \to \infty} \Pr\{N^t < x\} = F(x) \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1$$

(2.1)

where $F(x)$ is the limiting distribution function, and by $x \to \infty$ we understand that $x_j \to \infty$ for
all \( j \in \mathcal{M} = \{1, 2, \ldots, M\} \). If a weaker condition holds, namely,

\[
\lim_{x \to \infty} \liminf_{t \to \infty} \Pr[N_t < x] = 1
\]

(2.2)

then the process is called substable [LOY62]. Otherwise, the system is unstable (for more details see [LOY62, BOR76, BOR78]). The relationship between stability and substability is of course that a stable sequence is necessary substable, and a substable sequence is stable if the distribution function tends to a limit. For example, if \( N_t \) is an aperiodic and irreducible Markov chain, then substability is equivalent to stability (i.e., ergodicity) since a limiting distribution exists (it may be degenerate) for any such a Markov chain [CHU67].

Ergodicity. Now we assume that \( N_t \) is an irreducible aperiodic Markov chain [CHU67]. This postulate is assumed throughout this entire article. For Markov chains we define steady-state probabilities \( \pi_k \), as \( \pi_k = \lim_{t \to \infty} \Pr[N_t = k] \) where \( k \in \mathcal{J}^M \) and \( \mathcal{J} \) is the set of nonnegative integers.

A system is ergodic (stable) if and only if \( \pi_k > 0 \) and \( \sum_{k \in \mathcal{J}} \pi_k = 1 \). Sometimes a stronger condition is required, namely one needs that the rate of convergence to the steady-state distribution is fast enough. More precisely, let \( P_{i,k}^t = \Pr[N_{t+1} = k | N_t = i] \) be the transition probabilities of \( N_t \). Then, one requires that

\[
| P_{i,k}^t - \pi_k | = O(\eta^t)
\]

(2.3)

for some \( \eta < 1 \), that is, the convergence is geometrically fast (for more a precise definition see [TWE81, TWE82]). If (2.3) holds, then the Markov chain is called geometrically ergodic.

Finite moments. Let \( E N_i^t \) denote the \( l \)-th moment of the \( i \)-th component of \( N_t \) as \( t \to \infty \). A system is stable if for all \( i \in \mathcal{M} \) and given \( l \), the moments \( E N_1^l, E N_2^l, \ldots, E N_M^l \) exist and are finite [TWE83, SzR88].

Partial stability. In some systems, the steady-state distribution \( \pi_k, k \in \mathcal{J} \), may not exist for \( N_t \).
but marginal distributions of some components of $N'$ are still well-defined. Consider an example. Let $(N_1', N_2')$ be a two-dimensional Markov chain, and define $\pi_{k_1,k_2}$, $\pi_{k_1}$, $\pi_{k_2}$ as

$$\pi_{k_1,k_2} = \lim_{t \to \infty} Pr\{N_1' = k_1, N_2' = k_2\}, \quad \pi_{k_1} = \lim_{t \to \infty} Pr\{N_1' = k_1\}, \quad \pi_{k_2} = \lim_{t \to \infty} Pr\{N_2' = k_2\}.$$  

If $(N_1', N_2')$ is not ergodic then [CHU67] $\pi_{k_1,k_2} = 0$ for all $(k_1, k_2) \in \mathcal{E}^2$. But, by Fatou’s lemma

$$\pi_{k_1} = \lim_{t \to \infty} Pr\{N_1' = k_1\} = \lim_{t \to \infty} \sum_{k_2=0}^{\infty} Pr\{N_1' = k_1, N_2' = k_2\} \geq \sum_{k_2=0}^{\infty} \lim_{t \to \infty} Pr\{N_1' = k_1, N_2' = k_2\} = 0$$

hence $\pi_{k_1}$ or $\pi_{k_2}$ might be positive, and the marginal distribution may exist. To generalize it, let us consider a set of distinct indices $l_1, l_2, \ldots, l_n \in \mathcal{M}$, and for $k \in \mathcal{E}^n$. Then $\pi_{k} = \lim_{t \to \infty} Pr\{N_{l_1}' = k_{l_1}, N_{l_2}' = k_{l_2}, \ldots, N_{l_n}' = k_{l_n}\}$. Then, a system is partially ergodic if there exists an $n$-tuple $l=(l_1, \ldots, l_n)$ such that $\pi_{l} > 0$, $l \in \mathcal{E}^n$ and $\sum_{k \in \mathcal{E}^n} \pi_{k} = 1$.

Finally, define a function $f: \mathcal{E}^n \to \mathcal{R}$, where $\mathcal{R}$ is a set of real numbers. Consider $Ef(N')$. For example, if $f()$ is a projection on the $i$-th axis, then $f(N) = N_i'$, and $Ef(N) = EN_i$ is the average of $N_i$; if $f(N) = N_{i_1} + N_{i_2}$, $l_1, l_2 \in \mathcal{M}$, then $Ef(N) = EN_{i_1} + EN_{i_2}$ is the sum of average values of $N_{i_1}$ and $N_{i_2}$.

Practical Stability. Let $D$ be an average delay for a packet in a computer network with total input rate $\lambda$ packets per unit of time. In many applications one declares a system to be stable [KLE76, SZP83] if for a given $D_{\text{max}}$ the following holds $D \leq D_{\text{max}}$ for a set of input rates. Generalizing it, let $\lambda$ represent an input parameter and $\mathcal{A}$ be a set of admissible values of $\lambda$. Let also $c(\lambda)$ be a criterion function for system, e.g. delay, average queue length or a probability of loss. Define a set of required properties as $\mathcal{B} = \{c : c(\lambda) \leq c_{\text{max}}\}$. Then, we say a system is stable with respect to a function $f()$ if there exists a function $f()$ such that $Ef(N') < \infty$. This is known in control theory as practical stability [SZP83].
Shape of Steady-State Distribution. Assume for simplicity that $N^t$ is a one-dimensional Markov chain with finite state space $\mathcal{C} = \{k:0 \leq k \leq M\}$. Then, the steady-state probability vector $\pi = [\pi_0, \pi_1, \ldots, \pi_M]$ is a solution of a system of linear equation $\pi P = \pi$, where $P = \{p_{ij}\}_{i,j=0}^{M}$ is a transition matrix. Consider the probabilities $\pi_k$, $k \in \mathcal{C}$ as a function of $k$. We denote it as $\pi(k)$. Some properties of a system (e.g., bistability [KLE76, SZP83, SZP89a]) depend on the type (shape) of the function $\pi(k)$. It is important to know whether $\pi(k)$, $k \in \mathcal{C}$ is a unimodal function (only one maximum), bimodal (two maxima) or $n$-modal ($n$ maxima of $\pi(k)$) function. Bimodal distributions of $\pi(k)$ may produce a bistable behavior, which is obviously an undesirable phenomena. We say that a system is stable in the sense of shape of steady-state distribution if $\pi(k)$ is unimodal function for $k \in \mathcal{C}$. The problem is that we want to identify this stability without solving the system of linear equations, i.e., knowing only the transition matrix $P$ we investigate $\pi(k)$ as a function of $k \in \mathcal{C}$.

In this article we restrict our interest to the stability in the sense of the existence of stationary distribution (ergodicity) and finite moments. The interested reader may find some information on the other stabilities in the references. In particular, the practical stability adopted to the ALOHA system is discussed by the author in [SZP83], and the shape of steady-state distribution for some one-dimensional Markov chains is studied in [SZP89a].

3. STABILITY CRITERIA FROM A MARKOVIAN PERSPECTIVE

In this section we restrict our interest to multidimensional Markov chain $N^t$ with denumerable state space $\mathcal{C}$ and discrete time $t = 0, 1, \ldots$. In addition, it is assumed that $N^t$ is irreducible and aperiodic. Then, as discussed in Section 2, stability of such a process is the best represented by the concept of ergodicity and the existence of moments for $N^t$. In Section 3.1 we present criteria for ergodicity, geometric ergodicity and finiteness of moments. Section 3.2 is devoted to similar criteria for nonergodicity and nonexistence of moments. Finally, Section 3.3 deals with one-and-two dimensional Markov chains, because in these cases we can present
This section studies stability of Markov chains in a framework of a typical Markovian analysis. The primary tool in such an approach is the Lyapunov (test) function method [FOS53, TWE81, SZP88], which is the main topic of this section. A test or Lyapunov function $V(k)$, $k \in \mathcal{C}$ is any nonnegative real-valued function, that is, $V: \mathcal{C} \to \mathbb{R}_+$, where $\mathbb{R}_+$ represents the set of nonnegative real numbers. With every Markov chain $N'$ and Lyapunov function $V(k)$ we associate an operator $AV(k)$, which is also called the generalized drift, as follows

$$AV(k) = E\{V(N'+1) - V(N') | N' = k\} \quad k \in \mathcal{C}. \quad (3.1)$$

An interpretation of the operator $AV(k)$ is simple, namely it represents the average one-step change of the Markov chain $N'$ over the function $V(\cdot)$ assuming the process in initially at state $k$. It turns out that the sign of the operator $AV(k)$ is crucial to determine whether a Markov chain $N'$ is stable or not. We prove these and some other assertions using a unified approach which is based on the following lemma.

**Lemma 1.** Let $\tau$ be a Markov moment such that $Pr\{\tau < \infty\} = 1$. Then, for a Markov chain $N'$ and a Lyapunov function $V(k)$ such that $|AV(k)| < \infty$ for $k \in \mathcal{C}$ the following holds

$$EV(N') = EV(N^0) + E \sum_{j=0}^{\tau-1} AV(N^j). \quad (3.2)$$

**Proof.** Let $Z' = V(N') - \sum_{j=0}^{\tau-1} AV(N^j)$. In [SzR88], it is shown that $Z'$ is a martingal. Then by the optional sampling theorem [KaT81] $EZ' = EZ' = EV(N^0)$, and hence (3.2). $\blacksquare$

For future references we note that Lemma 1 implies the following useful formula

$$E_k V(N') = V(k) + E_k \sum_{j=0}^{\tau-1} AV(N^j), \quad (3.3)$$

where by definition $E_k V(N') = E\{V(N') | N^0 = k\}$.

### 3.1 Ergodicity and Existence of Finite Moments

Traditionally, since the seminal works of Foster [FOS53] and Pakes [PAK69], ergodicity
criteria are associated with the average (conditional) drift function. It is well known that a one-dimensional Markov chain is ergodic if the drift function $d(k) = E\{N^{t+1} - N^t | N^t = k\}$ is negative for sufficiently large $k$ [PAK69]. A generalization of this criterion to a multidimensional case is not easy, since for an $M$-dimensional Markov chain $N^t$ the drift $d(k)$ is an $M$-dimensional vector, and the $i$-th component of $d(k)$ is defined as $d_i(k) = E\{N_i^{t+1} - N_i^t | N^t = k\}$. Note that the $i$-th component $N_i^t$ of a multidimensional Markov chain $N^t$ is not a Markov process. This causes formidable difficulties, however, some solutions exist. In particular, the Lyapunov (test) function method proposed in 1953 by Foster [FOS53] can be easily generalized to multidimensional cases, however, the usage of this method in this case is much more restricted [SZP88].

A generalization of Foster's result can be summarized in the following theorem.

**Theorem 1.** If there exists a Lyapunov function $V(k)$, and if for a constant $\varepsilon > 0$ and a finite set $\mathcal{H} \subset \mathcal{C}$ the following holds

\[
|AV(k)| < \infty \quad \text{for all} \quad k \in \mathcal{H}
\]
\[
AV(k) \leq -\varepsilon \quad \text{for all} \quad k \in \mathcal{C} - \mathcal{H}
\]

then the Markov chain $N^t$ is ergodic.

**Proof.** It suffices to prove that the first re-entry time $\tau_{\mathcal{H}}$ to a finite set $\mathcal{H} \subset \mathcal{C}$ has finite first moment [TWE81, TWE82]. To prove this we use Lemma 1, and in particular (3.3). Let $\tau = \min\{T, \tau_{\mathcal{H}}\}$ for any $T > 0$. Then by Lemma 1

\[
0 \leq E_k V(N^t) \leq V(k) - \varepsilon E_k \tau \quad \text{for} \quad k \in \mathcal{C} - \mathcal{H}
\]

and

\[
0 \leq E_k V(N^t) \leq V(k) + AV(k) - \varepsilon E_k \tau - \varepsilon \quad \text{for} \quad k \in \mathcal{H}.
\]

Since $T$ is arbitrary, and in the presence of (3.4a) the above easily implies that $E_k \tau_{\mathcal{H}} < \infty$, hence $N^t$ is ergodic. 

The usage of Theorem 1 depends on the successful choice of the Lyapunov function $V(\cdot)$.
and one can hardly find any rules in selecting \( V(\cdot) \). A useful approach was proposed in 1961 by Kingman \([KIN61]\), which is discussed below. Let \( V(x) \) belong to \( C^2 \) class of function, that is, a set of functions having continuous partial derivatives of the second order. Then, Taylor's expansion of \( V(x) \) leads to the following.

\[
V(y) = V(x) + \nabla V(x)(y - x) + R(x,y)
\]

where \( \nabla V(x) = \left[ \frac{\partial f(x)}{\partial x_i} \right]_{i=1}^M \) is the gradient of \( V(\cdot) \), and the reminder \( R(x,y) \) is defined as

\[
R(x,y) = \frac{1}{2} (y - x)^T \nabla^2 V(x')(y - x).
\]

In the above, \( x' \) is a point in the interval \([x,y]\), and \( \nabla^2 V(x) \) denotes the Hessian of \( V(\cdot) \) (i.e., a matrix of all second derivatives of \( V(\cdot) \)). This and Theorem 1 lead to the next corollary.

**Corollary 1.** Let \( V(\cdot) \) belongs to \( C^2 \), and the other hypotheses of Theorem 1 hold, except that (3.4b) is replaced by

\[
\nabla V(k) \cdot d(k) \leq -\varepsilon \quad \text{for} \quad k \in \mathcal{G} - \mathcal{H}.
\]

If, in addition, \( ER(x,N^t) = o(1) \), then the Markov chain \( N' \) is ergodic.

**Proof.** Taking the conditional average of both sides of (3.5a) and using (3.5b), we obtain

\[
E(V(N^{t+1}) - V(N^t) | N^t = k) = \nabla V(k) \cdot d(k) + ER(x,N^{t+1}) \leq -\varepsilon
\]

since \( ER(x,N^{t+1}) = o(1) \). Therefore (3.4b) in Theorem 1 holds for sufficiently large \( k \), and this proves our corollary.

These two criteria are illustrated in the following two examples.

**EXAMPLE 3.1. Linear Lyapunov functions**

The Lyapunov function method is particularly appealing in the case of the linear function, that is, for \( k = (k_1, k_2, \ldots, k_M) \in \mathcal{G} \) we define \( V(k) = c_1 k_1 + \cdots + c_M k_M \), where \( c_i \), \( i \in \mathcal{M} \) are constants. The choice of these constants may be very crucial for some applications (see Section 5). Now it is easy to see that the operator \( AV(k) \) is a linear combination of drift components \( d_i(k) \) \([SZP88]\), that is
\[ AV(k) = \sum_{i=1}^{M} c_i d_i(k) \]  

This and Theorem 1, suggest the following corollary.

**Corollary 2.** Let for some finite \( \mathcal{H} \subset \mathcal{G} \) and \( \epsilon > 0 \) the next two conditions hold

\[
\left| \sum_{i=1}^{M} c_i d_i(k) \right| < \infty \quad \text{for} \quad k \in \mathcal{H}
\]
\[
\sum_{i=1}^{M} c_i d_i(k) \leq -\epsilon \quad \text{for} \quad k \in \mathcal{G} - \mathcal{H}
\]

then the Markov chain is ergodic. \( \square \)

In the next example, we show that the average drift is easy to compute for queueing models.

**EXAMPLE 3.2. A Multidimensional Queueing Model**

Let \( N' = (N'_1, N'_2, \ldots, N'_M) \) represent queue lengths in \( M \) buffers of a queueing system. Then,

\[ N'^{t+1} = N'^{t} + X'^{t} - Y'^{t} \]

where \( X' = (X'_1, \ldots, X'_M) \) and \( Y' = (Y'_1, \ldots, Y'_M) \) are arrival and departure processes. If \( X' \) and \( Y' \) are i.i.d. processes, then \( N' \) is a Markov chain (we implicitly assume that the time \( t \) is discrete). The average drift vector is easy to compute from the above

\[ d(k) = E\{N'^{t+1} - N'^{t} | N'^{t} = k\} = E\{X'^{t} | N'^{t} = k\} - E\{Y'^{t} | N'^{t} = k\} \]

But \( E\{X'_n | N'^{t} = k\} \) and \( E\{Y'_n | N'^{t} = k\} \) are simply the \( n \)-th components of the conditional input rate \( S^{in}(k) \) and the conditional throughput \( S^{th}(k) \) respectively. These quantities are easy to estimate, since they are one-step conditional changes in the input process and the departure process. For example, if the input process to the \( n \)-th buffer does not depend upon the queue lengths \( N' \) in all buffers, then \( S^{in}_n(k) = \lambda_n \), where \( \lambda_n \) is the average input rate to the \( n \)-th buffer. The conditional throughput \( S^{th}(k) \) usually depends on \( k \), but the dependency is straightforward and the evaluation of \( S^{th}(k) \) is rather not troublesome at all. For instance, in a standard MiCH1 queue, \( S^{th}(k) = \mu \) for all \( k > 0 \) and \( S^{th}(0) = 0 \); in the ALOHA system with single buffers [KLE76]
SO(k) = kr(1-r)^k-1 where r is the probability of a retransmission [SZP83], etc.

Before we study criteria for the existence of some moments of N', let us shortly elaborate on the rate of convergence of transition probabilities \( P'_{k,k} = Pr(N_{t+1} = k | N_t = i) \) to the steady-state probabilities \( \pi_k \). In particular, we discuss the geometric ergodicity which is defined in (2.3). The following theorem establishes easily verifiable criteria for this type of stability, and it extends Theorem 1.

**Theorem 2.** Let all hypotheses of Theorem 1 hold, except that (3.4b) is replaced by

\[
AV(k) \leq -c V(k) \quad \text{for} \quad k \in \mathcal{J} - \mathcal{K},
\]

and in addition, \( V(k) \geq 1 \) for \( k \in \mathcal{K} \) then \( N' \) is geometrically ergodic.

**Proof.** It suffices to prove that there exists a constant \( r > 1 \) such that \( Er_{\tau_{\mathcal{K}}} < \infty \) where \( \tau_{\mathcal{K}} \) is the first re-entry time to \( \mathcal{K} \). The proof uses Lemma 1 in the same manner as in the proof of Theorem 1 and it is left for the reader.

Finally, we address the issue of the existence of finite moments of \( N' \). In particular, a question arises whether ergodicity conditions expressed in Theorem 1 are sufficient for the existence of some moments of \( N' \). Obviously the answer is no, and the example below proves it.

**EXAMPLE 3.3. Ergodic Markov Chains Without Any Moment**

Let \( N' \) be a one-dimensional Markov chain with transition probabilities \( p_{n,m} = 0 \) for \( m \neq 0 \) and \( m \neq n+1 \), and \( p_{n,n+1} = (n-1)/(n+1) \) and \( p_{n,0} = 1 - p_{n,n+1} \) for \( n \geq 2 \), with \( p_{01} = p_{12} = \frac{1}{2} \).

The drift \( d(k) = E\{N'_{t+1} - N'_t | N'_t = k\} = -1 < 0 \) for \( k \geq 2 \), hence by Theorem 1 the Markov chain is ergodic, and there exists a stationary distribution \( \pi_k \). For this Markov chain, it is not difficult to prove that \( \pi_k = 0.5\pi_0/(k(k-1)) \) for \( k \neq 0 \). So the \( r \)-th stationary moment \( E\{\lim_{t \to \infty} N'_t\}^r \) of \( N' \) becomes

\[
E\{\lim_{t \to \infty} N'_t\}^r = \sum_{k=0}^{\infty} k^r \pi_k \geq \frac{\pi_0}{2} \sum_{k=2}^{\infty} k^{r-2} = \infty \text{ for all } r \geq 1.
\]
Therefore, for every \( r \geq 1 \) the Markov chain \( N^r \) is ergodic with the stationary distribution \( \pi_k \), but no moment exists. □

This example suggests that criteria of Theorem 1 are not sufficient for the existence of finite moments. The following strengthened conditions are, however, enough to assure finite moments.

**Theorem 3.** We assume that hypotheses of Theorem 1 holds, however, condition (3.4b) is strengthened to

\[
AV(k) \leq -\varepsilon f(k) \quad \text{for } k \in \mathcal{G} - \mathcal{H}
\]

where \( f(\cdot) \) is a nonnegative function. If, in addition, \( V(k) \geq f(k) \) for \( k \in \mathcal{G} - \mathcal{H} \) and

\[
\sum_{k \in \mathcal{H}} \pi_k f(k) < \infty
\]

where \( \pi_k \) is the stationary distribution of \( N^r \). Then,

\[
\sum_{k \in \mathcal{G}} \pi_k f(k) < \infty \tag{3.10}
\]

that is, \( E \left[ \lim_{t \to \infty} f(N^t) \right] < \infty \) and the \( f(\cdot) \) stationary moment of \( N^r \) exists.

**Proof.** Since \( f(k) \geq 0 \) for all \( k \in \mathcal{G} \), hence (3.8) implies (3.4b) and by Theorem 1 stationary distribution \( \pi_k \) exists, so (3.10) makes sense. The details of the proof can be found in [TWE83]. We only note here that (3.9) applied to (3.3) implies that for \( k \in \mathcal{G} - \mathcal{H} \)

\[
E_k \sum_{j=0}^{\tau_k - 1} f(N^j) \leq V(k)/\varepsilon
\]

This is sufficient for the so called ergodicity of order \( f(\cdot) \) (see Tweedie [TWE82] for more details).

An important application of Theorem 3 to random walk on \([0,\infty)\) is discussed in the example below.

**EXAMPLE 3.4. Random Walk on \([0,\infty)\)**

Let us consider a simple random walk \( W^r \) on \([0,\infty)\) defined as (see also (3.6))
\[ W^{t+1} = (W^t + Z')^+ \]  

(3.11)

where \( a^+ = \max \{0, a\} \), and \( Z' \) is a sequence of i.i.d. random variables with distribution function \( F(\cdot) \). From Example 3.2 and Theorem 1, we know that \( W^t \) is ergodic if \( EZ^t \leq -\varepsilon \) for \( k \) sufficiently large (in fact, it is enough to assume \( k > 0 \)). The next question is under what additional conditions \( W^t \) has finite moments. Let \( f(\cdot) \) be a nonnegative function, and we shall investigate the existence of \( E\left( \lim_{t \to \infty} f(W^t) \right) \). To avoid further unnecessary complications, we shall assume that \( W^t \) is continuous on \([0, \infty)\) (e.g., \( W^t \) represents waiting time in a queue instead of queue length \( N_t \) as in Example 3.2). We must verify (3.8) for suitable chosen Lyapunov function \( V(x), x \in \mathbb{R}_+ \). We select a Lyapunov function \( V(\cdot) \) such that 

\[ AV(x) = E\{ V(W^{t+1}) | W^t = x \} - V(x) = EV(Z' + x) - V(x) \]

But under appropriate conditions on \( f(\cdot) \), we obtain from Taylor’s expansion of \( V(Z' + x) \)

\[ AV(x) = f(x)EZ + \frac{1}{2} f'(x')EZ^2 \]

(3.12)

where \( x' \in (0, x) \), and \( Z \) is generic notation for \( Z' \). If \( EZ < -\varepsilon \) and \( EZ^2 < \infty \) then under suitable conditions on \( f(\cdot) \) we can easily assure that the LHS of (3.12) is smaller than \( -\varepsilon f(x) \), i.e., 

\[ AV(x) \leq -\varepsilon f(x) \]

as required in Theorem 3. The following is proved in [TWE83].

**Corollary 3.** Let \( f(\cdot) \) be defined as follows

\[ f(x) = x^{a}(\log x)^{\beta} \quad \alpha \geq 1, \quad \beta \geq 0 \]  

(3.13)

Then \( E\left( \lim_{t \to \infty} f(W^t) \right) < \infty \) provided that \( EZ < 0, EZ^2 < \infty \) and \( EZ^{a+1}(\log Z)^{\beta} < \infty \). ☐

### 3.2 Instability Criteria

Theorem 1 provides criteria for ergodicity of \( N^t \). These criteria are **not** necessary and this leads to major difficulties, in particular, for multidimensional processes. Since necessary conditions for ergodicity of \( N^t \) are equivalent to criteria for **nonergodicity** of \( N^t \), we shall study in this subsection, the latter problem.
A question arises whether a converse theorem to Theorem 1 leads to nonergodicity conditions. In other words, is it true that $AV(k) \geq \varepsilon$ for sufficiently large $k$ leads to nonergodicity? In general, the answer is in the negative, as it is illustrated below.

**EXAMPLE 3.5. Ergodic Markov Chain With Positive Drift**

Let $N^t$ be a one-dimensional Markov chain with transition probabilities as follows: $p_{0i} = 2^{-i}$ $(i \geq 1)$, and for each $i \geq 1$ we select $k(i) > 2i$ and let $p_{i0} = p_{i k(i)} = 1/2$. The average drift $d(k) = k(i)/2 - i > 0$ $(i \geq 1)$. But in [ShiT83] it proved that $N^t$ is ergodic since there exists a column (i.e., the $k(i)$-th column) whose entries are bounded away from zero for sufficiently large low index. \(\Box\)

The above example shows that some more restrictions are needed to assure nonergodicity. Kaplan in [KAP79] was the first who successfully attacked the problem and presented a solution. He has shown that the function $\psi(z)$, called Kaplan's function, and defined as follows $\psi(z) = -E(z^{N^{(t)}} - z^{N'}) | N' = k)/(1 - z)$ for $z \in [0,1)$, must be bounded from below, and this together with positivity of the drift $d(k)$ finally assure nonergodicity. This was subsequently generalized by Sennott et al [SHT83], Szpankowski [SZP85], and Szpankowski and Rego [SzR88]. Let a generalized Kaplan's function be defined as below.

$$\psi_k^V(z) = -E(z^{V^{(N^{(t)}}}) - z^{V^{(N')}} | N' = k)/(1 - z) \quad k \in \mathcal{C}$$

(3.14)

for $z \in [0,1)$. Note that by l'Hospital's rule $\lim_{z \to 1} \psi_k^V(z) = AV(k)$ provided $|AV(k)| < \infty$. The first main result of this section is presented next.

**Theorem 4.** Let $V(\cdot)$ be a Lyapunov function and $\mathcal{H}$ be a proper subset of $\mathcal{C}$ such that

$$\inf_{k \in \mathcal{C} - \mathcal{H}} V(k) > \sup_{k \in \mathcal{H}} V(k)$$

(3.15)

If $|AV(k)| < \infty$, $k \in \mathcal{C}$ and for finite $\mathcal{H}_1$ the following holds

$$AV(k) \geq 0 \quad \text{for } k \in \mathcal{C} - \mathcal{H}$$

(3.16)

$$\psi_k^V(z) \geq -B \quad \text{for } k \in \mathcal{C} - \mathcal{H}_1$$

(3.17)
for some constant $B \geq 0$, then the Markov chain is not ergodic.

**Proof.** Assume contrary that $N'$ is ergodic with $\pi_k$ being stationary distribution. Then (3.16), (3.17) and Fatou’s lemma imply

$$0 \geq \lim_{z \to \infty} \sum_{k \in \mathcal{K}} \pi_k \psi^y(z) \geq \sum_{k \in \mathcal{K}} AV(k) \pi_k = \sum_{k \in \mathcal{K}} AV(k) \pi_k + \sum_{k \in \mathcal{K}} AV(k) \pi_k > 0$$

where the last inequality follows from (3.17) and the fact that $\sum_{k \in \mathcal{K}} AV(k) \pi_k > 0$ proved in [SHT83, SzR88] under the condition (3.15). This is the desired contradiction. \[ \]

We note that in Theorem 4 we do not require the finiteness of the set $\mathcal{K}$ as it was postulated in Theorem 1.

A natural question arises, namely for which Markov chains the Kaplan’s condition (3.17) is automatically satisfied, and what kind of instability property one can expect if (3.17) is dropped in Theorem 4. As long as the first problem is concerned, the solution to it was already given by Kaplan. We follow his arguments and define a downward uniformly bounded Markov chain $N'$ as the one for which the downward transition probabilities are bounded from the below. More precisely, we require that $p_{ji} = 0$ if $j < i - m$ for some $m = (m_1, \ldots, m_M)$. Then, it is proved [KAP79, SHT83, SZP85].

**Corollary 4.** If $N'$ is a downward uniformly bounded Markov chain, then Kaplan’s condition (3.17) holds, and (3.16) alone implies nonergodicity of $N'$. \[ \]

**EXAMPLE 3.6. Linear Lyapunov Function Revisited**

As in Example 3.1, we consider linear Lyapunov function $V(k) = c_1 k_1 + c_2 k_2 + \cdots + c_M k_M$. Then (3.6) holds, and one immediately obtains from Theorem 4, the next corollary.

**Corollary 5.** Let $\mathcal{K}$ be a finite subset of $\mathcal{C}$, and $N'$ is a downward uniformly bounded Markov chain. Then, the following conditions

$$\left| \sum_{i=1}^{M} c_i d_i(k) \right| < \infty \quad k \in \mathcal{C}$$

(3.18a)
\[ \delta \sum_{l=1}^{M} c_l d_l(k) > 0 \quad k \in \mathcal{C} - \mathcal{H} \]  

imply nonergodicity of \( N^t \).

The second question above is a little more intricate, and was disused by Sennott et al. [SHT83]. A solution was proposed by Szpankowski and Rego [SzR88]. To recall the problem, we drop Kaplan’s condition from Theorem 4, and ask what kind of instability property condition (3.16) may imply. We shall show that (3.16) implies unboundness of \( V(\cdot) \) moments for the Markov chain, that is, \( \lim_{t \to \infty} EV(N(t)) = \infty \). This kind of instability is bad enough for systems encountered in practice, and we should avoid this kind of instability. The nice thing about it, is that the criteria for this instability do not include Kaplan’s condition (3.17), which is difficult to verify in practice, especially for multidimensional Markov chains.

More precisely, the main result regarding this kind of instability is formulated in the next theorem.

**Theorem 5.** Let hypotheses of Theorem 4 hold, except that (3.17) is dropped and (3.16) is replaced by a little stronger condition, namely

\[ AV(k) \geq \epsilon > 0 \quad k \in \mathcal{C} - \mathcal{H} \]  

If, in addition, \( \lim_{k \to \infty} V(k) = \infty \), then \( \lim_{t \to \infty} EV(N(t)) = \infty \).

**Proof.** We split the proof into two parts, and first we assume \( N^t \) is ergodic. Then stationary distribution exists and one finds from Fatou’s lemma that \( \lim_{t \to \infty} EV(N(t)) \geq \sum_{k \in \mathcal{C}} V(k) \pi_k = \overline{V} \). We prove that \( \overline{V} = \infty \). Assuming contrary that \( \overline{V} < \infty \), we find out that \( \overline{V} = \overline{V} + \sum_{k \in \mathcal{C}} \pi_k AV(k) \).

But, as in the proof of Theorem 4, \( \sum_{k \in \mathcal{C}} \pi_k AV(k) > 0 \), so this is the desired contradiction.

We now let \( N^t \) be nonergodic. Since \( \lim_{k \to \infty} V(k) = \infty \), hence there exists \( B > 0 \) such that \( V(k) > B \) for \( k > m \). From nonergodicity of \( N^t \) and the above, we conclude that

\[ \lim_{t \to \infty} E_k V(N(t)) \geq B \quad \lim_{t \to \infty} Pr\{N(t) > m|N^0 = k\} \geq B, \]
and this implies \( \lim_{t \to \infty} EV(N^t) = \infty \) as the consequence of an arbitrary \( B > 0 \). ■

We illustrate Theorems 4 and 5, as well as the others, in Section 5, where various applications arising in computer communication areas are discussed.

3.3 Criteria for One-and-Two Dimensional Markov Chains

We know quite a lot about stability conditions for one-dimensional Markov chains. In practice, Pakes lemma [PAK69] and Kaplan's theorem [KAP79] settle down sufficient and necessary conditions for stability (at least for applications arising in queueing theory). These two criteria are stated below in the form most often encountered in practical applications.

**Lemma 2.** Let \( d(k) \) be the average drift for a one-dimensional Markov chain \( N^t \) which is assumed to be aperiodic and irreducible. If \( |d(k)| < \infty \) for all \( k \) and

\[
\lim_{k \to \infty} \sup_{k \to \infty} d(k) < 0
\]

then \( N^t \) is ergodic. ■

**Lemma 3.** If \( N^t \) is downward uniformly bounded Markov chain and

\[
\lim_{k \to \infty} \inf_{k \to \infty} d(k) > 0
\]

then \( N^t \) is not ergodic. ■

The 'lim sup' and 'lim inf' in the above lemmas are important, as it is shown in the next example.

**EXAMPLE 3.7. Conflict Resolution Algorithm** [CAP79, FlM85, SZP87]

Let us consider Capetenakis-Tsybakov-Mikhailov blocked conflict resolution algorithm [CAP79, FlM85, SZP87]. The average drift \( d(n) = \lambda L_n - n \) where \( \lambda \) is the input rate and \( L_n \) denotes the conditional length of a conflict resolution session with the initial multiplicity equal to \( n \). Then Lemma 2 implies that for \( \lambda^{-1} < \lim_{n \to \infty} \sup_{n \to \infty} L_n/n \) the system is stable (ergodic), and from Lemma 3, we conclude that for \( \lambda^{-1} > \lim_{n \to \infty} \inf_{n \to \infty} L_n/n \) the system is unstable. But in
it is proved that

\[ L_n/n = 2/\ln 2 - F(\log n) + O(1) \]  \hspace{1cm} (3.22)

where \( F(\log n) \) is a fluctuating function with a small amplitude, hence, \( \lim \sup L_n/n > \lim \inf L_n/n \). More precisely, the system is stable for \( \lambda < \lambda_{\text{crit}} - 4 \times 10^{-7} \) and unstable for \( \lambda > \lambda_{\text{crit}} + 4 \times 10^{-7} \) where \( \lambda_{\text{crit}} = 0.346574 \). \( \Box \)

Two-dimensional Markov chains are by far much more difficult to analyze from the stability viewpoint. This is obvious in the presence of finiteness of the subset \( \mathcal{M} \) in Theorem 1 where condition \( A^V(k) \leq -\varepsilon \) can be violated. In most (queueing) applications, there is significantly different behavior of the average drift \( d(k) = (d_1(k), d_2(k)) \) on the \( N_1 \)-axis (i.e., on set \( x(0), x \geq 0 \) which is \textit{infinite}) and on the \( N_2 \)-axis (again infinite set). A sufficient and necessary stability condition are known only for the so called maximally homogeneous Markov chains [MAL72, MaM81, VaL88]. For these chains, the average drift is assumed to be constant on \( N_1 \)-axis, \( N_2 \)-axis and in the region of \( N_1 > 0, N_2 > 0 \). More precisely, the drift vector \( d(k) = (d_1(k), d_2(k)) \) takes only three distinct values, namely \( d(0,0) = (d_1(0,0), d_2(0,0)) \), \( d(0,1) = (d_1(0,1), d_2(0,1)) \) and \( d(1,1) = (d_1(1,1), d_2(1,1)) \) defined respectively on the following sets \( (k,0), (0,k), (k,t), \) and \( k > 0, t > 0 \). In 1972 Malyshev [MAL72], using Lyapunov function (Theorem 1), proved the following theorem. Below by \( x < y \) we mean component-wise inequality, that is, \( x_1 < y_1 \) and \( x_2 < y_2 \).

**Theorem 6a.** For a maximally homogeneous two-dimensional Markov chain \( N' \), we assume that \( N'^{k+1} - N'k \) is uniformly bounded in \( k \), that is, for some \( K \) and \( N' = k \)

\[ ||N'^{k+1} - N'k|| < K \quad \text{uniformly in} \quad k \in \mathcal{C} \]  \hspace{1cm} (3.23)

where \( ||x|| \) is Euclidean norm. We consider four cases.

A. If \( d(1,1) > 0 \), then \( N' \) is transient.

B. If \( d(1,1) < 0 \), then \( N' \) is positive recurrent if and only if
\[ d_{1}(1,1) \cdot d_{2}(1,0) - d_{2}(1,1) \cdot d_{1}(1,0) < 0, \quad d_{2}(1,1) \cdot d_{1}(0,1) - d_{1}(1,1) \cdot d_{2}(0,1) < 0 \] (3.24)

and recurrent if the strict inequality in (3.24) is replaced by the weak inequality \( \leq 0 \).

C. If \( d_{1}(1,1) \geq 0 \) and \( d_{2}(1,1) < 0 \), then \( N^{i} \) is positive recurrent if and only if

\[ d_{1}(1,1) \cdot d_{2}(1,0) - d_{2}(1,1) \cdot d_{1}(1,0) < 0, \quad (3.25) \]

recurrent if equality holds in (3.25) and transient in the remaining cases.

D. Symmetric to C. □

The boundness condition (3.23) is rather restrictive, and for example, it rules out the Poisson arrival process. But a generalization of Malyshev's theorem is not easy and twenty six years passed since Rosenkrantz [ROS89] successfully "bite" the problem. Rosenkrantz considered a Lyapunov function \( V(\cdot) \) that is twice differentiable so that Corollary 1 can be applied. To construct the function we follow Rosenkrantz, and first we define another function \( \psi(\cdot, \cdot) \) which in polar coordinates \((r, \theta)\) is equal to \( \psi(r, \theta) = r^{\alpha} \cos(\theta - \omega) \) where \( \alpha \) and \( \omega \) depend on the angels that the vector \( d(k) \) makes with vectors \( n_{1} \) and \( n_{2} \) respectively inward pointing normals to the coordinates \( N_{1} \) and \( N_{2} \). Then the Lyapunov function is defined as \( V(\cdot) = \psi(\cdot) \) for \( \alpha > 0 \) and \( V(\cdot) = \psi^{-1}(\cdot) \) for \( \alpha < 0 \). Using such a construction, Rosenkrantz [ROS89] proved the following theorem.

**Theorem 6b.** Let hypotheses of Theorem 6a hold with the boundness condition (3.23) replaced by the following two requirements. Let \( K_{1} \) and \( K_{2} \) be constants and for \( N^{i} = k \) we postulate

\[ E\{||N^{i+1} - N^{i}||^{2}\} \leq K_{1} \]

(3.26a)

\[ N^{i+1} - N^{i} \geq c \quad \text{for} \quad ||k|| \geq K_{2} \]

(3.26b)

where \( c = (c_{1}c_{2}) \) and \( c_{1} > -\infty, c_{2} > -\infty \). Then:

A. As in part A of Theorem 6a.

B. If \( d(1,1) < 0 \), then (3.25) implies that \( N^{i} \) is recurrent, but not necessarily positive recurrent. The latter holds, however, if \( 1 \leq \alpha \leq 2 \). In addition, both LHS of (3.24) are strictly positive, then the chain is transient.
C. If \( d_1(1,1) \geq 0 \) and \( d_2(1,1) < 0 \), then (3.25) implies only recurrence of \( N' \). If, in addition, \( 1 \leq \alpha \leq 2 \), then (3.25) is sufficient for positive recurrence. Finally, if the LHS of (3.25) is strictly positive, then the chain is transient.

D. Symmetric to C.

Some applications of these theorems to stability problems in computer communication systems are presented in Section 5.

As the last tool of this section, we discuss the so called comparison tests [SZP88] for Markov chains. The idea is the following. Since we know explicit stability conditions for one- and two-dimensional Markov chains, we may use them to assess stability of multidimensional Markov chains by upper bounding components of the multidimensional process by some one- or two-dimensional Markov chains. More precisely, let \( \mathcal{M} = \{1,2,\ldots,M\} \) be an index set and we define a cover set \( \mathcal{P}_n = \{\sigma_1,\sigma_2,\ldots,\sigma_n\} \) such that \( \sigma_i \subset \mathcal{M} \) for every \( i \leq 1 \leq n \) and \( \bigcup_{i=1}^n \sigma_i = \mathcal{M} \). For example, if \( \mathcal{M} = \{1,2,3,4\} \), then a possible cover \( \mathcal{P}_3 \) of \( \mathcal{M} \) can be \( \mathcal{P}_3 = \{\sigma_1 = (1,2,3), \sigma_2 = (2,4), \sigma_3 = (3)\} \). In addition, from an \( M \)-dimensional Markov chain \( N' \) we extract a process \( N'_\sigma \) that consists of \( \sigma \) coordinates of \( N' \). Note that \( N'_\sigma \) is not a Markov chain. So let us define on the same set of indices \( \sigma \), and the same state space two \( 1 \)-dimensional Markov chains \( N'_\sigma \) and \( \overline{N}'_{\sigma} \) such that

\[
N'_\sigma \preceq N'_\sigma \preceq \overline{N}'_{\sigma}
\]

where \( N'_\sigma \preceq \overline{N}'_{\sigma} \) means stochastically smaller [STO83]. We note here that by the sample path comparison theorem [STO83] \( N'_\sigma \preceq \overline{N}'_{\sigma} \) implies that one can construct two other processes with the same distribution as \( N' \) and \( \overline{N}' \), but for these two processes the inequality \( \preceq \) holds for every path. This simplifies radically many proofs [SZP88, RaE89]. Having this in mind, we can formulate the comparison tests theorem.

Theorem 7. Let \( N' \) be a Markov chain and \( \mathcal{P}_n = \{\sigma_i\}_{i=1}^n \) be a cover of the index set \( \mathcal{M} \) of the
chain $N'$. 

(i) If there exists Markov chains $\overline{N}'_\sigma$ such that $N'_\sigma \leq \overline{N}'_\sigma$ for every $\sigma \in \mathcal{P}_n$, and $\overline{N}'_\sigma$ is ergodic for every $\sigma \in \mathcal{P}_n$, then the $M$-dimensional Markov chain $N'$ is ergodic too.

(ii) If there exist a nonergodic Markov chain $N'_\sigma$ for some $\sigma^* \in \mathcal{P}_n$ such that $N'_\sigma \leq N'_\sigma^*$, then the Markov chain $N'$ is not ergodic too.

**Proof.** The proof is rather easy and it imitates our other proofs of Lemmas 4 and 5 of the next section. Therefore, it is left for the interested reader. The detailed proof is presented in [SZP88].

**NOTES**

(i) The concept of (test) Lyapunov function was introduced by Foster [FOS53] and Theorem 1, practically speaking, was due to him. It was generalized to a continuous space by Tweedie [TWE76]; see also [MAE81, PAK69, SZP88]. Corollary 1 is due to Kingman [KING1]; see also [CLA86, ROS89]. Corollary 2 was presented by Szpankowski in [SZP88]. Geometric ergodicity and the existence of finite moments (Theorems 2 and 3) have been studied extensively by Tweedie [TWE81, TWE82, TWE83]. Example 3.3 is from Szpankowski and Rego [SzR88]. Corollary 3 is proved in [TWE83].

(ii) Nonergodicity criteria were initialized by a seminal paper of Kaplan [KAP79]. His result was generalized by Sennott et al. [SHT83], Szpankowski [SZP85] and Szpankowski and Rego [SzR88]. Example 3.5 is from [SHT83]. Theorem 4 is proved in [SzR88] and Corollary 5 from [SZP88]. The study of conditions for infinite moments is initialized in [SHT83] and developed in [SzR88] from which Theorem 5 comes.

(iii) Pakes lemma (Lemma 2) is from Pakes [PAK69]. Example 3.7 comes from Szpankowski [SZP87]. Sufficient and necessary conditions for ergodicity of a two-dimensional Markov chain (Theorem 6a) was presented by Malyshev in [MAL72]. The generalization given in Theorem 6b has been recently proved by Rosenkrantz [ROS89]. Another approach to this problem with the weakest assumption regarding the two-dimensional process $N'$ is suggested by Vaninskii and Lazareva [VaL88]. Finally the comparison tests (Theorem 7) were introduced by Szpankowski in [SZP88].

4. **STABILITY CRITERIA FROM A NON-MARKOVIAN PERSPECTIVE**

This section is devoted to a special multidimensional Markov chains $N'$ arising often in queueing models, e.g., $N'_i$ may represent the queue length in the $i$-th buffer in a multiqueue system. This restriction enables us to derive sufficient and necessary conditions for the discussed case. At first, we prove that for stability of $N'$, understood in the sense of (2.1) and (2.2), it is
required that every component $N^i_t, 1 \leq i \leq M$ of $N'$ is stable. This isolation lemma allows us to consider every non-Markovian queue $N^i_t$ in an isolation. But for a single general $GI_GI_H$ queue, Loynes [LOY62] and Borovkov [BOR78] presented sufficient and necessary conditions for stability, and this together with the isolation lemma is used to derive sufficient and necessary stability conditions for the multidimensional Markov chains $N'$.

We start with "isolation" lemmas that allow us to study every (non-Markovian) component $N^i_t$ of $N'$ separately. To recall, we investigate the concept of stability defined in (2.1) and (2.2), that is, we study the existence of a stationary distribution. Then, one proves the following two isolation lemmas.

**Lemma 4.** If for all $j = 1, 2, \ldots, M$, the one dimensional process $N^j_t$ is stable, then the $M$-dimensional process $N' = (N'_1, N'_2, \ldots, N'_M)$ is substable (see definition (2.2)).

**Proof.** Since each component of the process $N'$ is stable, then by definition (2.1) for all $j \in \mathcal{M} = \{1, 2, \ldots, M\}$

$$\lim_{x_j \rightarrow \infty} \lim_{t \rightarrow \infty} Pr\{N^j_t > x_j\} = 0$$

But

$$1 \geq \lim_{x \rightarrow \infty} \lim_{t \rightarrow \infty} Pr\{N^j_t \leq x, \text{for } j = 1, 2, \ldots, M\} \geq 1 - \sum_{j=1}^{M} \lim_{x_j \rightarrow \infty} \lim_{t \rightarrow \infty} Pr\{N^j_t > x_j\} = 1$$

Thus

$$\lim_{x \rightarrow \infty} \lim_{t \rightarrow \infty} Pr\{N' < x\} = 1$$

and $N'$ is substable by (2.2). If $N'$ is a Markov chain, then substability implies stability.

**Lemma 5.** If for some $j$, say $j^*$, $N^j_{t^*}$ is unstable, then $N'$ is also unstable.

**Proof.** Since $N^j_{t^*}$ is unstable, hence by (2.1) and (2.2)

$$\lim_{x_{j^*} \rightarrow \infty} \lim_{t \rightarrow \infty} \inf Pr\{N^j_{t^*} < x_{j^*}\} < 1$$

Then
\[
\lim_{x_p \to \infty} \lim_{t \to \infty} \inf \Pr \{N^t < x_p\} \leq \lim_{x_p \to \infty} \lim_{t \to \infty} \inf \Pr \{N_{j,p}^t < x_p\} < 1
\]
which proves Lemma 5. ■

Let us now assume that \(N^t = (N^t_1, N^t_2, \ldots, N^t_M)\) represents queue lengths in \(M\) buffers in a queueing model. By the isolation lemmas, we know that \(N^t\) is stable if and only if every queue \(N^t_j\) is stable. We note that the process \(N^t_j\) describing the queue length in the \(j\)-th buffer is not Markovian. In particular, the interarrival times \(\{A^n\}_{n=0}^\infty\) and service times \(\{S^n\}_{n=0}^\infty\) might not be i.i.d., and in addition \(\{S^n\}\) may depend on \(\{A^n\}\). What can be said about stability of such a general GlGII queue? In 1962 Loynes proved the following result.

**Theorem 8.** Let the pair \((A^n, S^n)_{n=0}^\infty\) be strictly stationary and ergodic (metrically transitive) process. We denote by \(E A\) and \(E S\) the average interarrival time and service time. Then the following holds

(i) if \(E A < E S\), then the GlGII queue is stable in the sense of definition (2.1),

(ii) if \(E A > E S\), then the GlGII queue is unstable,

(iii) if \(E A = E S\) the queue may be stable, substable or unstable. If \(\{S^n\}\) and \(\{A^n\}\) are independent of each other, and one of them is formed of non-constant mutually independent random variables, then the queue is unstable. ■

This important result of Loynes was not applied for decades to stability analysis of queueing models. In fact, the first important generalization is due to Borovkov [BOR76, BOR78] who has weakened the strict stationarity of Theorem 8 to asymptotic stationarity (see below). Nevertheless, even with this kind of stationarity, the application of the result might be troublesome. This suggests to look at the stability of GlGII from a new perspective. One possible approach is to consider another stability concept which is equally important in practice. For example, one may consider the first moment of the queue length \(E N^t\) as a criterion for stability. Then, it is natural to say that a GlGII queue is stable, if \(\lim_{t \to \infty} \sup_{x_p} E N^t < \infty\), and unstable other-
wise. If so, some new approaches are possible. Recently, Szpankowski proved the following result, which hopefully is "a piece of iceberg" for many new developments.

Theorem 9. A GIGI queue is analyzed without any assumptions regarding interarrival and service processes. Let $A^t$ represent the number of arrivals during the $t$-th epoch defined arbitrary, e.g., $t$ may represent the arrival instance of a customer. Then

(i) if $\lim_{t \to \infty} \inf EA^t > \lim_{t \to \infty} \sup Pr\{N^t > 0\}$, then $\lim_{t \to \infty} \sup EN^t = \infty$,

(ii) if $\lim_{t \to \infty} \sup EA^t < \lim_{t \to \infty} \inf Pr\{N^t > 0\}$, then $\lim_{t \to \infty} \sup EN^t < \infty$.

Proof. A GIGI queue can describe by $N^{t+1} = N^t + A^t - D^t$ where $A^t$ and $D^t$ represent the number of arrivals and departures during the $t$-th epoch. Then

$$EN^{t+1} - EN^t = EA^t - Pr\{N^t > 0\},$$

and the theorem follows from the following easy to prove results on any sequence $a_n \geq 0$:

(i) if $\lim_{n \to \infty} \sup a_n = \infty$, then $\lim_{n \to \infty} \sup (a_{n+1} - a_n) \geq 0$, 

(ii) if $\lim_{n \to \infty} \sup a_n < \infty$, then $\lim_{n \to \infty} \inf (a_{n+1} - a_n) \leq 0$.

This completes the proof. 

Before we leave a single GIGI queue, we notice that 9(i) can be equivalently expressed that $\lim_{t \to \infty} \inf EA^t > 1$ implies $\lim_{t \to \infty} \sup EN^t = \infty$, that is, instability. Theorem 9(ii) implies stability in the sense of the existence of a stationary distribution. However, in this case it might be difficult to assess the probability $\lim_{t \to \infty} \inf Pr\{N^t > 0\}$.

Now we can present our main contribution to the stability of multidimensional processes, and for simplicity of further considerations we shall concentrate on a generic queueing model which is used throughout this section to describe a large class of computer communication systems. Let us consider a distributed system with $M$ users that require the use of a single scarce resource. In queueing terminology, we say that customers (messages) from $M$ queues compete
for access to a single server. Each queue has an infinite capacity. The arrival process to the $j$-th queue is Poisson with parameter $\lambda_j$, $j \in M$. Messages arriving to the $j$-th queue, possess independent lengths that form an i.i.d random sequence with distribution function $H_j(\cdot)$. The average message length is denoted by $h_j$, and the first two moments of the service times are assumed to be finite. The server works in a distributed fashion. While visiting the $j$-th nonempty queue, the server removes at most one message at time. A server may visit a queue and not remove a message from the queue (e.g., the user is "down" or the algorithm does not allow the user to transmit, as is done in the ALOHA system [SaE81, SZP86, RaE89, SHA89].

To avoid confusion, we coin the term successful visit if the server visits a queue and either removes a message or the queue is empty.

The model can be described by an $M$-dimensional process $N_t = (N_j^t, N_j^t, \ldots, N_j^t)$ where $N_j^t$ represents the queue length in the $j$-th queue at time $t$. For a queue, say $j$, let $\tau_{j,n}$, $n = 0, 1, \ldots$, denote the end of the $n$-th successful visit of the server. We define the $n$-th cycle time $C_{j,n}$, as $C_{j,n} = \tau_{j,n+1} - \tau_{j,n}$. In addition, we define a so called modified service time. For that purpose, we choose from the sequence $\tau_{j,n}$ of successful visits a subsequence $\tau_{j,n_k}$, $k = 0, 1, \ldots$, such that at a time $\tau_{j,n_k}$ the $j$-th queue is nonempty, i.e., $N_j^t > 0$ for $t = \tau_{j,n_k}$ (the queue is nonempty after the service). We further denote this sequence of successful visits to the $j$-th nonempty queue as $\tau_{j,k}$. Then, the modified service time is defined as

$$C_{j,k}^* = \tau_{j,n_k+1} - \tau_{j,n_k} \quad k = 0, 1, \ldots$$

(4.1)

that is, during the time $C_{j,k}^*$ exactly one message is removed from the $j$-th queue, and hence $C_{j,k}^*$ may be interpreted as a new modified service time. Note that at time $\tau_{j,n_k+1}$ the queue may or may not be empty. On the other hand, if the queue is empty at a successful visit time $\tau_{j,n}$ (i.e., $n \neq n_k$), then the time elapsed until the next successful visit of the server is called a vacation time. More specifically, we define the $l$-th vacation time as $V_{j,l}$, where $l = 0, 1, \ldots$, and

$$V_{j,l} = \tau_{j,n_l+1} - \tau_{j,n_l} \quad \text{for} \quad n_l \neq n_k$$

(4.2)
where at time $\tau_{j,n}$, the queue is empty, i.e., $N_j^r=0$ for $r=\tau_{j,n}$. Naturally, if a customer arrives during a vacation it cannot be served until the end of this vacation.

In order to illustrate the above definitions, we show in Figure 1 a time diagram for one isolated queue of the distributed system. It is not difficult to conclude that this queue behaves as an M/G/1 queue with vacation [FuC85, DOS85, SHA88]. Indeed, in a queuing system with vacation it is assumed that a single server of walking type serves, each time it visits a nonempty queue, one customer for a service time $H_n$, and then takes a rest period $T_n$. If the queue is empty when the server returns, then the server takes off for a vacation period $V_n$. Any isolated queue in our generic distributed model works exactly in this manner. For example, the modified service time $C^*_n$ is equal to $H_n+T_n$. Note that in our distributed system the server visits other queues during the rest time $T_n$ or the vacation time $V_n$. Finally, we point out that in the case where the vacation distribution is exactly the same as the idle time distribution, then the queue reduces to the simple M/G/1 queue without vacation.

In summary, the evolution of the $j$-th queue in our generic model can be described by a stochastic equation

$$N_j^r = [N_j^{r-1} - 1]^+ + X_j(\tau_{j,n+1}, \tau_{j,n}) \quad n = 0, 1, \ldots,$$

where $X_j(\tau_{j,n+1}, \tau_{j,n})$ stands for the number of new arrivals to the $j$-th queue during the cycle time $C_{j,n}=(\tau_{j,n+1}, \tau_{j,n})$. Note that the length of the cycle $C_{j,n}=(\tau_{j,n+1}, \tau_{j,n})$ is equal either to the modified service time $C^*_{j,k}$ (if the queue is nonempty at time $\tau_{j,n}$) or to the vacation time $V_{j,k}$ (if the queue is empty at time $\tau_{j,n}$). In general the distribution of $X_j(\tau_{j,n+1}, \tau_{j,n})$ depends whether the interval $(\tau_{j,n+1}, \tau_{j,n})$ is the service time or the vacation time. In the course of our analysis we adopt the following three assumptions.

A1. The sequence $\{C^*_{j,k}\}$ is a strictly stationary (ergodic) random sequence with average $C^*_j = E(C^*_{j,k})$.

Note that "boldface" $C^*_{j,k}$ denotes a random sequence while "romanface" $C^*_j$ denotes the average of $C^*_{j,k}$. 
A2. The evolution of the system up to time $t$ is independent of the arrival process in $(t, \infty)$.

A3. The sequence $\{V_{j,t}\}$ of vacation times is a strictly stationary sequence with finite mean, that is, $E V_{j,t} < \infty$.

The next two examples specify possible distributed algorithms for server behavior and illustrates the definition of $C_{j,n}$, $C_{j,k}^*$ and $V_{j,t}$.

EXAMPLE 4.1. *Token passing ring* [KUE79, WAT84, BoG87, BoG88].

In this system, $M$ queues (users) are handled by a single token (server), which visits the queues in a cyclic order. It is assumed that a *walking time*, $W_j$, is required to switch from queue $j$ to $(j + 1) \mod M$. More specifically, when the server visits the $j$-th queue, it serves at most one customer, then walks in time $W_j$ to the $(j + 1)$-st queue, etc. The sequence $C_{j,n}$ is defined as the sequence of time intervals which have elapsed between two consecutive visits of the server to the $j$-th queue. The vacation $V_{j,t}$ is the time the server is away from the $j$-the empty queue, and the modified service time $C_{j,k}^*$ represents the period of time the token is away from the $j$-th nonempty queue. Actually, the behavior of any queue in the system is illustrated in Figure 1.

EXAMPLE 4.2. *Buffered ALOHA system* [SaE81, SZP86, RaE89].

There are $M$ distributed users, each having an infinite buffer for storing fixed-length packets. The packets are transmitted through a broadcast channel. The channel is slotted, and a slot duration is equal to a packet transmission time. Each nonempty user transmits a packet with a probability $r_i$ in a slot, where $i \in M$. If two or more users transmit simultaneously, then a collision occurs and the packets must be retransmitted in the future. When exactly one packet is transmitted in a slot, then a successful transmission takes place. Referring to our multiqueue model, we say that the server (channel) visits all queues simultaneously at the end of each slot. However, a successful visit occurs if and only if, successful transmission takes place or the
queue is empty. The end of a successful transmission or the end of a slot in which a new customer arrives to an empty queue is denoted by \( \tau^*_j \). Therefore, \( C^*_j \) is the time between the end of a successful transmission or the end of a slot in which a newly arrived customer found the queue empty, and the end of the next successful transmission. The vacation time falls into the idle time, so any queue in this system can be interpreted as a synchronized (slotted) M/G/1 queue without vacation, but with dependent service times. □

Now we are ready to present our stability condition for the process \( N' = (N^1_1, \ldots, N^m_k) \).

Our main result of this section is as follows.

**Theorem 10.** Under assumptions A1, A2 and A3, the process \( N' \) satisfying (4.3) is substable if

\[
\lambda_j C^*_j < 1 \quad \text{for all} \quad j \in \mathcal{M} 
\]

and is unstable if

\[
\lambda_j C^*_j > 1 \quad \text{for at least one} \quad j \in \mathcal{M} 
\]

**Proof.** Before we give a proof, let us briefly explain the idea. Let us concentrate on one queue, say \( j = 1 \). From the description of our model, we know that (4.3) holds for \( j = 1 \), whence \( C^*_1 \), as defined in (4.1), can be interpreted as a (modified) service time in an M/G/1 queue with vacation \( V_1 \), as defined in (4.2). The process represented by (4.3) is not Markovian. Assume, however, for a moment that (4.3) represents a Markov chain. Then by Theorem 1 and Theorem 4 with a linear Lyapunov function (see Lemma 2 and 3) such a queue is stable if and only if \( E\{X_1(\tau_{n+1}, \tau^*_n)\} < 1 \). But with A2, we have \( E \{X_1(\tau_{n+1}, \tau^*_n)\} = \lambda_1 C^*_1 \), as required in (4.4).

Fortunately, under assumption A1 Theorem 8 shows that for the M/G/1 queue without vacation, (4.4a) (for \( j = 1 \)) is sufficient for stability, and (4.4b) (for \( j = 1 \)) is sufficient for instability of (4.3), even when the service times are dependent, as long as A1 holds. If, in addition, A3 holds, then the same conditions (4.4a) and (4.4b) are sufficient for stability and instability respectively of an M/G/1 queue with vacation. The proof of this fact trivially extends the result of Loynes given in Theorem 8, and can be extracted from the discussion in [DOS85] (see
Theorem 1 and hypothesis H2 in [DOS85] and [BOR76]. This confirms our intuition and the well known result from queues with renewal processes, namely, that finite vacations do not affect stability conditions. Finally, by the isolation Lemma 4, condition (4.4a) is sufficient for substability of $N_t$, and by Lemma 5 condition (4.4b) is sufficient for instability.

In some applications the assumptions A1 and A3 regarding strict stationarity of the modified service times and vacation times are too strong, hence Theorem 9 can be useful if one agrees to switch to a different stability concept. Otherwise, we may apply a result of Borovkov [BOR76] who extended Theorem 8 proving that strict stationarity of the interarrival times and service times in a single GI/GI/1 queue can be replaced by asymptotic stationarity [BOR76, p. 12]. So, we can relax our assumptions A1 and A3, and adopt the following two modified postulates.

A1' The sequence of modified service times is asymptotically stationary, that is, the sequence \[
\{C_{j,k+N,k>0}\}
\]
converges as a process with $N \to \infty$ to a strictly stationary sequence \[
\{C^*_{j,k}, k>0\}.
\]

A3' The sequence of vacations is asymptotically stationary.

Then the following corollary to Theorem 10 can be established.

Corollary 6. If A1' and A3' replace assumptions A1 and A3 in Theorem 10, then the thesis (4.4) of the theorem holds.

While stability criteria (4.4) appear to be simple, complications arise when one attempts to compute the average modified service time $C^*_{j,k}$ for a particular system since this quantity may depend on input rates $\lambda_k$, $k \neq j$ and the conditional behavior of some subsystems of the system.

---

† In [DOS85] Doshi shows that the waiting time $W_k$ in an M/GI/1 with vacation and the waiting time $w_k$ in a queue without vacation are related by the following stochastic formula $W_k = w_k + D_k$ where $D_k$ depends on the vacation time and the idle times. This formula, first of all, satisfies the monotonicity criterion required in the Loynes proof, and secondly under assumption A3 Doshi shows that $D_k$ has a limit as $k \to \infty$, so the stability condition for $W_k$ and $w_k$ is the same.
Nevertheless, the criteria (4.4) establish the ultimate goal we need to achieve in order to prove stability. If, for some reason, $C_j^*$ is difficult to compute, the Theorem 10 and Corollary 6 can be used to derive sufficient conditions for stability and sufficient conditions for instability. Indeed, let us assume we can bound the average $C_j^*$ by $\overline{C}_j^*$ from below, and by $\overline{C}_j$ from above, that is, $C_j^* \leq C_j^* \leq \overline{C}_j$. Then $\lambda_j \overline{C}_j < 1$ for all $j \in \mathcal{M}$ implies that $\lambda_j C_j^* < 1$, hence stability. On the other hand, if for some $j$, $\lambda_j C_j^* > 1$, then $\lambda_j C_j^* > 1$, and instability follows.

Corollary 7. Let $C_j^* \leq C_j^* \leq \overline{C}_j^*$. (i) If for all $j \in \mathcal{M}$

$$\lambda_j \overline{C}_j < 1$$

(4.5)

then the system is stable.

(ii) If for some $j \in \mathcal{M}$

$$\lambda_j C_j^* > 1$$

(4.6)

then the system is unstable. ■

It must be stressed, however, that verifying stationarity assumptions A1 and A3 can lead to major difficulties in assessing stability of some computer communication systems. Therefore, we present below a set of conditions which are sufficient to verify (asymptotic) stationarity of the modified service times $C_{j,k}$ and vacation times $V_{j,l}$. More precisely, we shall show that replacing assumptions A1 and A3 by some other hypotheses, which are easier to verify in practice, leads also to stability condition (4.4). In particular, it turns out that to establish the fact that condition (4.4a) is necessary for stability of $N'$ is a rather easy task, and this can be done under a fairly general hypothesis. We adopt the following two assumptions.

(A) Let $Y'_i = (Y'_1, Y'_2, \ldots, Y'_k)$ where $Y'_i = \chi(N'_i)$ with $\chi(0) = 0$ and $\chi(x) = 1$ for $x > 0$.

Then the modified service time $C_{j,k}$ and the vacation time $V_{j,l}$ can be represented as $C_{j,k} = f(Y'_1, X'_1, \tau_{j,k} < t < \tau_{j,t+1})$ and $V_{j,l} = g(Y'_1, X'_1, \tau_{j,l} < t < \tau_{j,t+1})$. The process $X'_1$ is a stationary sequence, and $\tau_{j,k}$ and $\tau_{j,l}$ are the $k$-th and the $l$-th successful visits to a
nonempty and empty $j$-th queue, respectively. (For example, in the token passing ring from Example 4.1, one shows that $C_{j,k}^* = S_j + W_k^0 + \sum_{l=1, l \neq j}^M Y_l^k \cdot S_j^l$ for some $t_l$ such that $\tau_{j,k}^* < t_l < \tau_{j,n+1}$, where $S_j, W_0^0$ are stationary random sequences representing the service times at the $j$-th station and the total walking time, respectively.)

(B) $N'$ is an aperiodic irreducible Markov chain.

Then, one can easily prove the following lemma.

**Lemma 6.** If $N'$ is stable and (A), (B) and A2 hold, then $\lambda C_j^* \leq 1$ for all $j \in \mathcal{M}$, that is, condition (4.4a) is satisfied.

**Proof.** Since $N'$ is stable and it is a Markov chain, hence a stationary distribution exists. Let the initial distribution of this Markov chain be the same as the stationary one. Then, $N'$ is stationary, and by (A) the process $Y'$ is stationary. This implies that A1 and A3 hold, and this completes the proof. ■

These two assumptions are not yet sufficient to establish stability. To prove sufficient conditions for stability we must show that the sequence $Y'$ converges in distribution to a stationary one. This has to be proved in the the case when some queues are stable while the others are unstable. To formulate it more rigorously, let us partition the set of all queues $\mathcal{M} = \{1, 2, \ldots, M\}$ into two disjoint sets, namely, a set of stable queues $\mathcal{A}$, and a set of unstable queues $\mathcal{U}$, i.e., $\mathcal{M} = \mathcal{A} \cup \mathcal{U}$. In the same manner, we partition the processes $N'$ and $Y'$, that is, $N'=(N_{\mathcal{U}'}, N_{\mathcal{A}'})$ and $Y'=(Y_{\mathcal{U}'}, Y_{\mathcal{A}'})$. In addition, to describe all possible states of the process $Y'$, we introduce an $M$ dimensional zero-one vector $z = (z_1, \ldots, z_M)$ such that for every $j \in \mathcal{M}$ the $j$-th component $z_j$ is either zero or one, i.e., $z_j \in \{0, 1\}$. The set of all zero-one $M$-tuples is denoted as $\Theta_M$, i.e.,

$$\Theta_M = \{z: z = (z_1, \ldots, z_M), \quad z_j \in \{0, 1\}, \quad j \in \mathcal{M}\} \quad (4.7)$$

Moreover, any vector $z$ can be partitioned as $z=(z_{\mathcal{U}}, z_{\mathcal{A}})$, where $z_{\mathcal{U}}$ represents states of the
process $Y^t$ while $z_\Delta$ describes states for $Y^t$. By $0=(0_u, 0_\Delta)$ and $1=(1_u, 1_\Delta)$ we mean the all-zeros and all-ones vectors, respectively. To prove a sufficient condition for stability we need to show that $\lim_{t \to \infty} Pr\{ Y^t = z \}$ exists. Therefore, we adopt one more assumption, namely

(C) Let $\mathcal{U} \neq \emptyset$ and $\mathcal{U} \neq \mathcal{M}$. Then, for every $k \in \mathcal{U}$ we assume the following

$$\lim_{t \to \infty} Pr\{ N_k^t = 0 \} = 0$$  \hspace{1cm} (4.8)

Then, we can prove

Lemma 7. If (A), (B) and (C) hold, then condition (4.4a) is sufficient for stability of the process $N^t$.

Proof. We must prove that $\lim_{t \to \infty} Pr\{ Y^t = z \}$ exists for every $z \in \Theta_M$. Then our theorem follows from Corollary 6. The value of the probability $Pr\{ Y^t = (z_u, z_\Delta) \}$ depends on whether $z_u = 1_u$ or not. If $z_u \neq 1_u$, then there exists a $k \in \mathcal{U}$ such that $z_k = 0$. It can be proved [SzR88] that $\lim_{t \to \infty} Pr\{ Y^t = z \} = 0$ in this case. So, now we turn to the case $z_u = 1_u$. Then,

$$Pr\{ Y_\Delta = z_\Delta \} - \sum_{k \in \mathcal{U}} Pr\{ Y_k^t = 0 \} \leq Pr\{ Y^t = (1_u, z_\Delta) \} \leq Pr\{ Y^t \sb{\Delta} = z_\Delta \}$$

But, condition (4.8) from assumption (C) implies that the LHS of the above is equal to the RHS, so $\lim_{t \to \infty} Pr\{ Y^t = (1_u, z_\Delta) \} = \lim_{t \to \infty} Pr\{ Y^t \sb{\Delta} = z_\Delta \}$, and the latter limit exists since $Y^t \sb{\Delta}$ is a stable process. $\blacksquare$

There are, however, situations when checking condition (4.8) in assumption (C) is rather troublesome. Therefore, we suggest yet another approach, which ideally applies to the systems we plan to study in this section. We replace assumption (C) by a more restrictive one, namely

(C') Let for every $k \in \mathcal{U}$ the $k$-th queue be never empty, that is, $N_k^t \geq 1$ for every $t = 0, 1, \cdots$.

We denote such a modified queue by $\tilde{N}^t_k$ for $k \in \mathcal{U}$. Naturally,

$$[N_k^t]_{k \in U} \leq [	ilde{N}^t_k]_{k \in \mathcal{U}}$$  \hspace{1cm} (4.9)
where $\leq$ means "stochastically smaller" [ST83]. In addition, we assume that the remaining queues form an $\mathcal{J}$-dimensional Markov chain denoted by $\overline{N}_\mathcal{J} = (\overline{N}_t)_{t \in \mathcal{J}}$ and 

$$\{N_t\}_{t \in \mathcal{J}} \leq \{\overline{N}_t\}_{t \in \mathcal{J}} \quad (4.10)$$

Also, for the process $\overline{N} = (\overline{N}_t, \overline{N}_p)$ we denote by $\overline{C}_j$ the average modified service time, and 

$$C_j^* \leq \overline{C}_j^* \quad (4.11)$$

is assumed.

Note that (4.9) implies condition (4.8) required for Lemma 7. Since $N' \leq \overline{N}'$, then by our comparison test Theorem 7 stability of $\overline{N}'$ implies stability of the original process $N'$. Together with (4.11) we immediately find the following conclusion.

**Corollary 8.** Assume that (A), (B) and (C') hold. If $\lambda_j \overline{C}_j^* < 1$ for all $j = 1, 2, \ldots, M$, then the original process $N'$ is stable. \[ \blacksquare \]

If the process $\overline{N}'$ bounds the original process $N'$ very tightly, then we may expect that for some systems $C_j^* = \overline{C}_j^*$. In such a case the condition (4.4a) is sufficient for stability of $N'$.

**NOTES**

This section is almost exclusively based on Szpankowski and Rego paper [SzR87]. In particular, Lemmas 4 and 5, the main result Theorem 10, and Lemmas 6 and 7 come from this work. However, study on this kind of stabilites (i.e., stationary distribution in general stochastic models) was initiated by Loynes in 1962 in his seminal paper [LOY62]. Theorem 8 was proved there. His work was continued and extended by Borovkov [BOR76, BOR78] (see also [ROL81, BaB87, WAR88]). Finally, another twist in the stability area is suggested in Theorem 9 proved by Szpankowski in his preliminary technical report [SZP89b].

**5. APPLICATIONS TO SOME COMPUTER AND COMMUNICATION SYSTEMS**

In this section, we apply our criteria from Sections 3 and 4 to establish stability conditions for some computer and communication systems such as token passing rings (Sec. 5.1), coupled-processor systems (Sec. 5.2), buffered ALOHA systems (Sec. 5.3), and a decentralized dynamic control multiaccess protocol (Sec. 5.4). These applications will undoubtedly show a
superiority of the non-Markovian approach from Section 4. However, this approach might not
give explicit formulas in some cases (see Sec. 5.3). Then, methods of Section 3 are very useful.

5.1 Token passing ring [KUE79, WAT84, BoG87, BoG88]

We analyze the token passing ring system described in Example 4.1. Briefly, we recall
that the system consists of \( M \) users each containing an infinite capacity buffer. A server (token)
visits all queues in a cyclic order. The average transmission time (service time) is denoted by
\( h_j, j \in M \), and the walking time required to switch from queue \( j \) to \( j + 1 \) mod \( M \), is denoted by
\( W_j \). We establish stability of this type of system by appealing to our non-Markovian approach.
In particular, we shall use Theorem 10 and Corollary 6. However, in order to circumvent sta-

tionarity requirement A1 and A3 we shall apply Lemma 6, Lemma 7 and Corollary 8.

We first establish a necessary condition for stability by appealing to Lemma 6. The sys-
tem is described by an \( M \)-dimensional process \( N_t' = (N'_1, \ldots, N'_M) \) where \( N'_j \) is the queue
length at the \( j \)-th user at time \( t \). In general, \( N'_t \) is not a Markov chain, but \( N'_t \) becomes a Mark-
kov chain if one imbeds the process at the token scan instants of all queues. Naturally, assump-
tion (A) is satisfied, so we can refer to our Lemma 6. Therefore, we assume that the process \( N'_t \)
is stationary by selecting an appropriate initial distribution. To evaluate \( c_j^* \), we need a little bit
of notation. As before, \( \Theta_M \) defined in (4.7) represents the set of zero-one \( M \)-tuples. In addition,
\( z^{(j)} \in \Theta_{M-1} \), denotes an \((M - 1)\)-tuple with the \( j \)-th coordinate missing, that is,

\[
z^{(j)} = (z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_M) \in \Theta_{M-1}
\]

(5.1)

Finally, since only empty and nonempty buffers are important for stability we adopt the follow-
ing definition

\[
P(z^{(j)}) = Pr\{ Y^*_k = z_k, \quad k \in M - \{j\} | N^*_j > 0 \}
\]

(5.2)

where \( \tau^*_j < t < \tau^*_j,n+1 \). For example, for \( M = 3 \) and \( z^{(2)} = (1, 0) \)

\[
P(z^{(2)}) \overset{\text{def}}{=} P_2(1, 0) = Pr\{ N'_1 \geq 1, N'_3 = 0 | N^*_2 > 0 \},
\]

and this represents the conditional proba-
bility that the first buffer is nonempty, while the third is empty. We emphasize here the fact that \( P(x(t)) \) does not depend upon the time \( t \) since the process \( Y_t = \chi(N_t^1) \) is stationary by selecting an appropriate initial distribution (such a distribution exists since by our assumption the process is a stable Markov chain).

By the virtue of the above, the average of the modified service time \( C_j^* \) for the \( j \)-th user is

\[
C_j^* = \sum_{x^{(0)} \in \Theta_{x^{(1)}}} P(x^{(0)}) \sum_{k=1}^{M} \left[ \chi(z_k) h_k + w_k \right] + w^*_j + h_j
\]  

(5.3)

and \( w_i = E W_i \) with \( w_0 = \sum_{i=1}^{M} w_i \). Note also that

\[
\sum_{\{x^{(0)} z_0 = 1\}} P(x^{(0)}) = \Pr\{N_k^1 \geq 1 | N_j^1 > 0\}
\]  

(5.4)

So, after grouping all probabilities with the coefficient \( h_k, k \neq j \), one finds

\[
C_j^* = w_0 + h_j + \sum_{k=1}^{M} h_k \Pr\{N_k^1 \geq 1 | N_j^1 > 0\}
\]

Now, for a stable (and unstable) system, by balance flow arguments [KUE79, WAT84], the following holds

\[
\Pr\{N_k^1 \geq 1 | N_j^1 > 0\} = \min\{1, \lambda_k C_j^*\}
\]  

(5.5)

Hence, after some manipulation, one proves

\[
C_j^* = \frac{w_0 + h_j}{1 - \rho_0 + \rho_j}
\]  

(5.6)

where \( \rho_j = \lambda_j h_j \) and \( \rho_0 = \sum_{j=1}^{M} \rho_j \). By Lemma 6, \( \lambda_j C_j^* \leq 1 \) for all \( j \in M \) is necessary for stability of the system, if one understands \( C_j^* \) in the sense of (5.6). It is not difficult in this case to show that (5.6) is also necessary for stability (for details see [SzaR87]). This proves that the following condition

\[
\lambda_j < \frac{1 - \rho_0 + \rho_j}{w_0 + h_j} \quad j = 1, 2, \ldots, M
\]  

(5.7)
is sufficient and necessary for stability of the system (except when there is an equality in (5.7), however, it is reasonable to conjecture that the system is unstable in this case too).

NOTES

(i) The stability conditions (5.7) for a token passing ring with all infinite buffers have been intuitively derived by Kuehn [KUE79]. As it was pointed out by Watson [WAT84], it is convenient to derive such stability conditions for other modified token passing rings, based on Kuehn's analysis, however, without proof. Some generalization of these conditions are possible. For example, if one assumes that at most \( s_j \) messages are removed from the \( j \)-th queue, then the modified service time, \( C_j^* \), becomes

\[
C_j^* = \frac{w_0 + s_j h_j}{1 - \rho_0 + \rho_j}
\]

and \( \lambda_j C_j^* < s_j \) for all \( j \in M \) is the stability condition.

(ii) It is important to understand why in the case of the token passing ring, we have been able to compute exact stability conditions, that is, to evaluate \( C_j^* \). Note that knowing the vector \( x^{(0)} \) (i.e., under the condition that \( x(N_1) \leq \ldots \leq x(N_{j-1}) \leq x(N_{j+1}) \leq \ldots \leq x(N_M) \) = \( x^{(0)} \) the conditional modified service time for the \( j \)-th station is a linear function of the average service times of those stations for which the buffer is nonempty. This allows us to group the joint probabilities \( P(x^{(0)}) \) such that the coefficient at \( h_k \) is a one dimensional probability (5.5), which is easy to evaluate. If the above grouping does not work, then joint distributions appear in the expression for \( C_j^* \) and this causes additional difficulties.

5.2 Coupled-Processors System

In [Fal79] (see also [CoB83]), Fayolle and Iasnogrodski described a coupled-processor system. A queueing model for this consists of two M/M/1 queues with infinite capacities. The service rate of each server is \( \mu_1 \) and \( \mu_2 \) respectively, if the queues are nonempty. If the second queue is empty, then the service rate for the first queue is \( \mu_1^* \); and reverse, the second queue serves with rate \( \mu_2^* \) if the first queue is empty. To establish stability condition we may apply either Rosenkrantz's Theorem 6b or our non-Markovian approach from Section 4. Since Theorem 6b requires some algebra, we rather use Theorem 10. It is easy to check that assumption (A), (B) and (C) are satisfied in this system, so Lemmas 6 and 7 give ready to apply approach.

For stability purposes, it is convenient to deal with the modified service rate, i.e., \( 1/C_j^* \), \( j = 1, 2 \). For obvious reasons, we have

\[
\frac{1}{C_1^*} = \mu_1^* Pr \{ N_2^* > 0 | N_1^* > 0 \} + \mu_1^* Pr \{ N_1^* = 0 | N_1^* > 0 \}
\]

(5.8)
\[ \frac{1}{C^2_2} = \mu_2 \Pr \{ N_1 > 0 \mid \bar{N}_2^* > 0 \} + \mu_2^* \Pr \{ N_1^* = 0 \mid \bar{N}_2^* > 0 \} \quad (5.9) \]

But, the following holds

\[ \Pr \{ N_2^* = 0 \mid \bar{N}_1^* > 0 \} = \max \{ 0, 1 - \lambda_2 / \mu_2 \} \quad (5.10) \]

\[ \Pr \{ N_1^* = 0 \mid \bar{N}_2^* > 0 \} = \max \{ 0, 1 - \lambda_1 / \mu_1 \} \quad (5.11) \]

Therefore, from the above and Theorem 10 we immediately show that the system is stable if and only if

\[ \lambda_1 < \mu_1^* + \frac{\lambda_2}{\mu_2} (\mu_1 - \mu_1^*) \quad (5.12) \]

\[ \lambda_2 < \mu_2^* + \frac{\lambda_1}{\mu_1} (\mu_2 - \mu_2^*) \quad (5.13) \]

**NOTES**

(i) Conditions (5.13) coincide with the stability criteria established in [Fal79] and [CoB84]. Note, however, that the authors of [Fal79] and [CoB84] used the Riemann-Hilbert problem to obtain (5.13). Other systems described in [CoB84] can be analyzed, from the stability viewpoint, in a similar way. A generalization to \( M \) coupled processors, as described in [SZP88], is possible.

(ii) Both inequalities, (5.12) and (5.13) must be satisfied simultaneously for establishing stability regions. Note, however, that during the course of the derivation, we have to concentrate on one queue, say the first one. Then, according to (5.10) and (5.11), two regions must be considered, \( \lambda_2 \leq \mu_2 \) and \( \lambda_2 > \mu_2 \). In the first region (see (5.12))

\[ \lambda_1 < \mu_1^* + \frac{\lambda_2}{\mu_2} (\mu_1 - \mu_1^*) \quad (5.14a) \]

while in the second region (i.e., \( \lambda_2 > \mu_2 \))

\[ \lambda_1 < \mu_1 \quad (5.14b) \]

Nevertheless, the union of the regions (5.14a) and (5.14b) is contained in the intersection of the regions from (5.12) and (5.13). If we reverse the queues and concentrate on the second queue, we obtain two inequalities similar to (5.14), that is, one as (5.13b) and the second \( \lambda_2 < \mu_2 \). The intersection of these regions and the one established in (5.14), coincides with (5.12) - (5.13).

5.3 Buffered ALOHA system [TsM79, SaE81, SZP86, RaE89]

The buffered ALOHA system was described in Example 4.2. It consists of \( M \) buffered users. The channel (server) is slotted and the duration of a slot is equal to a fixed-packet length transmission time. At the beginning of a slot, the \( j \)-th user with a nonempty buffer transmits
with probability \( r_j \), and delays transmission for one slot with probability \( \overline{r}_j = 1 - r_j \). If two or more users transmit simultaneously, then a collision occurs (unsuccessful transmission) and the colliding users repeat transmission in the future according to the above described random procedure.

The system is described by an \( M \)-dimensional Markov chain \( N' = (N'_1, \ldots, N'_M) \) where \( N'_j \) represents the number of packets in the \( j \)-th queue at the \textit{end} of the \( t \)-th slot, \( t = 1, 2, \ldots \). We first deal with the necessary stability condition, so Lemma 6 is applied. Actually, we assume a stationary version of the stable system, and for stability we compute the average \( C_j^* \) of the modified service time, which is the average time between two successful transmissions from the \( j \)-th user. In this case, however, it is more convenient to deal with the probability of a successful transmission \( P^{(j)}_{\text{success}} \) (in a slot), instead of \( C_j^* \). These two quantities are related by \( C_j^* = 1/P^{(j)}_{\text{success}} \).

The probability \( P^{(j)}_{\text{success}} \) is a conditional probability of a successful transmission from the \( j \)-th user under the condition that \( N'_j > 0 \). By Theorem 10, stability of the ALOHA system implies that

\[
\lambda_j < P^{(j)}_{\text{success}} \quad \text{for all} \quad j \in \mathcal{M}
\]  

(5.15)

In order to evaluate \( P^{(j)}_{\text{success}} \) we note that it depends only on the probabilities of emptiness of the other buffers, so notation from the previous sections is adopted here. In particular, we define the probability \( P(z^{(j)}) \) as

\[
P(z^{(j)}) = Pr\{ Y_k^j = z_k, \quad k \in \mathcal{M} - \{j\} \mid N'_j > 0 \}
\]  

(5.16)

where, as before, \( Y_k^j = \chi(N_k^j) \), and \( t = 0, 1, \cdots \). Then, one immediately obtains

\[
P^{(j)}_{\text{success}} = r_j \sum_{z^{(j)} \in \Theta_{\mathcal{M} - j}} P(z^{(j)}) \prod_{k=1}^{M} (1 - r_k)^{\chi(k)}
\]  

(5.17)

As long as a sufficient condition is concerned, we adopt the approach from Corollary 8. To recall, we divide the set of users \( \mathcal{M} - \{j\} \) into stable and unstable subsets. For unstable users we
assume that they are never empty (for example, by transmitting dummy packets). Then the system of stable queues is a Markov chain [SZP86, RaE89], and the original process is upper bounded by the modified process $N'$ as defined in assumption (C'). Let $\tilde{P}_{\text{succ}}(\cdot)$ be the probability of success in the modified system for any partition of the user set into stable and unstable queues. This probability is given exactly by the same formula as $P_{\text{succ}}(\cdot)$ (see (5.17)) except that the probability $P(\cdot)$ is replaced by $\tilde{P}(\cdot)$ for the upper bounding system (for details see Section 5.0). In particular, Corollary 8 implies that

$$\lambda_j < \tilde{P}_{\text{succ}}^{(j)} \quad \text{for all } j = 1, 2, \ldots, M$$

(5.18)
is sufficient for ergodicity of the system. We shall not argue here whether $\tilde{P}_{\text{succ}}^{(j)}$ is equal to $P_{\text{succ}}(\cdot)$ or not since none of these probabilities, as we shall see, can be computed. We shall use some other arguments to obtain computable stability conditions. We must, however, mention here that (5.15) is sufficient and necessary for $M=2$ and $M=3$ users (for details see [SzR87]).

The case $M=2$ can be analyzed by Rosenkrantz's Theorem 6b, however, we shall show below how our non-Markovian analysis can be applied to get the same result. We know that (5.15) is sufficient and necessary for stability in this case. In particular, (5.17) implies

$$\lambda_1 < P_{\text{succ}}^{(1)} = r_1 \{ P_1(0) + \bar{r}_2 P_1(1) \}$$

(5.19)

where $P_1(0) = 1 - P_1(1) = \Pr \{ N_2 = 0 \mid N_1 > 0 \}$. Since the first buffer is nonempty, this probability can be easily computed from statistical equilibrium arguments, that is,

$$P_1(0) = \max \{ 0, 1 - \frac{\lambda_2}{r_2(1 - r_1)} \}$$

Two cases must be considered: (i) $\lambda_2 < r_2 \bar{r}_1$ and (ii) $\lambda_2 > r_2 \bar{r}_1$. In the first case, the second queue is stable (precisely: conditionally stable), while in the second case, the second queue is unstable. For $\lambda_2 < r_2 \bar{r}_1$, (5.19) implies

$$P_{\text{succ}}^{(1)} = r_1(1 - \lambda_2 \bar{r}_1)$$

(5.20a)

while for $\lambda_2 > r_2 \bar{r}_1$, 
Reversing the queues, one immediately obtains the following stability region

\begin{align}
\lambda_1 &< r_1 (1 - \lambda_2 / \bar{r}_1) \\
\lambda_2 &< r_2 (1 - \lambda_1 / \bar{r}_2)
\end{align}

where both conditions (5.21) and (5.22) must be simultaneously satisfied.

Now we consider the case of \( M = 3 \) users which is by far the more difficult. We focus our attention on the first user. Then (5.15) and (5.17) imply

\begin{equation}
\lambda_1 < P_{\text{succ}}^{(1)} = r_1 [P_1(0, 0) + \bar{r}_2 P_1(1, 0) + \bar{r}_3 P_1(0, 1) + \bar{r}_1 \bar{r}_2 P_1(1, 1)] = \\
r_1 [1 - r_2 Pr(N_2 > 0|N_1 > 0) - r_3 Pr(N_3 > 1|N_1 > 0) + r_2 r_3 P_1(1, 1)]
\end{equation}

where the notation was explained earlier. We consider three cases (i) both queues, the second and the third, are unstable (i.e., \( \lambda_2 \) and \( \lambda_3 \) are "large"), (ii) either the second or the third queue is stable and the other unstable and (iii) both queues are stable (\( \lambda_1 \) and \( \lambda_3 \) are "small"). The third case is the most difficult to analyze. For the first case we easily show that

\begin{align}
P_1(0, 0) = P_1(0, 1) = 0 \quad P_1(1, 1) = 1 \\
\bar{P}_1(1, 1) = 1 - P_1(1, 0) = Pr(N_3 > 0|N_1 > 0) = \lambda_3 / \bar{r}_1 \bar{r}_2.
\end{align}

Moreover, (5.23) implies

\begin{equation}(5.24a)\end{equation}

\begin{align}
\lambda_1 &< r_1 \bar{r}_2 \bar{r}_3
\end{align}

is sufficient for ergodicity. We shall soon see that this case is, in fact, entirely covered by the second case, which is discussed next.

In the second case, we can safely apply Lemma 6 since assumption \((C')\) holds in this case, so (5.23) is sufficient and necessary for stability of the system. Let us assume that the third queue is unstable and the second queue is stable. Then \( P_1(0, 0) = P_1(0, 1) = 0 \) and

\begin{align}
P_1(1, 1) = 1 - P_1(1, 0) = Pr(N_3 > 0|N_1 > 0) = \lambda_3 / \bar{r}_1 \bar{r}_2.
\end{align}

Moreover, (5.23) implies

\begin{equation}(5.24b)\end{equation}

\begin{align}
\lambda_1 &< r_1 \bar{r}_2 \left[ 1 - \frac{\lambda_3}{\bar{r}_1 \bar{r}_2} \right]
\end{align}

and reversing the condition imposed on the second and third queue, one obtains

\begin{equation}(5.24c)\end{equation}

\begin{align}
\lambda_1 &< r_1 \bar{r}_3 \left[ 1 - \frac{\lambda_2}{\bar{r}_1 \bar{r}_2} \right]
\end{align}
In the third case, we must compute the joint probabilities $P_1(0, 0), P_1(1, 0), P_1(0, 1)$ and $P_1(1, 1)$. Note that these probabilities are estimated under the condition that the first queue is nonempty. There is a relationship between these probabilities, that is, $P_1(1, 0), P_1(0,1),$ and $P_1(1,1)$ can be expressed as a function of $P_1(0,0)$. The latter probability can be, on the other hand, computed as in [NAI85] (see (4.10) in [NAI85]), where Nain solved (exactly) a two-user buffered ALOHA system. After some algebra, the following stability region was derived in [SzR87]

$$\lambda_1 < P_{\text{succ}}^{(1)} = r_1 \left\{ 1 - \frac{\lambda_2 \overline{r}_2 \overline{r}_1 + \lambda_3 \overline{r}_3 \overline{r}_1}{1 - r_2 - r_3} [P_1(0, 0) - 1] \right\} \quad (5.25)$$

$$\lambda_2 < P_{\text{succ}}^{(2)} = r_2 \left\{ 1 - \frac{\lambda_1 \overline{r}_2 \overline{r}_2 + \lambda_3 \overline{r}_3 \overline{r}_2 + r_1 r_3 [P_2(0, 0) - 1]}{1 - r_1 - r_3} \right\} \quad (5.26)$$

$$\lambda_3 < P_{\text{succ}}^{(3)} = r_3 \left\{ 1 - \frac{\lambda_1 \overline{r}_3 \overline{r}_3 + \lambda_2 \overline{r}_2 \overline{r}_3 + r_1 r_2 [P_3(0, 0) - 1]}{1 - r_1 - r_2} \right\} \quad (5.27)$$

Figure 2 presents boundary lines of the stability region for $M = 3$ with points $\omega = (\lambda_1, \lambda_2, \lambda_3) = (r_1 \overline{r}_2 \overline{r}_3, \overline{r}_1 r_2 \overline{r}_3, \overline{r}_1 \overline{r}_2 \overline{r}_3)$, $A = (r_2 \overline{r}_2, \overline{r}_1 r_2, 0)$, $B = (r_1, 0, 0)$, $C = (r_1 \overline{r}_3, 0, \overline{r}_1 r_3)$, $D = (0, 0, r_3)$, $E(0, r_2 \overline{r}_3, \overline{r}_2 r_3)$ and $F = (0, r_2, 0)$, explicitly shown.

The ultimate stability criterion (5.17) for the ALOHA system requires to estimate the probabilities $P(\omega^D)$. This is difficult as shown in the case $M = 3$, and there is no hope for computing these probabilities for higher dimensional cases. Therefore, another approach needs to be investigated, namely the one that concentrates on bounds for stability region of the ALOHA system. Such bounds can be derived from the Markovian methodology presented in Section 3.

At first, we investigate the application of the Lyapunov function method to estimate stability region of the system. In particular, we apply Corollary 2 and Corollary 5, taking into account Example 3.2. The example shows that the $m$-th component of the average drift can be computed as the difference between the conditional input rate and the condition throughput
$S_m^2(k)$. But, it is easy to see that for $k \in M$

$$S_m^2(k) = r_m \chi(k_m) \prod_{j=1}^{M} r_j \chi(k_j)$$

where $\chi(0) = 0$ otherwise $\chi(x) = 1$ for all $x > 0$. Letting $\mathcal{H} = (0, 0, \ldots, 0)$ in Corollaries 2 and 5, we know that $N'$ is ergodic if

$$\sum_{i=1}^{M} c_i \lambda_i < \sum_{i=1}^{M} c_i S^2_i(k)$$

and nonergodic if

$$\sum_{i=1}^{M} c_i \lambda_i \geq \sum_{i=1}^{M} c_i S^2_i(k)$$

for any choice of the constants $c_i$. But, by the property of the function $\chi(k)$, we may restrict a set of $k \in \mathcal{E}$ satisfying (5.28) to $\mathcal{E} \supset \mathcal{B} = \{(k_1, \ldots, k_M): k_i = 1$ or $k_i = 0$, $i \in M\}$.

Since the LHS of (5.28) is the same for all $k \in \mathcal{B}$, then for (5.28a) we must find the smallest value of the RHS of the inequality, while for (5.28b) the greatest value for the RHS of it must be determined. After some algebra we can prove the following.

**Property 1.** Let $\sum r_i \leq 1$. Then,

(i) $N'$ is ergodic if

$$\sum_{i=1}^{M} \frac{\lambda_i}{r_i} < 1$$

(ii) $N'$ is not ergodic if

$$\sum_{i=1}^{M} r_i \lambda_i \geq \prod_{i=1}^{M} r_i \sum_{j=1}^{M} r_j$$

or if for an $m \in M$

$$\lambda_m + \frac{r_m}{r_m} \sum_{j \neq m} \lambda_j \geq r_m$$

(ii) Suppose that for all $n, m \in M$, $r_n + r_m \geq 1$. Then $N'$ is not ergodic if
Proof. For (i) we set \( c_i = 1/r_i \) and for (ii) formula (5.29b) we assume \( c_i = \bar{r}_i \) for all \( i \in \mathcal{M} \). In the case of (5.29c) we put \( c_m = 1 \) and \( c_i = r_m/\bar{r}_m \) for \( i \in \mathcal{M} - \{ m \} \). Finally, (iii) is proved in the same manner as (i). For more details see [SZP88].

The bounds just derived are rather good for very asymmetric case, that is, when characteristics of one user differ significantly from the other users. However, these bounds are not very tight in a "semi-symmetric" case. To improve this situation we shall consider another approach, namely the one which is based on the comparison tests (Theorem 7) and the known stability conditions for one and two dimensional Markov chains (Lemma 2 and 3, and Theorem 6b). In particular, let \( \bar{N}_m^t \) and \( N_m^t \) denote the queue lengths in the \( m \)-th buffer of modified ALOHA systems assuming all other queues are never empty and always empty respectively. Then, naturally \( N_m^t \leq \bar{N}_m^t \) and \( \bar{N}_m^t \leq N_m^t \), so by comparison tests can be applied. We define \( q_m \) as

\[
q_m = \prod_{j=1}^{M} \frac{r_j}{\bar{r}_j}
\]

Then, Lemmas 2 and 3 together with Theorem 7 imply the following property.

**Property 2.** (i) The ALOHA system is ergodic if for all \( m \in \mathcal{M} \)

\[
\lambda_m < q_m
\]  

(ii) The ALOHA system is not ergodic if

\[
\lambda_m \geq r_m
\]

holds for an \( m \in \mathcal{M} \).}

Proving Property 2 we found that one dimensional Markov chains \( \bar{N}_m^t \) and \( N_m^t \) upper and lower bounded the \( m \) component \( N_m^t \) of \( N^t \). In other words, as a cover of \( \mathcal{M} \) we chose 1-tuples \( \sigma_1 = (1), \sigma_2 = (2), \ldots, \sigma_M = (M), P_M = \{ \sigma_1, \sigma_2, \ldots, \sigma_M \} \). However, due to Rosenkrantz (see Theorem 6b) we also know sufficient and necessary conditions for ergodicity of a class of two-dimensional Markov chains. In particular, we can prove a sufficient and necessary stability
conditions for a two-queue ALOHA system, which coincide with our stability conditions (5.21) and (5.22) derived above. To extend this result to arbitrary \( M \) we apply again comparison tests.

Let us define two dimensional Markov chains \( \overline{N}_{mn} \) and \( \underline{N}_{mn} \) as follows. The two dimensional Markov chain \( \overline{N}_{mn}(t) \) represents queue lengths in the \( n \)-th and the \( m \)-th queues under the condition that all other queues in the ALOHA system are never empty while \( \underline{N}_{mn}(t) \) models the queue lengths in the system under the condition that all other queues are always empty (or in other words, \( \overline{N}_{mn}(t) \) represents the ALOHA system with two queues, \( n \) and \( m \)). In addition, we introduce some more notations. Let for \( n, m \in \mathcal{M} \)

\[
\begin{align*}
    d_n &= \lambda_n - r_n \prod_{j=1}^{M} r_j, \\
    d'_n &= \lambda_n - r_n r_m \\
    a_n(m) &= d_n r_m + d_m r_n, \\
    a'_n(m) &= d'_n r_m + d'_m r_n
\end{align*}
\]

Then, using Theorem 7 and Theorem 6b we obtain the following refinement of Property 2.

Property 3. (i) The ALOHA system is ergodic if for every pair \( n, m \in \mathcal{M} \) the following holds

\[
\begin{align*}
    (i) & \quad a_n(m) < 0 \text{ and } a_m(n) < 0 \text{ if } r_n + r_m \leq 1 \quad (5.32a) \\
    (ii) & \quad a_n(m) < 0 \text{ or } a_m(n) < 0 \text{ if } r_n + r_m > 1
\end{align*}
\]

(ii) The Markov chain \( N' \) is not ergodic if the next condition is satisfied

\[
\begin{align*}
    (i) & \quad a'_n(m) \geq 0 \text{ or } a'_m(n) \geq 0 \text{ if } r_n + r_m \leq 1 \quad (5.32b) \\
    (ii) & \quad a'_n(m) \geq 0 \text{ and } a'_m(n) \geq 0 \text{ if } r_n + r_m > 1
\end{align*}
\]

Proof. Using the sample path arguments we show that the process \( N'_{mn} = (N'_m, N'_n) \) satisfies:

\( N'_{mn} \preceq_{st} \overline{N}_{mn} \) and \( \underline{N}_{mn} \preceq_{st} N'_{mn} \). Hence, Theorem 7 may be applied to determine stability conditions of \( N' \), and using Theorem 6b we prove the property. For more details see [SZP86, SZP88].

To obtain a more sophisticated bounds on stability one needs to introduce tighter upper and lower bounded systems. Recently, Rao and Ephremides [RaE89] have suggested a very tight dominant system of the ALOHA system. They define a sequence of dominant systems, and
in particular, $\Delta^j$ is a system that satisfies the following properties. The arrival processes to $\Delta^j$ and the ALOHA are (pathwise) exactly the same as well as the retransmission attempts. In addition,

- for $i > j$ users $i$ behaves exactly the same way as in the original ALOHA

- for $i \leq j$, user $i$ attempts to transmit "dummy" packets when empty (i.e., the user is never empty) according to the following rules: with the aid of a "genie" the $i$-th user is informed whether any user $k$, with $k < i$, will attempt a transmission in the slot; if yes, the user $i$ refrains from attempting to transmit; if no it attempts to transmit a "dummy" packet with probability $r_i$.

Using this type of dominant systems Rao and Ephremides [RaE89] derived the below stability conditions.

**Property 4.** The ALOHA system is stable if $\lambda_j < b_j$ for all $j \in M$ where

$$b_j = r_j \prod_{i=1, i\neq j}^M \overline{r_i} + \sum_{i=j+1}^M r_j (1 - \lambda_i / b_i) \prod_{k=1, k \neq i, j}^M \overline{r_k}$$

and $b_M = r_M \prod_{i=1}^M \overline{r_i}$.

**Proof.** This proof is much more complicated than the ones presented so far, and details can be found in [RaE89]. The idea is to show that the throughput of the dominant system $\Delta^j$ is smaller than equal to $b_j$ given by (5.33). This is done by appealing to Loynes’ result (Theorem 8), and the proof resembles our analysis from Section 4. □

Finally, using a different approach (in [SZP88] it is called the random walk method) we can also prove the following.

**Property 5.** The ALOHA system is not ergodic if

$$\lambda_m > q_m$$

for all $m \in M$ □
This and Property 2 formula (5.30) imply sufficient and necessary stability conditions for symmetric ALOHA system.

Property 6. Let \( \lambda = \lambda_m \) and \( r = r_m \) for all \( m \in M \). Then \( N^\prime \) is ergodic if and only if
\[
\lambda < r \gamma^{M-1}
\]
which settles stability condition for the symmetric case. ■

NOTES

The ergodicity analysis of the buffered slotted ALOHA system was initiated by Tsybakov and Mikhailov [TsM79] who obtained the bound (5.30) from Property 2. In particular, using Malyshev's condition [MAL72], they established the exact stability condition for \( M = 2 \) users, but for uniformly bounded arrival process. The ultimate stability criteria for ALOHA (5.17) are derived by Szpankowski and Rego in [SzR87]. They also present the exact stability conditions for \( M = 2 \) and \( M = 3 \) (see (5.21)-(5.22) and (5.25)-(5.27)). Properties 1 and 3 are established by Szpankowski in [SZP88]. Property 4 is proved by Rao and Ephremides [RaE89]. Property 5 and Property 6 are derived by Tsybakov and Mikhailov in [TsM79] and Szpankowski [SZP88]. Finally, Sharma in [SHA89] proved the bound (5.30) holds also for general stationary nonindependent arrival process.

5.4 A Decentralized Dynamic Control Algorithm [Hal82, KEL85, MIK89]

There is a variety of protocols for resolving collisions in a broadcast packet communications [KLE76], and they differ depending on how feedback information for a channel is used to resolve the collisions. To avoid subsequent collisions, a probability of retransmitting a collided packet is introduced (e.g., see Section 5.3 for ALOHA protocol), which controls the number of retransmissions. This probability depends on the outcomes of the channel, the time, the number of users involved in a collision, etc. In this section, we assume, in addition, an infinite population of users with single buffers. We illustrate stability analysis on a system with decentralized dynamic control algorithm proposed by Hajek and van Loon [HaL82] In a system implementing such a protocol every user contains a counter, \( S^t, t = 0, 1, \ldots \), which is updated recursively at the end of each slot according to some rules common for all users. For example, Hajek and van Loon [HaL82] assumed \( S^{t+1} = \max(S^t, 1, a(Z^t)) \) where \( Z^t \) is an outcome from the channel (idle, success or collision) and \( a(\cdot) \) is a function of \( Z^t \). Kelly [KEL85] proposed
\[
S^{t+1} = \max\{1, S^t + a I[Z^t = 0] + b I[Z^t = 1] + c I[Z^t = \text{collision}]\},
\]
where \( a, b, c \) are
constants and \( I(\cdot) \) is an indicator function of an event. The probability of transmitting a packet is the same for all users at a time \( t \) and is equal to \( f^t = 1/S' \).

The system is described by a two dimensional Markov chain \((N^t, S^t)\) where \( N^t \) is the backlog and \( S^t \) is the counter discussed above. The \( N \)-th component of the average drift is given by (see Kelly [KEL85])

\[
d(n, s) = E(N^t + 1 - N^t | N^t = n, S^t = s) = \lambda - n/s \cdot (1 - 1/s)^{n-1}. \tag{5.36}
\]

We prove that

**Property 7.** (i) The system is geometrically ergodic if \( \lambda < e^{-1} \).

(ii) If \( \lambda > e^{-1} \), then for any recursive formula on \( S^t, S' \geq 1 \), the average backlog is infinite, that is, \( \lim_{t \to \infty} E N^t = \infty \).

**Proof:** The part (i) of the proof was established in 1982 by Hajek [HAJ82] by a method of the average drift. A geometric proof of this was recently proposed by Mikhailov [MIK89]. The interested reader is referred to these papers. Here, we shall concentrate only on the proof of part (ii), and in particular we illustrate the usage of Theorem 5 with infinite set \( \mathcal{X} \). Let \( V(n, s) = n \) be the Lyapunov function required in Theorem 5, and we define for any number \( M > 0 \) an (infinite) set \( \mathcal{X}_M \) as \( \mathcal{X}_M = \{ n, s \): \( n < M \} \). It is easy to check that all hypotheses of Theorem 5 are satisfied, and to apply it we only need to verify condition (3.19). Naturally, \( AV(n, s) = d(n, s) \) given by (5.36). Using the inequality \( 1 - x \leq e^{-x} \) we find that

\[
n/s \cdot (1 - 1/s)^{n-1} \leq n/s \cdot \exp \left[ (n-1)/s \right] = \overline{f}_n(s). \tag{5.37}
\]

But, by simple algebra we obtain

\[
\max \overline{f}_n(s) = e^{-1} + e^{-1/(n-1)} \text{ for } s \geq 1. \tag{5.38}
\]

Let \( \varepsilon = \lambda - e^{-1} > 0 \). For \( \delta = \varepsilon/2 \) and for \( M > 1 + \frac{1}{\delta e} \), we bound the second term in (5.38) above by \( \delta \). Then if \( \lambda > e^{-1} \), we obtain, for \( (n, s) \in \mathcal{C} - \mathcal{X}_M \) and \( M > 1 + \frac{1}{\delta e} \), that
\[ AV(n, s) \geq \lambda - e^{-1} - \delta \geq \varepsilon - \delta = \varepsilon / 2 > 0. \] Hence condition (3.19) is satisfied, and therefore

\[ \lim_{t \to \infty} E(V(N', S')) = \lim_{t \to \infty} E(N') = \infty. \]

**NOTES**

The system discussed in this section was first described by Hajek and Loon [HaL82]. Hajek in [Haj82] presented first stability analysis of the system (Property 7 (i)). Also, a geometric approach to the stability analysis of the algorithm is discussed by Mihkalov in a number of papers, however, a good account of his methodology is presented in [MIK89]. The instability criterion of Property 7 (ii) is derived by Szpankowski and Rego [SzR88].

**REFERENCES**


[SZP89b] Szpankowski, W., Some thoughts on the ultimate stability condition for G/GI/1 queue, Purdue University, CSD TR-876, 1989.

[SzR87] Szpankowski, W., Rego, V., Ultimate stability conditions for some multidimensional distributed systems, Purdue University, CSD TR-715, 1987.


Figure 1. Illustration of \( \{\tau_n\}_{n=0}^\infty \) and \( \{\tau^*_n\}_{n=0}^\infty \) in an MGI1 queue with vacation.
Figure 2. Stability region for $M = 3$ users in slotted ALOHA system.